Guida per l'utente

Piattaforma Claude su AWS

Piattaforma Claude su AWS: Guida per l’utente

Table of Contents

Cos'è Claude Platform su AWS?

Accedi alle funzionalità complete della piattaforma Claude tramite AWS con Anthropic-managed infrastruttura.

Claude Platform su AWS ti offre l'esperienza completa della piattaforma Anthropic, tra cui l'API Messages, Agent Skills, l'esecuzione del codice e le funzionalità beta, accessibili tramite il tuo account AWS. A differenza di Amazon Bedrock, dove AWS gestisce lo stack di inferenza, Anthropic gestisce Claude Platform su AWS. AWS fornisce il livello di autenticazione (SigV4 o chiave API), il controllo degli IAM-based accessi e l'integrazione della fatturazione tramite AWS Marketplace.

> **Note**
>
> Gli SDK Anthropic supportano Claude Platform su AWS. Per la disponibilità dei client per lingua, consulta la documentazione degli SDK di Anthropic Client.

Come funziona l'integrazione della piattaforma

I modelli di Claude funzionano sull' Anthropic-managed infrastruttura. Si tratta di un'integrazione commerciale per la fatturazione e l'accesso tramite AWS. Sia Anthropic che AWS agiscono come processori di dati indipendenti. I termini di servizio AWS regolano Claude Platform su AWS. Si applicano anche i termini di servizio commerciali, l'addendum sull'elaborazione dei dati e la politica di utilizzo di Anthropic.

Nota le seguenti caratteristiche operative. I dati potrebbero non risiedere in AWS. L'inferenza può essere indirizzata al cloud principale di Anthropic. I sottoservizi possono cambiare senza preavviso. Imposta il `inference_geo` parametro per richiesta per associare l'inferenza a una geografia specifica. Vedi Data residency per i dettagli.

La piattaforma Claude su AWS segue la stessa politica di conservazione dei dati dell'API Claude di prima parte. Zero Data Retention (ZDR) è disponibile su richiesta. Contatta il rappresentante del tuo account Anthropic per abilitarlo per il tuo account.

Caratteristiche della piattaforma Claude su AWS

La piattaforma Claude su AWS offre le seguenti funzionalità:

- Same-day accesso a modelli e funzionalità: i nuovi modelli e le nuove funzionalità API di Claude sono disponibili lo stesso giorno del lancio sull'API Claude di prima parte.
- Agent Skills: utilizza le competenze predefinite per la generazione di documenti (ExcelPowerPoint, Word, PDF) e crea competenze personalizzate.
- Esecuzione del codice: esegui il codice nella sandbox gestita di Anthropic.
- Pensiero esteso: abilita il pensiero esteso per attività di ragionamento complesse.
- Streaming ed elaborazione in batch: utilizza lo streaming SSE in tempo reale o invia richieste in batch per carichi di lavoro ad alta produttività.
- Memorizzazione rapida nella cache: strumenti di cache, prompt di sistema e cronologia dei messaggi per ridurre latenza e costi.
- API Files: carica e fai riferimento ai file tra le richieste.
- Integrazione con AWS IAM: controlla l'accesso con policy IAM standard e autenticazione SigV4.
- Fatturazione AWS unificata: paga tramite il tuo account AWS esistente con misurazione basata sul consumo (Marketplace come backend di fatturazione).

Per l'elenco completo delle funzionalità supportate, consulta Feature support.

Come è organizzata questa guida

Questa guida tratta i seguenti argomenti:

- Guida introduttiva: configurazione dell'account, prerequisiti, autenticazione e installazione dell'SDK.
- Utilizzo dell'API: modelli disponibili, invio di richieste e supporto delle funzionalità.
- Gestione dell'ambiente: residenza dei dati, aree di lavoro, console Claude, limiti di tariffa e fatturazione.
- Monitoraggio e operazioni: CloudTrail registrazione e tracciamento degli ID delle richieste.
- Migrazione: passaggio da Amazon Bedrock a Claude Platform su AWS.
- Sicurezza e controllo degli accessi: riferimento alle politiche e alle azioni IAM.

Piattaforma Claude su AWS e Amazon Bedrock

Entrambe le offerte consentono di utilizzare Claude tramite AWS, ma differiscono in modo significativo per architettura, superficie API e disponibilità delle funzionalità.

	Piattaforma Claude su AWS	Amazon Bedrock
Chi gestisce lo stack	Anthropic	AWS
Superficie API	API per messaggi antropici () `/v1/messages`	API Bedrock (Converse/) InvokeModel
Disponibilità delle funzionalità	Same-day come Claude API	Dipende dalla tempistica di integrazione con AWS
Competenze degli agenti	Disponibilità	Non disponibile (richiede l'esecuzione di codice)
Funzionalità beta	Passa attraverso `anthropic-beta` gli header (vedi Supporto delle funzionalità)	Richiede il supporto di AWS
Autenticazione	AWS IAM/SigV4 o chiave API	AWS IAM/SigV4 o token bearer (solo SDK C#, Go e Java)
URL di base	`aws-external-anthropic.{region}.api.aws`	`bedrock-runtime.{region}.amazonaws.com`
Client SDK	Platform-specific classe client (ad esempio, `AnthropicAWS` in Python), in versione beta	`AnthropicBedrock` /Bedrock SDK
Console	Console Claude (`platform.claude.com` accesso tramite la console AWS)	Console Bedrock

	Piattaforma Claude su AWS	Amazon Bedrock
Limiti tariffari e quote	Gestito da Anthropic	Gestito da AWS
Processori di dati	Anthropic e AWS	AWS

Se hai bisogno di AWS-operated Claude con l'API Bedrock, consulta Amazon Bedrock.

> **Important**
>
> Anthropic fornisce Claude Platform su AWS come offerta di terze parti. Non è coperto da programmi di conformità, certificazioni o report di audit standard di AWS (come l'idoneità SOC, ISO o HIPAA). I clienti sono gli unici responsabili dell'esecuzione della propria due diligence per garantire che questa offerta di terze parti soddisfi i loro requisiti normativi, legali e di conformità.

Quando scegliere Bedrock: scegli Amazon Bedrock se la tua organizzazione richiede AWS-operated inferenza, AWS come unico processore di dati o copertura nell'ambito dei programmi di conformità AWS (tra cui l'idoneità a FedRAMP High, IL4, IL5, SOC, ISO e HIPAA). Bedrock funziona interamente su AWS-controlled un'infrastruttura con AWS come parte operativa.

Configurare l'account

La configurazione della piattaforma Claude su AWS prevede quattro fasi: registrazione alla console AWS, configurazione dell'organizzazione Anthropic, annotazione dell'ID dell'area di lavoro e accesso alla console Claude.

> **Note**
>
> La registrazione tramite la console AWS fornisce una nuova organizzazione Anthropic collegata al tuo account AWS. Questa organizzazione è separata da tutte le organizzazioni esistenti che la tua azienda ha con Anthropic, comprese le organizzazioni Claude Enterprise acquistate tramite AWS Marketplace. Le chiavi API, le aree di lavoro e le impostazioni della Claude Console di un'organizzazione Anthropic proprietaria non vengono trasferite.
> Se hai già un'offerta privata Amazon Bedrock, contatta il tuo account executive Anthropic o AWS prima di registrarti, in modo che lo sconto venga applicato dalla prima richiesta. Gli sconti non possono essere applicati retroattivamente all'utilizzo effettuato prima dell'accettazione dell'offerta privata. Vedi le offerte private.

Fase 1: Registrati nella console AWS

1. Apri la console AWS e accedi alla pagina del servizio Claude Platform on AWS.
2. Scegli Sign up.
3. Scegli Continua.

La pagina mostra un banner Sign-up in corso. Rimani sulla pagina. Sign-up impiega alcuni minuti mentre AWS gestisce l'abbonamento AWS Marketplace per te, quindi ti reindirizza automaticamente.

Se la tua organizzazione ha un'offerta privata di Anthropic, la Console la cerca e ti chiede di accettarla in AWS Marketplace. Per maggiori dettagli, consulta la sezione Offerte private.

> **Note**
>
> Se utilizzi Claude Platform su AWS, i tuoi contenuti (come istruzioni e completamenti) vengono elaborati da Anthropic all'esterno di AWS. Consulta le politiche di utilizzo dei dati di Anthropic per dettagli su come i contenuti e i metadati vengono elaborati e archiviati.

Passaggio 2: configura la tua organizzazione Anthropic

Una volta completata la registrazione, verrai reindirizzato a. `platform.claude.com/partner-signup`

1. Inserisci l'indirizzo email del proprietario dell'organizzazione e scegli Inizia.
2. Controlla nella casella di posta elettronica il link di configurazione e seguilo. Se il tuo browser mostra la pagina Accedi come account diverso, scegli Esci e continua.
3. Completa il modulo con i dettagli dell'organizzazione (nome dell'organizzazione, tipo di entità, paese, uso previsto) e scegli Configurazione completa.

Il completamento della configurazione crea la tua organizzazione Anthropic. I termini di servizio AWS regolano Claude Platform su AWS. Si applicano anche i termini di servizio commerciali, l'addendum sull'elaborazione dei dati e la politica di utilizzo di Anthropic. La pagina del servizio della console AWS ora mostra una navigazione a sinistra con Home, chiavi API, Quickstart e Workspaces.

Passaggio 3: annota l'ID del tuo spazio di lavoro

Al momento della registrazione, viene fornito automaticamente uno spazio di lavoro predefinito. È possibile creare aree di lavoro aggiuntive per regione; consulta Workspaces per i dettagli sull'associazione delle regioni, sulla gestione degli spazi di lavoro e sull'ambito delle risorse IAM.

Trova l'ID dello spazio di lavoro nella pagina del servizio Workspaces sulla console AWS Claude Platform on AWS o nella console Claude (vedi Uso della console Claude). Gli ID di Workspace utilizzano il formato `wrkspc_` seguito da un identificatore alfanumerico.

Passaggio 4: accedi alla console Claude

L'accesso alla console Claude è federato tramite AWS IAM:

1. Assumi un ruolo IAM con l'`aws-external-anthropic:AssumeConsole`autorizzazione. Vedi le azioni IAM.
2. Dalla pagina del servizio Claude Platform on AWS, scegli Accedi. La console AWS emette un JWT e ti reindirizza a. `platform.claude.com`
3. Al primo accesso, ti viene richiesto un indirizzo e-mail. Inserisci la tua email di lavoro. La piattaforma effettua il provisioning dell'utente della Claude Console in tempo utile.

Quando effettui l'accesso tramite la console AWS, la console Claude si riferisce alla tua organizzazione Claude Platform on AWS. Un indicatore Account managed by AWS viene visualizzato nella barra laterale della console Claude.

Risoluzione dei problemi di configurazione dell'account

- "Sign-up fallito: Attivazione non riuscita OutboundWebIdentityFederation «: Se vedi questo banner rosso al primo invio, scegli di nuovo Continua. La propagazione dell'abilitazione IAM può richiedere alcuni istanti.
- Nessun indicatore di avanzamento durante la registrazione: Sign-up richiede alcuni minuti. La pagina mostra un banner statico Sign-up in corso senza barra di avanzamento mentre AWS effettua il provisioning del tuo account.
- «Accedi come account diverso» dopo aver seguito il link di configurazione: scegli Esci e continua. La pagina ti autentica nuovamente con l'indirizzo email che hai inserito.
- Messaggio «Non trovato» durante l'accesso: questo messaggio potrebbe apparire brevemente durante il reindirizzamento. Puoi ignorarlo.
- La pagina di utilizzo non mostra dati dopo la prima chiamata all'API: i dati di utilizzo possono richiedere alcuni minuti prima che vengano visualizzati nella console Claude.

Prerequisiti

Prima di effettuare chiamate API, assicurati di avere:

1. Un account AWS attivo con un abbonamento a Claude Platform on AWS (vedi Configurazione dell'account).
2. L'AWS CLI è stata installata e configurata.
3. La federazione delle identità Web in uscita è abilitata sul tuo account AWS (una sola operazione di configurazione; consulta Abilitare la federazione delle identità Web in uscita).
4. Il tuo ID dello spazio di lavoro (vedi Ottieni l'ID del tuo spazio di lavoro).

Abilita la federazione delle identità Web in uscita

La piattaforma Claude sul gateway AWS chiama `sts:GetWebIdentityToken` lato server per coniare un JWT che inoltra ad Anthropic. Questa funzionalità STS è disabilitata per impostazione predefinita su ogni account AWS. Attivala una volta per account:

```
aws iam enable-outbound-web-identity-federation
```

Se la risposta è`[ERROR] (FeatureEnabled) … already enabled` , l'impostazione è già attiva per il tuo account e puoi andare avanti. Verifica e recupera l'URL emittente del tuo account:

```
aws iam get-outbound-web-identity-federation-info
```

> **Warning**
>
> Senza questo passaggio, ogni richiesta viene restituita. `"Outbound web identity federation is disabled for your account"` Questo è l'errore di configurazione più comune.

Ottieni l'ID del tuo spazio di lavoro

Al momento della registrazione viene fornito automaticamente uno spazio di lavoro predefinito in ogni regione (vedi Configurazione dell'account). Gli spazi di lavoro sono associati a un'unica regione

AWS. Puoi trovare l'ID dello spazio di lavoro nella console Claude sotto Workspaces o nella sezione Workspaces della pagina del servizio della console AWS. Consulta Workspaces per l'associazione delle regioni e l'ambito delle risorse IAM.

Imposta le variabili `ANTHROPIC_AWS_WORKSPACE_ID` e di `AWS_REGION` ambiente in modo che i client SDK le leggano automaticamente:

```
export ANTHROPIC_AWS_WORKSPACE_ID='wrkspc_01AbCdEf23GhIj'
export AWS_REGION='us-west-2'
```

La regione è obbligatoria. Il client SDK viene lanciato se non è impostata alcuna regione. Passa`aws_region`/`awsRegion`al costruttore o imposta `AWS_REGION` (or). `AWS_DEFAULT_REGION` Tutte le regioni commerciali AWS sono supportate; le regioni opt-in richiedono prima l'attivazione dell'account nella regione.

Autenticazione

La piattaforma Claude su AWS supporta due metodi di autenticazione: AWS IAM con firma delle richieste SigV4 (primaria) e autenticazione con chiave API. Entrambi utilizzano lo stesso URL di base e lo stesso formato di richiesta.

Gli SDK Anthropic (vedi Installare un SDK) implementano il flusso di autenticazione e la risoluzione delle credenziali descritti di seguito. Se non utilizzi un SDK Anthropic, devi creare SigV4-signed richieste (o presentare la tua chiave API come token portante) sull'endpoint regionale. `https://aws-external-anthropic.<region>.api.aws` Per la configurazione SDK-specific del client e l'ordine autorevole di risoluzione delle credenziali, consulta la guida all'installazione di Claude on AWS nella documentazione di Anthropic.

Autenticazione SigV4

SigV4 è il percorso aziendale nativo e si integra con le policy, i ruoli e gli audit AWS IAM esistenti. Configura le credenziali AWS utilizzando qualsiasi metodo supportato dalla catena di provider di credenziali AWS predefinita:

- Variabili di ambiente (`AWS_ACCESS_KEY_ID`,,`AWS_SECRET_ACCESS_KEY`) `AWS_SESSION_TOKEN`
- File di credenziali condiviso () `~/.aws/credentials`
- File di configurazione condiviso (`~/.aws/config`) che include SSO e `credential_process`
- Identità Web (`AWS_WEB_IDENTITY_TOKEN_FILEeAWS_ROLE_ARN`) per IRSA e Actions GitHub
- Credenziali del contenitore ECS
- Servizio di metadati delle istanze EC2 (IMDS)

Per verificare che Anthropic SDK stia raccogliendo le tue credenziali e si stia risolvendo nell'endpoint regionale corretto, effettua una richiesta di test seguendo gli esempi in Effettuare richieste. Una risposta corretta conferma che le credenziali, la regione, l'URL di base e l'ID dell'area di lavoro sono tutti configurati correttamente.

Autenticazione con chiave API

Per percorsi di integrazione più semplici (sviluppo locale e script), puoi autenticarti con una chiave API anziché SigV4. Imposta la variabile di `ANTHROPIC_AWS_API_KEY` ambiente

o passa al costruttore SDK`apiKey`. L'SDK Anthropic invia la chiave come token portante all'endpoint Claude Platform on AWS; IAM autorizza la richiesta tramite l'`aws-external-anthropic:CallWithBearerToken`enazione (vedi policy IAM).

Genera chiavi API nella console AWS sotto Claude Platform su AWS → chiavi API. Scegli Genera una chiave, quindi copia il valore della chiave. Concedi l'azione `aws-external-anthropic:CallWithBearerToken` IAM ai principali che dovrebbero essere autorizzati a utilizzare l'autenticazione tramite chiave API.

> **Note**
>
> Solo le chiavi API create nella console AWS sotto la piattaforma Claude su AWS funzionano con questo servizio.
>
> - Le chiavi create nella console Claude standard per l'accesso alle API di prime parti non funzionano con la piattaforma Claude sull'endpoint AWS.
> - Anche le chiavi API di Amazon Bedrock non funzionano: Bedrock utilizza un endpoint, un flusso di autenticazione e uno spazio dei nomi IAM separati.

Short-term Chiavi API

Nei casi in cui desideri l'autenticazione tramite chiave API ma senza la durata di una chiave a lunga durata, Anthropic pubblica librerie di generatori di token che coniano chiavi API a breve termine dalle tue credenziali AWS IAM. Per impostazione predefinita, i token hanno una durata di 12 ore e possono essere assegnati a uno spazio di lavoro specifico e all'azione. `aws-external-anthropic:CallWithBearerToken` Installa il generatore per il tuo linguaggio, scambia le credenziali IAM con una chiave API e passa la chiave all'SDK Anthropic.

- Python: -generatore-per-aws-external-anthropic-python aws/token
- JavaScript / TypeScript: aws/token-generatore-per-aws-esternal-anthropic-js
- Java: aws/token -generatore-per-aws-java-esterno-antropico

Short-term Le chiavi API richiedono comunque che il principale IAM del chiamante rimanga nell'area di lavoro di destinazione. `aws-external-anthropic:CallWithBearerToken` Scadono da sole e non devono essere ruotate o revocate in modo esplicito.

Priorità delle credenziali e risoluzione della regione

Il client Claude Platform on AWS di Anthropic SDK risolve le credenziali e la regione utilizzando un ordine di precedenza definito. I nomi degli argomenti seguono le convenzioni di ogni linguaggio: CamelCase TypeScript per e PHP, snake_case per Python e PascalCase Ruby, per Go e modelli di proprietà o builder per C# e Java.

Per l'ordine di precedenza autorevole, le variabili di ambiente supportate e gli argomenti del costruttore per ogni SDK, consulta Claude sulla configurazione di AWS nella documentazione di Anthropic. In generale, gli argomenti espliciti del costruttore hanno la precedenza sulle variabili di ambiente e sulla catena di `ANTHROPIC_AWS_API_KEY` provider di credenziali AWS predefinita. La regione è obbligatoria; a differenza di `AnthropicBedrock` (che ricade su`us-east-1`), il client Claude on AWS genera se nessuna regione viene fornita dal costruttore o da/. `AWS_REGION` `AWS_DEFAULT_REGION`

ID WorkSpace

Ogni richiesta del piano dati deve includere l'ID dello spazio di lavoro nell'intestazione. `anthropic-workspace-id` Per impostazione predefinita, gli SDK Anthropic lo leggono dalla variabile di `ANTHROPIC_AWS_WORKSPACE_ID` ambiente, oppure puoi passare`workspaceId`/`workspace_id`al costruttore del client. Se utilizzi il client di base (non il `Anthropic` Claude-on-AWS client), passa l'intestazione in modo esplicito. Vedi Richieste di esempi per lingua e Workspaces per sapere come individuare l'ID del tuo workspace.

Installa un SDK di

Gli SDK client di Anthropic supportano Claude Platform su AWS. Tutti e sette gli SDK forniscono una classe client specifica per la piattaforma; in alternativa, puoi configurare il client di base con la piattaforma Claude Platform sull'URL di base AWS.

Python

```
pip install -U "anthropic[aws]"
```

> **Tip**
>
> Su macOS con Homebrew Python o altri ambienti `pip install` Python gestiti esternamente, potrebbe fallire con un errore PEP 668. `externally-managed-environment` Crea e attiva prima un ambiente virtuale:. `python3 -m venv .venv && source .venv/bin/activate`

TypeScript

```
npm install @anthropic-ai/aws-sdk
```

C#

```
dotnet add package Anthropic.Aws
```

Go

```
go get github.com/anthropics/anthropic-sdk-go
```

Java

```
implementation("com.anthropic:anthropic-java-aws:2.27.0")
```

```
<dependency>
  <groupId>com.anthropic</groupId>
  <artifactId>anthropic-java-aws</artifactId>
```

```
  <version>2.27.0</version>
</dependency>
```

PHP

```
composer require anthropic-ai/sdk aws/aws-sdk-php
```

Ruby

```
gem install anthropic aws-sdk-core
```

> **Note**
>
> I client SDK per Claude Platform su AWS sono in versione beta.

Modelli disponibili

Claude Platform su AWS offre lo stesso set di modelli Claude dell'API Claude di prima parte.

Per l'elenco aggiornato di modelli, ID dei modelli, dimensioni delle finestre di contesto e funzionalità, consulta la panoramica dei modelli nella documentazione di Anthropic.

ID modello

Gli ID modello su Claude Platform su AWS sono identici a quelli utilizzati nell'API Claude di prime parti (ad esempio,`claude-opus-4-6`,`claude-sonnet-4-6`). `claude-haiku-4-5` Non ci sono Bedrock-style ARN o `anthropic.` prefissi: usa l'ID del modello esattamente come lo pubblica Anthropic.

Esecuzione di richieste

La piattaforma Claude su AWS utilizza l'API Anthropic Messages (`/v1/messages`), la stessa superficie API della piattaforma proprietaria Claude. Le differenze sono l'URL di base, il metodo di autenticazione e un'`anthropic-workspace-id`intestazione obbligatoria che identifica l'area di lavoro a cui è destinata la richiesta.

Shell

```
curl "https://aws-external-anthropic.us-west-2.api.aws/v1/messages" \
  --aws-sigv4 "aws:amz:us-west-2:aws-external-anthropic" \
  --user "$AWS_ACCESS_KEY_ID:$AWS_SECRET_ACCESS_KEY" \
  -H "x-amz-security-token: $AWS_SESSION_TOKEN" \
  -H "content-type: application/json" \
  -H "anthropic-version: 2023-06-01" \
  -H "anthropic-workspace-id: $ANTHROPIC_AWS_WORKSPACE_ID" \
  -d '{
    "model": "claude-sonnet-4-6",
    "max_tokens": 1024,
    "messages": [
      {"role": "user", "content": "Hello!"}
    ]
  }'
```

Python

```
from anthropic import AnthropicAWS

client = AnthropicAWS()

message = client.messages.create(
    model="claude-sonnet-4-6",
    max_tokens=1024,
    messages=[{"role": "user", "content": "Hello!"}],
)
print(message)
```

TypeScript

```
import AnthropicAws from "@anthropic-ai/aws-sdk";
```

```
const client = new AnthropicAws();

const message = await client.messages.create({
  model: "claude-sonnet-4-6",
  max_tokens: 1024,
  messages: [{ role: "user", content: "Hello!" }]
});
console.log(message);
```

C#

```
using Anthropic;
using Anthropic.Aws;

var client = new AnthropicAwsClient();

var message = await client.Messages.Create(new()
{
    Model = "claude-sonnet-4-6",
    MaxTokens = 1024,
    Messages = [new() { Role = "user", Content = "Hello!" }]
});

Console.WriteLine(message);
```

Go

```
client, err := anthropicaws.NewClient(context.Background(),
 anthropicaws.ClientConfig{})
if err != nil {
    panic(err)
}

message, err := client.Messages.New(context.Background(), anthropic.MessageNewParams{
    Model:     anthropic.ModelClaudeSonnet4_6,
    MaxTokens: 1024,
    Messages: []anthropic.MessageParam{
        anthropic.NewUserMessage(anthropic.NewTextBlock("Hello!")),
    },
})
if err != nil {
    panic(err)
```

```
}

fmt.Println(message.Content[0].Text)
```

Java

```
import com.anthropic.aws.backends.AwsBackend;
import com.anthropic.client.AnthropicClient;
import com.anthropic.client.okhttp.AnthropicOkHttpClient;
import com.anthropic.models.messages.Message;
import com.anthropic.models.messages.MessageCreateParams;
import com.anthropic.models.messages.Model;

AnthropicClient client = AnthropicOkHttpClient.builder()
    .backend(AwsBackend.fromEnv())
    .build();

Message message = client.messages().create(
    MessageCreateParams.builder()
        .model(Model.CLAUDE_SONNET_4_6)
        .maxTokens(1024)
        .addUserMessage("Hello!")
        .build()
);

IO.println(message);
```

PHP

```
use Anthropic\Aws\Client;

$client = new Client();

$message = $client->messages->create(
    model: 'claude-sonnet-4-6',
    maxTokens: 1024,
    messages: [['role' => 'user', 'content' => 'Hello!']],
);

echo $message;
```

Ruby

```
require "anthropic"

client = Anthropic::AWSClient.new

message = client.messages.create(
  model: "claude-sonnet-4-6",
  max_tokens: 1024,
  messages: [{ role: "user", content: "Hello!" }]
)

puts message
```

Il client legge `AWS_REGION` (o`AWS_DEFAULT_REGION`) e `ANTHROPIC_AWS_WORKSPACE_ID` dall'ambiente. È possibile eseguire l'override passando`aws_region`/`awsRegion`o`workspace_id`/`workspaceId`al costruttore. Sono obbligatori sia l'ID della regione che l'ID dell'area di lavoro; il costruttore genera se uno dei due non può essere risolto da queste fonti.

> **Note**
>
> L'`x-amz-security-token`intestazione (curl) è richiesta solo per credenziali temporanee come ruoli IAM, SSO o STS. Omettilo quando usi credenziali utente IAM a lungo termine. I client SDK lo gestiscono automaticamente in base alla fonte delle credenziali.

Il `--aws-sigv4` valore segue il formato. `aws:amz:<region>:<service>` Il nome del servizio SigV4 è e `aws-external-anthropic` la regione deve corrispondere alla regione nell'URL dell'endpoint. Una mancata corrispondenza in uno dei due genera un errore generico di rifiuto della firma anziché una diagnostica specifica.

Utilizzo del client Anthropic di base

Se preferisci non aggiungere la dipendenza dell'SDK AWS, puoi utilizzare il client di base`Anthropic`. Questo percorso utilizza i nomi delle variabili di ambiente Vanilla SDK anziché i `ANTHROPIC_AWS_*` nomi letti dal client dedicato:

```
export ANTHROPIC_API_KEY='your-api-key'
export ANTHROPIC_BASE_URL='https://aws-external-anthropic.us-west-2.api.aws'
export ANTHROPIC_WORKSPACE_ID='your-workspace-id'
```

Gli SDK Python e Go leggono e vengono letti `ANTHROPIC_API_KEY` `ANTHROPIC_BASE_URL` automaticamente; per gli altri linguaggi, passali al costruttore. TypeScript In ogni lingua, passa l'ID dell'area di lavoro nell'intestazione. `anthropic-workspace-id`

Python

```
import os
from anthropic import Anthropic

client = Anthropic(
    # ANTHROPIC_API_KEY and ANTHROPIC_BASE_URL are read automatically
    default_headers={"anthropic-workspace-id": os.environ["ANTHROPIC_WORKSPACE_ID"]},
)

message = client.messages.create(
    model="claude-sonnet-4-6",
    max_tokens=1024,
    messages=[{"role": "user", "content": "Hello!"}],
)
print(message.content[0].text)
```

TypeScript

```
import Anthropic from "@anthropic-ai/sdk";

const client = new Anthropic({
  // ANTHROPIC_API_KEY and ANTHROPIC_BASE_URL are read automatically
  defaultHeaders: { "anthropic-workspace-id": process.env.ANTHROPIC_WORKSPACE_ID }
});

const message = await client.messages.create({
  model: "claude-sonnet-4-6",
  max_tokens: 1024,
  messages: [{ role: "user", content: "Hello!" }]
});
console.log(message.content);
```

Supporto funzionalità

Claude Platform su AWS utilizza direttamente l'API Anthropic Messages. Ottieni la piena parità di funzionalità con l'API Claude di prima parte, tranne dove indicato in Caratteristiche non attualmente disponibili:

- Funzionalità beta: passa l'`anthropic-beta`intestazione standard per accedere alle funzionalità beta, proprio come faresti con l'API Claude.
- Agent Skills: utilizza Agent Skills predefinite e personalizzate con gli stessi `container.skills` parametri e le stesse intestazioni beta dell'API Claude. Tutte le competenze predefinite (ExcelPowerPoint, Word, PDF) funzionano immediatamente.
- Esecuzione del codice: esegui il codice nella sandbox gestita di Anthropic utilizzando lo strumento di esecuzione del codice.
- Utilizzo degli strumenti: sono disponibili l'uso del computer e tutte le altre funzionalità di utilizzo degli strumenti.
- Pensiero esteso: abilita il pensiero esteso con gli stessi parametri dell'API Claude.
- Streaming: supporto completo allo streaming SSE per risposte in tempo reale.
- Elaborazione in batch: invia richieste in batch per carichi di lavoro ad alta produttività.
- Memorizzazione rapida nella cache: strumenti di cache, prompt di sistema e cronologia dei messaggi per ridurre latenza e costi. Sono disponibili tutte le funzionalità di memorizzazione rapida nella cache (TTL a 5 minuti, TTL a 1 ora e memorizzazione nella cache automatica).
- API Files: carica e fai riferimento ai file tra le richieste.
- Tag Workspace con integrazione IAM: le aree di lavoro possono essere etichettate e i tag passano a IAM per il controllo degli accessi basato sugli attributi e ai report di costo e utilizzo di AWS per l'allocazione dei costi. Il tagging dello spazio di lavoro è una funzionalità GA su Claude Platform su AWS (è una funzionalità beta sull'API Claude di prima parte).

Agenti gestiti da Claude

I Managed Agent di Claude sono disponibili sulla piattaforma Claude su AWS. Agenti, sessioni, ambienti, archivi di credenziali e archivi di memoria sono risorse IAM di prima classe. Vedi Azioni IAM per il riferimento all'azione e Uso della console Claude per le pagine corrispondenti. L'SDK Anthropic Agent funziona con Claude Platform su AWS senza alcuna integrazione aggiuntiva. Indirizzalo

all'URL di base della piattaforma Claude su AWS ed esegui l'autenticazione con SigV4 o una chiave API.

Le sessioni autonome su Claude Platform su AWS richiedono una nuova autenticazione ogni 6 ore. Long-running le esecuzioni degli agenti devono aggiornare le credenziali SigV4 o la chiave API all'interno di quella finestra, altrimenti la sessione termina.

Le seguenti funzionalità di Claude Managed Agents non sono attualmente disponibili su Claude Platform on AWS:

- Risultati: il monitoraggio dei risultati per le sessioni degli agenti non è disponibile.
- Multi-agent sessioni: le sessioni con più agenti interagenti non sono disponibili.

Conformità

> **Important**
>
> Anthropic fornisce questo servizio come offerta di terze parti. Non è coperto da programmi di conformità, certificazioni o report di audit standard di AWS (come l'idoneità SOC, ISO o HIPAA). I clienti sono gli unici responsabili dell'esecuzione della propria due diligence per garantire che questa offerta di terze parti soddisfi i loro requisiti normativi, legali e di conformità. Prima di trattare dati sensibili o regolamentati, è necessario consultare la documentazione ufficiale di conformità di Anthropic tramite l'Anthropic Trust Center. Per ulteriori informazioni, consulta i termini del servizio AWS.

Come viene modellata l'iscrizione allo spazio di lavoro

Nell'API Claude di prima parte, l'accesso è regolato dall'appartenenza degli utenti agli spazi di lavoro: un utente viene aggiunto a uno spazio di lavoro ed eredita i diritti di accesso dell'area di lavoro. Claude Platform su AWS sostituisce quel modello con l'autorizzazione IAM: non esiste un elenco di utenti per area di lavoro. Invece, i responsabili IAM (utenti o ruoli) dispongono di politiche che `aws-external-anthropic` concedono azioni contro ARN di aree di lavoro specifiche. La valutazione delle policy IAM determina cosa può fare ogni responsabile in ogni spazio di lavoro.

Concedi e revoca l'accesso allo spazio di lavoro modificando le policy IAM (SCP, limiti di autorizzazione, politiche legate all'identità o condizioni a livello di risorsa) anziché gestire l'appartenenza per area di lavoro nella Console Claude. Consulta le politiche IAM per alcuni esempi.

Funzionalità non attualmente disponibili

Le seguenti funzionalità non sono attualmente disponibili su Claude Platform on AWS:

- Conformità all'HIPAA: il HIPAA-ready programma di Anthropic non è disponibile su Claude Platform on AWS. I clienti con requisiti HIPAA dovrebbero invece valutare Claude in Amazon Bedrock.
- Livelli di servizio oltre Standard e Batch: il livello prioritario non è disponibile.
- API di amministrazione: endpoint per membri dell'organizzazione, membri dell'area di lavoro, invito, chiave API, report sull'utilizzo, report sui costi e report sui limiti di velocità: questi endpoint dell'API di amministrazione non sono attualmente disponibili. Workspace CRUD e rename endpoints (`/v1/organizations/workspaces`) sono disponibili tramite il namespace; consulta le azioni IAM. `aws-external-anthropic` L'iscrizione è modellata tramite AWS IAM anziché tramite elenchi di membri di workspace (vedi la sezione precedente). Il ciclo di vita delle chiavi API è gestito nella console AWS sotto la piattaforma Claude su AWS → chiavi API. Per i dati su utilizzo, costi e limiti di velocità, utilizza le viste della console Claude (vedi Monitoraggio e registrazione) o l'API Anthropic Usage and Cost.
- Limiti di spesa: non disponibili. Affidati invece ai controlli di fatturazione di AWS.
- Spazio di lavoro Claude Code e API Analytics: l'area di lavoro Claude Code con limiti di velocità automatici non è disponibile. L'utilizzo di Claude Code viene visualizzato nella visualizzazione di utilizzo generale anziché in una schermata dedicata.
- Autenticazione OAuth: non supportata. Usa SigV4 o l'autenticazione con chiave API.
- OpenAI-compatible Endpoint API: non disponibile su Claude Platform on AWS.
- Workspace-level controlli geografici di inferenza: `allowed_inference_geos` e `default_inference_geo` non sono disponibili. Imposta `inference_geo` invece su ogni richiesta.

Residenza dei dati

Claude Platform su AWS supporta le seguenti aree geografiche di inferenza:

- Stati Uniti: l'inferenza rimane all'interno dei data center statunitensi. Si applica un moltiplicatore di prezzo 1,1x.
- Globale: l'inferenza può essere indirizzata a qualsiasi data center in tutto il mondo. Anthropic-operated Si applica il prezzo standard.

Imposta la geografia di inferenza per richiesta con il `inference_geo` parametro:

> **Note**
>
> Il `inference_geo` parametro è supportato su Claude Opus 4.6, Claude Sonnet 4.6 e modelli successivi. Le richieste con `inference_geo` Claude Opus 4.5, Claude Sonnet 4.5 o Claude Haiku 4.5 restituiscono un errore 400. Vedi Data residency per i dettagli sulla disponibilità del modello.

Shell

```
curl "https://aws-external-anthropic.us-west-2.api.aws/v1/messages" \
  --aws-sigv4 "aws:amz:us-west-2:aws-external-anthropic" \
  --user "$AWS_ACCESS_KEY_ID:$AWS_SECRET_ACCESS_KEY" \
  -H "x-amz-security-token: $AWS_SESSION_TOKEN" \
  -H "content-type: application/json" \
  -H "anthropic-version: 2023-06-01" \
  -H "anthropic-workspace-id: $ANTHROPIC_AWS_WORKSPACE_ID" \
  -d '{
    "model": "claude-sonnet-4-6",
    "max_tokens": 1024,
    "inference_geo": "us",
    "messages": [
      {"role": "user", "content": "Hello!"}
    ]
  }'
```

Python

```
from anthropic import AnthropicAWS

client = AnthropicAWS(aws_region="us-west-2")
message = client.messages.create(
    model="claude-sonnet-4-6",
    max_tokens=1024,
    inference_geo="us",
    messages=[{"role": "user", "content": "Hello!"}],
)
print(message)
```

TypeScript

```
import AnthropicAws from "@anthropic-ai/aws-sdk";
const client = new AnthropicAws({ awsRegion: "us-west-2" });
const message = await client.messages.create({
  model: "claude-sonnet-4-6",
  max_tokens: 1024,
  inference_geo: "us",
  messages: [{ role: "user", content: "Hello!" }]
});
console.log(message);
```

C#

```
using Anthropic;
using Anthropic.Aws;

var client = new AnthropicAwsClient(new() { AwsRegion = "us-west-2" });

var message = await client.Messages.Create(new()
{
    Model = "claude-sonnet-4-6",
    MaxTokens = 1024,
    InferenceGeo = "us",
    Messages = [new() { Role = "user", Content = "Hello!" }]
});

Console.WriteLine(message);
```

Go

```
client, err := anthropicaws.NewClient(context.Background(),
 anthropicaws.ClientConfig{})
if err != nil {
    panic(err)
}

message, err := client.Messages.New(context.Background(), anthropic.MessageNewParams{
    Model:        anthropic.ModelClaudeSonnet4_6,
    MaxTokens:    1024,
    InferenceGeo: anthropic.String("us"),
    Messages: []anthropic.MessageParam{
        anthropic.NewUserMessage(anthropic.NewTextBlock("Hello!")),
    },
})
if err != nil {
    panic(err)
}

fmt.Println(message)
```

Java

```
import com.anthropic.aws.backends.AwsBackend;
import com.anthropic.client.AnthropicClient;
import com.anthropic.client.okhttp.AnthropicOkHttpClient;
import com.anthropic.models.messages.Message;
import com.anthropic.models.messages.MessageCreateParams;
import com.anthropic.models.messages.Model;
import software.amazon.awssdk.regions.Region;

AnthropicClient client = AnthropicOkHttpClient.builder()
    .backend(AwsBackend.builder().region(Region.of("us-west-2")).build())
    .build();

Message message = client.messages().create(
    MessageCreateParams.builder()
        .model(Model.CLAUDE_SONNET_4_6)
        .maxTokens(1024)
        .inferenceGeo("us")
        .addUserMessage("Hello!")
        .build()
);
```

```
IO.println(message);
```

PHP

```
use Anthropic\Aws\Client;

$client = new Client(awsRegion: 'us-west-2');

$message = $client->messages->create(
    model: 'claude-sonnet-4-6',
    maxTokens: 1024,
    inferenceGeo: 'us',
    messages: [['role' => 'user', 'content' => 'Hello!']],
);

echo $message;
```

Ruby

```
require "anthropic"

client = Anthropic::AWSClient.new(aws_region: "us-west-2")

message = client.messages.create(
  model: "claude-sonnet-4-6",
  max_tokens: 1024,
  inference_geo: "us",
  messages: [{ role: "user", content: "Hello!" }]
)

puts message
```

Se si ometteinference_geo, la richiesta viene impostata come predefinita. global

Workspace-level i controlli di inferenza geografica (allowed_inference_geosedefault_inference_geo) non sono disponibili su Claude Platform on AWS. Imposta invece inference_geo su ogni richiesta.

WorkSpace

Le richieste di inferenza e risorse su Claude Platform su AWS hanno come target uno spazio di lavoro. L'ID dell'area di lavoro viene passato nell'`anthropic-workspace-id`intestazione di queste chiamate API. Gli ID dell'area di lavoro utilizzano il formato con tag `wrkspc_` seguito da un identificatore alfanumerico (ad esempio,). `wrkspc_01AbCdEf23GhIj` Vedi Ottieni l'ID del tuo spazio di lavoro se non lo possiedi ancora.

Ambito dell'area di lavoro

Gli spazi di lavoro sono associati a un'unica regione AWS. È `us-west-2` possibile accedere a uno spazio di lavoro creato in solo tramite l'`us-west-2`endpoint. L'utilizzo, le quote, i costi, i file, i batch e le competenze sono tutti raggruppati per area di lavoro, offrendoti suddivisioni per regione nella console Claude.

Gli spazi di lavoro fungono anche da risorsa IAM principale per Claude Platform su AWS. Puoi concedere o negare l'accesso a spazi di lavoro specifici tramite le policy di AWS IAM utilizzando l'ARN dell'area di lavoro. Il segmento di risorse dell'ARN è lo stesso ID con `wrkspc_` prefisso che passi nell'intestazione: `anthropic-workspace-id`

```
arn:aws:aws-external-anthropic:{region}:{account-id}:workspace/{workspace-id}
```

Ad esempio: `arn:aws:aws-external-anthropic:us-west-2:123456789012:workspace/wrkspc_01AbCdEf23GhIj`

Consulta le politiche IAM per esempi di policy.

Creazione di spazi di lavoro

Uno spazio di lavoro predefinito viene fornito automaticamente al momento dell'iscrizione. È possibile creare, aggiornare, rinominare e archiviare aree di lavoro aggiuntive tramite gli `/v1/organizations/workspaces` endpoint, autorizzate dalle `aws-external-anthropic` azioni corrispondenti (`CreateWorkspace`,,); vedi Azioni IAM. `UpdateWorkspaceArchiveWorkspace`

Impostazione dell'ID dello spazio di lavoro nel client

Ogni richiesta del piano dati deve includere l'ID dell'area di lavoro nell'intestazione`anthropic-workspace-id`. Quando utilizzi un Claude-on-AWS client di Anthropic SDK, ci sono tre modi per fornirlo:

1. Variabile d'ambiente: `ANTHROPIC_AWS_WORKSPACE_ID` impostata e il client la legge automaticamente.

   ```
   export ANTHROPIC_AWS_WORKSPACE_ID='wrkspc_01AbCdEf23GhIj'
   ```

2. Argomento del costruttore: passa`workspaceId`/`workspace_id`(il nome segue la convenzione linguistica dell'SDK) quando si crea un'istanza del client.
3. Per-request override: passa l'intestazione direttamente su una singola richiesta se devi indirizzare più aree di lavoro dello stesso client.

Se utilizzi il `Anthropic` client di base (senza il backend AWS), imposta l'`anthropic-workspace-id`intestazione in modo esplicito su ogni richiesta (vedi Fare richieste per esempi per lingua). Per i nomi SDK-specific degli argomenti, consulta la documentazione degli SDK di Anthropic Client.

Taggare le aree di lavoro

I tag Workspace sono coppie chiave-valore collegate a un'area di lavoro. Si integrano con AWS IAM per il controllo degli accessi basato sugli attributi e con AWS Cost and Usage Reports per l'allocazione dei costi (vedi Monitoraggio e registrazione per i dettagli del CUR). Il tagging è GA su Claude Platform su AWS.

Aggiorna i tag sull'area di lavoro esistente con `TagResource` e `UntagResource` nel namespace. `aws-external-anthropic` Le stesse chiavi di tag possono determinare le condizioni IAM e l'allocazione dei costi senza configurazioni aggiuntive.

Tag-based controllo degli accessi

È possibile modificare le `aws-external-anthropic` azioni sui valori dei tag del workspace utilizzando la chiave di contesto della `aws:ResourceTag/<key>` condizione. La seguente politica consente l'inferenza solo sulle aree di lavoro contrassegnate: `Environment=production`

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "aws-external-anthropic:CreateInference",
        "aws-external-anthropic:CreateBatchInference",
        "aws-external-anthropic:CountTokens"
      ],
      "Resource": "arn:aws:aws-external-anthropic:*:*:workspace/*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/Environment": "production"
        }
      }
    }
  ]
}
```

Usa `aws:RequestTag/<key>` e attiva i `aws:TagKeys` tasti `TagResource` per applicare le politiche di etichettatura (ad esempio, richiedendo lo spazio di lavoro per il trasporto e l'etichettatura). `CostCenter` `Owner` Consulta le politiche IAM per altri esempi.

Utilizzo della console Claude

La piattaforma Claude su AWS utilizza la console Claude standard all'indirizzo platform.claude.com. Quando accedi dalla console AWS, nella barra laterale della console Claude viene visualizzato un indicatore Account managed by AWS. La console si riferisce alla tua organizzazione Claude Platform on AWS. Fornisce analisi di utilizzo, analisi dei costi, visibilità dei limiti di velocità, visibilità dell'area di lavoro e pagine per la gestione di file, Agent Skills, processi in batch, agenti, sessioni, ambienti, archivi di credenziali e archivi di memoria.

Accesso

L'accesso alla console Claude è federato tramite AWS IAM. Vedi Configurazione dell'account per il flusso completo di accesso al primo accesso. In breve:

1. Assumi un ruolo IAM con l'`aws-external-anthropic:AssumeConsole`autorizzazione. Vedi le azioni IAM.
2. Vai alla pagina Claude Platform on AWS nella Console AWS.
3. Scegli Open Claude Console. La console AWS emette un JWT e ti reindirizza a. `platform.claude.com`
4. Al primo accesso, ti viene richiesto un indirizzo e-mail; inserisci l'email di lavoro. La piattaforma effettua il provisioning dell'utente della Claude Console in tempo utile.

Sono disponibili due ruoli di Claude Console: Amministratore e Sviluppatore. Il ruolo di amministratore consente l'accesso a tutte le pagine e le impostazioni della console Claude disponibili per Claude Platform su AWS. Il ruolo Sviluppatore consente l'accesso in lettura a informazioni su utilizzo, costi, limiti di tariffa e spazio di lavoro. Contatta il rappresentante del tuo account Anthropic per assegnare il ruolo di Amministratore o Sviluppatore a un responsabile.

Pagine disponibili

La colonna Via AWS gateway indica se la pagina legge e scrive dati tramite il gateway AWS (ed è quindi governata da azioni IAM). Le pagine contrassegnate con No leggono i metadati a livello di organizzazione direttamente tramite le API Anthropic e aggirano i controlli delle azioni IAM.

Pagina	Disponibilità	Tramite gateway AWS	Note
Utilizzo	Sì	No	Visualizza l'utilizzo dei token per modello, area di lavoro e dimensione. La visualizzazione dei dati potrebbe richiedere alcuni minuti dopo una richiesta.
Costo	Sì	No	Visualizza la suddivisione dei costi per modello e area di lavoro. AWS Cost Explorer mostra l'elemento di riga aggregato della CCU.
Limiti	Sì	No	Visualizza i limiti di tariffa (sola lettura).
Spazi di lavoro	Sì	No	Visualizza gli spazi di lavoro per regione (sola lettura).
Files	Sì	Sì	Visualizza e gestisci i file caricati.
Competenze	Sì	Sì	Visualizza e gestisci le competenze degli agenti.
Batch	Sì	Sì	Visualizza e gestisci i processi di elaborazione in batch.
Agents (Agenti)	Sì	Sì	Visualizza e gestisci le definizioni degli agenti.
Sessioni	Sì	Sì	Visualizza le sessioni degli agenti e la cronologi a degli eventi.
Ambienti	Sì	Sì	Visualizza e gestisci le configurazioni dei container cloud per le sessioni.
Archivi di credenziali	Sì	Sì	Visualizza e gestisci gli archivi di credenziali per l'autenticazione delle sessioni.

Pagina	Disponibilità	Tramite gateway AWS	Note
Archivi di memoria	Sì	Sì	Visualizza e gestisci la memoria persistente degli agenti.
Chiavi API	No	N/A	Gestisci le chiavi API nella console AWS (Claude Platform on AWS → chiavi API). Vedi Autenticazione.
Membri	No	N/A	Non applicabile. AWS IAM gestisce l'accesso.
Fatturazione	No	N/A	Non applicabile. AWS Marketplace gestisce la fatturazione e la fatturazione. Visualizza le ripartizioni dei costi nella pagina Costi.
Codice Claude	No	N/A	Visualizza l'utilizzo del codice Claude nella pagina Utilizzo.

Cambio di organizzazione

La console Claude non supporta il cambio di organizzazione per Claude Platform su AWS. Per accedere a un'altra organizzazione, esci ed esegui nuovamente l'autenticazione tramite la console AWS utilizzando il ruolo IAM per l'account AWS di quell'organizzazione.

Limiti tariffari e quote

Claude Platform su AWS assegna limiti di tariffa Tier 1 al momento dell'abbonamento. Anthropic gestisce i limiti di velocità direttamente, non tramite i sistemi di quote AWS.

Limiti predefiniti

La piattaforma Claude su AWS utilizza la pianificazione dei livelli standard di Anthropic, identica all'API Claude di prima parte. I limiti del livello 1 si applicano per ogni spazio di lavoro. I limiti sono raggruppati per famiglia di modelli. Ad esempio, tutti i modelli Opus condividono un limite combinato, tutti i modelli Sonnet ne condividono un altro e tutti i modelli Haiku ne condividono un terzo.

Per gli attuali valori Tier 1 (RPM, ITPM, OTPM) e le soglie di livello superiore, consulta Rate limits sul sito web di documentazione di Anthropic. La pagina Anthropic è la fonte della verità e viene aggiornata quando i limiti cambiano.

Intestazioni relative ai limiti di velocità

Ogni risposta include intestazioni che riportano lo stato attuale del limite di velocità. Intestazioni chiave:

- `anthropic-ratelimit-requests-limit`— Numero massimo di richieste al minuto
- `anthropic-ratelimit-requests-remaining`— Richieste rimanenti nella finestra corrente
- `anthropic-ratelimit-requests-reset`— Ora in cui viene ripristinato il limite di richieste (RFC 3339)
- `anthropic-ratelimit-tokens-limit`— Numero massimo di token combinati (ingresso +uscita) al minuto
- `anthropic-ratelimit-tokens-remaining`— Token combinati rimanenti nella finestra corrente
- `anthropic-ratelimit-tokens-reset`— Ora in cui viene ripristinato il limite combinato dei token (RFC 3339)
- `anthropic-ratelimit-input-tokens-limit`//`-remaining`— intestazioni `-reset` Input-token-specific
- `anthropic-ratelimit-output-tokens-limit`/`-remaining`/`-reset`— Output-token-specific intestazioni

- `retry-after`— In una risposta 429, il numero di secondi di attesa prima di riprovare

Per il set completo, consulta le intestazioni di Response sul sito Web della documentazione di Anthropic.

Richiesta di limiti più elevati

A differenza dell'API Claude di prima parte, l'avanzamento automatico dei livelli non si applica alla piattaforma Claude su AWS. Per richiedere limiti più elevati, contatta il rappresentante dell'account Anthropic con l'ID del tuo spazio di lavoro e la velocità effettiva desiderata. Per le soglie dei livelli e altri dettagli, consulta Rate limits sul sito web della documentazione di Anthropic.

Errori relativi ai limiti di velocità

Quando superi un limite di velocità, l'API restituisce HTTP 429 con un `rate_limit_error` tipo. Implementa il backoff esponenziale con jitter nella logica dei tentativi. L'`retry-after`intestazione indica quanti secondi attendere prima di riprovare.

Fatturazione

Claude Platform su AWS utilizza AWS Marketplace come backend di fatturazione per la misurazione basata sul consumo. I contatori antropici vengono utilizzati ogni ora tramite AWS Marketplace e AWS emette fatture mensili sulla fattura AWS esistente.

La fatturazione avviene solo per arretrati (paghi solo quello che hai usato) e al lordo delle imposte (AWS applica le impostazioni fiscali del tuo account).

L'utilizzo è denominato in Claude Consumption Units (CCU) a 0,01 USD per CCU. Il prezzo della CCU è fisso e mai scontato. Anthropic valuta l'utilizzo dei token in USD alle tariffe standard per modello e per funzionalità, applica qualsiasi sconto negoziato, quindi converte il risultato in CCU a 0,01 USD per CCU. Gli sconti si traducono in un minor numero di CCU misurate, non in un prezzo della CCU inferiore. Le CCU non sono crediti prepagati; non è previsto alcun saldo o impegno per la CCU. Consulta la sezione Prezzi per la definizione della CCU e le tariffe dei token per modello.

Monitoraggio e registrazione dei log

AWS CloudTrail può acquisire tutte le richieste di Claude Platform su AWS. Per impostazione predefinita, le operazioni di workspace e vault vengono registrate come eventi di gestione. Tutte le altre operazioni (inferenza, batch, file, skill, modello, profilo utente e Claude Managed Agents tranne i vault) sono eventi Data. La loro registrazione richiede una configurazione esplicita e comporta costi aggiuntivi. CloudTrail Consulta le azioni IAM per la classificazione completa dei tipi di evento e la CloudTrail documentazione AWS per i dettagli di configurazione.

Quando CloudTrail configuri la registrazione degli eventi relativi ai dati, seleziona il tipo di risorsa. `AWS::AWSExternalAnthropic::Workspace` Questo è il tipo di CloudTrail risorsa per Claude Platform sugli spazi di lavoro AWS ed è necessario per acquisire gli eventi del piano dati (inferenza, batch, file e altre operazioni per area di lavoro). Se desideri registrare l'attività per un sottoinsieme di aree di lavoro anziché per l'intero account, consoni i selettori di eventi di dati a ARN specifici dello spazio di lavoro.

Richiedi ID

Ogni risposta include due ID di richiesta nelle intestazioni di risposta:

- AWS request ID (**`x-amzn-requestid`**): l'ID primario, indicizzato in. CloudTrail Usalo quando esamini le richieste tramite gli strumenti AWS o quando contatti il supporto AWS.
- Anthropic request ID (**`request-id`**): l'ID secondario. Usalo quando contatti l'assistenza Anthropic.

Python

```
from anthropic import AnthropicAWS

client = AnthropicAWS(aws_region="us-west-2")

response = client.messages.with_raw_response.create(
    model="claude-sonnet-4-6",
    max_tokens=1024,
    messages=[{"role": "user", "content": "Hello!"}],
)

print(response.headers.get("x-amzn-requestid"))  # AWS request ID
```

```
print(response.headers.get("request-id"))  # Anthropic request ID

message = response.parse()
print(message.content)
```

TypeScript

```
import AnthropicAws from "@anthropic-ai/aws-sdk";

const client = new AnthropicAws({ awsRegion: "us-west-2" });

const { data: message, response } = await client.messages
  .create({
    model: "claude-sonnet-4-6",
    max_tokens: 1024,
    messages: [{ role: "user", content: "Hello!" }]
  })
  .withResponse();

console.log(response.headers.get("x-amzn-requestid")); // AWS request ID
console.log(response.headers.get("request-id")); // Anthropic request ID
console.log(message.content);
```

C#

```
using Anthropic;
using Anthropic.Aws;

var client = new AnthropicAwsClient(new() { AwsRegion = "us-west-2" });

var response = await client.WithRawResponse.Messages.Create(new()
{
    Model = "claude-sonnet-4-6",
    MaxTokens = 1024,
    Messages = [new() { Role = "user", Content = "Hello!" }]
});

Console.WriteLine(response.Headers.GetValues("x-amzn-requestid").First()); // AWS
 request ID
Console.WriteLine(response.Headers.GetValues("request-id").First()); // Anthropic
 request ID
Console.WriteLine(response.Value.Content);
```

Go

```
client, err := anthropicaws.NewClient(context.Background(),
 anthropicaws.ClientConfig{})
if err != nil {
    panic(err)
}

var response *http.Response
message, err := client.Messages.New(
    context.Background(),
    anthropic.MessageNewParams{
        Model:     anthropic.ModelClaudeSonnet4_6,
        MaxTokens: 1024,
        Messages: []anthropic.MessageParam{
            anthropic.NewUserMessage(anthropic.NewTextBlock("Hello!")),
        },
    },
    option.WithResponseInto(&response),
)
if err != nil {
    panic(err)
}

fmt.Println(response.Header.Get("x-amzn-requestid")) // AWS request ID
fmt.Println(response.Header.Get("request-id"))       // Anthropic request ID
fmt.Println(message.Content[0].Text)
```

Java

```
import com.anthropic.aws.backends.AwsBackend;
import com.anthropic.client.AnthropicClient;
import com.anthropic.client.okhttp.AnthropicOkHttpClient;
import com.anthropic.core.http.HttpResponseFor;
import com.anthropic.models.messages.Message;
import com.anthropic.models.messages.MessageCreateParams;
import com.anthropic.models.messages.Model;

AnthropicClient client = AnthropicOkHttpClient.builder()
    .backend(AwsBackend.fromEnv())
    .build();

HttpResponseFor<Message> response = client.messages().withRawResponse().create(
```

```
    MessageCreateParams.builder()
        .model(Model.CLAUDE_SONNET_4_6)
        .maxTokens(1024)
        .addUserMessage("Hello!")
        .build()
);

IO.println(response.headers().values("x-amzn-requestid")); // AWS request ID
IO.println(response.requestId()); // Anthropic request ID
IO.println(response.parse().content());
```

PHP

```
use Anthropic\Aws\Client;

$client = new Client();

$response = $client->messages->raw->create(
    model: 'claude-sonnet-4-6',
    maxTokens: 1024,
    messages: [['role' => 'user', 'content' => 'Hello!']],
);

echo $response->getHeaderLine('x-amzn-requestid') . "\n"; // AWS request ID
echo $response->getHeaderLine('request-id') . "\n"; // Anthropic request ID
echo $response->parse();
```

Ruby

Attualmente l'SDK di Ruby non espone un accessor a risposta grezza, quindi gli ID delle richieste non possono essere letti dalle intestazioni di risposta in Ruby. Se hai bisogno della registrazione dell'ID della richiesta, usa uno degli altri linguaggi sopra indicati o acquisisci gli ID a livello di proxy HTTP.

Anthropic consiglia di registrare la propria attività su base continuativa per almeno 30 giorni per comprendere i modelli di utilizzo e indagare su eventuali problemi.

> **Note**
>
> AWS CloudTrail è configurato all'interno del tuo account AWS. L'abilitazione della registrazione non fornisce ad AWS o Anthropic l'accesso ai tuoi contenuti oltre a quanto necessario per la fatturazione e il funzionamento del servizio.

Metriche e dashboard di utilizzo

Claude Platform su AWS non pubblica i parametri di utilizzo per area di lavoro su Amazon. CloudWatch Il numero di token, il volume delle richieste e l'utilizzo per modello sono invece disponibili tramite le superfici di utilizzo di Anthropic:

- Visualizzazioni sull'utilizzo della Claude Console: quando un principale si unisce alla Claude Console con `aws-external-anthropic:AssumeConsole` (vedi le politiche IAM), le dashboard di utilizzo mostrano il volume delle richieste, il conteggio dei token e le suddivisioni per modello per aree di lavoro accessibili. Account-wide le visualizzazioni di utilizzo richiedono la funzionalità della console di amministrazione.
- API Anthropic Usage and Cost: per l'accesso programmatico ai dati di utilizzo e costo, inclusa l'aggregazione per area di lavoro e per modello, consulta l'API Usage and Cost nella documentazione di Anthropic.

Per il monitoraggio operativo (tassi di errore, latenza, intestazioni dei limiti di velocità), utilizza le intestazioni di risposta restituite su ogni richiesta (vedi Limiti e quote di velocità) insieme agli ID di richiesta AWS acquisiti in. CloudTrail

Allocazione e monitoraggio dei costi

I costi di Claude Platform on AWS sono indicati nella fattura AWS sotto il servizio Claude Platform on AWS, con dimensioni di utilizzo per modello. Utilizza l'AWS Cost and Usage Report (CUR) combinato con i tag dell'area di lavoro per un'allocazione dettagliata dei costi:

1. Etichetta i tuoi spazi di lavoro con le dimensioni di allocazione che ti interessano (team, applicazione, ambiente, centro di costo). Vedi Workspaces per i dettagli sui tag.
2. Nella console di fatturazione AWS, attiva ogni chiave tag come tag di allocazione dei costi. Dopo l'attivazione, l'utilizzo effettuato alla o dopo la data di attivazione viene suddiviso per tag nel CUR.
3. In CUR o AWS Cost Explorer, raggruppa per `resourceTags/user:<tag-key>` colonna per suddividere la spesa per tag dell'area di lavoro.

I tag Workspace passano agli elementi di riga CUR per tutto l'utilizzo di Claude Platform su AWS. Le stesse chiavi di tag determinano sia il controllo degli IAM-based accessi che l'allocazione dei costi. Consulta i tag di allocazione dei costi nella documentazione di AWS Billing per le fasi di configurazione e i ritardi di attivazione.

Migrazione da Amazon Bedrock

Se attualmente utilizzi Claude su Bedrock, la migrazione a Claude Platform su AWS richiede modifiche durante l'integrazione. La firma SigV4 rimane supportata, ma cambiano il contesto di firma, l'URL di base, il formato API, gli ID dei modelli, il client e il pacchetto SDK, il formato di streaming, le intestazioni delle richieste e la disponibilità della regione. La tabella seguente riassume le differenze.

Cosa cambia

	Amazon Bedrock	Piattaforma Claude su AWS
URL di base	`bedrock-runtime.{region}.amazonaws.com`	`aws-external-anthropic.{region}.api.aws`
Formato API	Converse Bedrock/ InvokeModel	API per messaggi antropici () `/v1/messages`
ID dei modelli	`anthropic.claude-opus-4-7-v1`	`claude-opus-4-7`
Client SDK	`AnthropicBedrock` /Bedrock SDK	Platform-specific client (`AnthropicAWS` ,`AnthropicAws` , ecc.)`AnthropicAwsClient` , in versione beta
pacchetto SDK	`anthropic[bedrock]` `@anthropic-ai/bedrock-sdk`, ecc.	`anthropic[aws]` `@anthropic-ai/aws-sdk` , ecc. (vedi Installare un SDK)
nome del servizio SIGv4	`bedrock`	`aws-external-anthropic`
Formato di streaming	AWS EventStream	SSE (come Claude API)

	Amazon Bedrock	Piattaforma Claude su AWS
Intestazione dell'area di lavoro	Non applicabile	`anthropic-workspace-id` obbligatorio
Disponibilità nelle Regioni	Diverse aree commerciali	Tutte le regioni commerciali AWS (le regioni con opt-in richiedono l'attivazione dell'account)

Cosa si guadagna

- Accesso in genere lo stesso giorno a nuovi modelli e funzionalità, senza una fase separata di integrazione con i partner
- Agent Skills per la generazione di documenti (ExcelPowerPoint, Word, PDF)
- Esecuzione del codice nella sandbox gestita di Anthropic
- Funzionalità beta tramite l'`anthropic-beta`intestazione (vedi Funzionalità non attualmente disponibili)
- Claude Console per la visibilità delle quote e l'analisi dell'utilizzo
- Supporto diretto di Anthropic
- Autenticazione con chiave API come alternativa a SigV4 (vedi Autenticazione)

Cosa rimane lo stesso

- Autenticazione AWS IAM (SigV4)
- Fatturazione tramite il tuo account AWS
- AWS Commitment Retirement

Le insidie della migrazione

> **Warning**
>
> Abilita innanzitutto la federazione delle identità web in uscita. Se il tuo account AWS non ha mai utilizzato Claude Platform on AWS, devi abilitare la federazione delle identità web in uscita una volta per account prima di effettuare richieste. Senza questo passaggio, tutte le richieste hanno esito negativo e si verifica un errore di federazione. Questo passaggio non è richiesto per Bedrock.

> **Warning**
>
> Zero Data Retention (ZDR) è un'opzione opzionale su Claude Platform su AWS. Su Bedrock, AWS è l'elaboratore di dati e Anthropic non conserva input o output di inferenza; il programma ZDR di Anthropic non si applica a questo caso. Su Claude Platform su AWS, Anthropic è l'elaboratore dei dati e ZDR segue il modello di API Claude di prima parte: è disponibile su richiesta tramite il rappresentante dell'account Anthropic. Conferma l'iscrizione a ZDR prima di migrare i carichi di lavoro di produzione che dipendono dalle garanzie di conservazione dei dati.

Considerazioni commerciali

- Termini di servizio: i termini di servizio AWS regolano Claude Platform su AWS. Si applicano anche i termini di servizio commerciali, l'addendum sull'elaborazione dei dati e la politica di utilizzo di Anthropic.
- Sconti e offerte private: gli sconti negoziati e le offerte private di AWS Marketplace non vengono trasferiti automaticamente tra Bedrock e Claude Platform su AWS. Collabora con il rappresentante del tuo account Anthropic per impostare le condizioni commerciali per Claude Platform su AWS.

Policy IAM

La piattaforma Claude su AWS si integra con AWS IAM per il controllo degli accessi. Puoi concedere o negare l'accesso a specifiche azioni API su aree di lavoro specifiche utilizzando la sintassi standard delle policy IAM.

Il nome del servizio SigV4 e lo spazio dei nomi delle azioni IAM sono. `aws-external-anthropic` Le azioni seguono lo schema `aws-external-anthropic:<Action>` (ad esempio,). `aws-external-anthropic:CreateInference`

Esempio: negare l'inferenza in batch

La seguente policy consente l'inferenza in tempo reale bloccando l'elaborazione in batch, un requisito comune per i carichi di lavoro: ZDR-sensitive

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "aws-external-anthropic:CreateInference",
        "aws-external-anthropic:CountTokens",
        "aws-external-anthropic:GetModel",
        "aws-external-anthropic:ListModels",
        "aws-external-anthropic:GetWorkspace",
        "aws-external-anthropic:ListWorkspaces"
      ],
      "Resource": "arn:aws:aws-external-anthropic:*:*:workspace/*"
    },
    {
      "Effect": "Deny",
      "Action": [
        "aws-external-anthropic:CreateBatchInference",
        "aws-external-anthropic:GetBatchInference",
        "aws-external-anthropic:ListBatchInferences"
      ],
      "Resource": "*"
    }
  ]
```

```
}
```

L'`GetBatchInference`azione autorizza sia il percorso dei metadati del batch che il percorso dei risultati del batch. Negarlo, inoltre`ListBatchInferences`, blocca sia le letture che l'enumerazione in batch.

L'`Allow`istruzione enumera azioni e specifiche anziché utilizzare caratteri jolly. `Get*` `List*` I caratteri jolly consentirebbero `GetFile` (ossia il download di byte di file) e altre letture non desiderate; hanno comunque la precedenza, ma la forma `Deny` esplicita è il modello più `Allow` sicuro da modellare.

Esempio: inferenza sincrona su un singolo spazio di lavoro

Concede le autorizzazioni minime per un principale IAM che esegue l'inferenza su un workspace di produzione:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "aws-external-anthropic:CreateInference",
        "aws-external-anthropic:CountTokens",
        "aws-external-anthropic:Get*",
        "aws-external-anthropic:List*"
      ],
      "Resource": "arn:aws:aws-external-anthropic:us-west-2:123456789012:workspace/
wrkspc_01AbCdEf23GhIj"
    }
  ]
}
```

> **Note**
>
> Corrisponde `ListWorkspaces` anche la `List*` jolly contenuta in questa policy, che si riferisce all'account. Il vincolo ARN dell'area di lavoro lo filtra automaticamente, quindi questo criterio non autorizza l'elenco degli spazi di lavoro. Se il tuo account di servizio deve enumerare le aree di lavoro, aggiungi un'istruzione separata per with. `Allow` `ListWorkspaces` `Resource: "*"`

Questa policy presuppone l'autenticazione AWS SigV4. Se il principale si autentica con una chiave API, `aws-external-anthropic:CallWithBearerToken` concedi anche (vedi Autenticazione).

Esempio: isolamento dell'area di lavoro per cliente

Limita un ruolo a un singolo spazio di lavoro:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "aws-external-anthropic:*",
      "Resource": "arn:aws:aws-external-anthropic:us-west-2:123456789012:workspace/
wrkspc_01AbCdEf23GhIj"
    },
    {
      "Effect": "Allow",
      "Action": [
        "aws-external-anthropic:CallWithBearerToken",
        "aws-external-anthropic:AssumeConsole"
      ],
      "Resource": "*"
    }
  ]
}
```

Note

Il carattere `aws-external-anthropic:*` jolly nella prima istruzione include azioni con ambito account (`CreateWorkspace`,`ListWorkspaces`) che il vincolo ARN dell'area di lavoro filtra silenziosamente. Ciò è coerente con l'intento di «isolamento»: il ruolo non può creare o enumerare aree di lavoro, ma la politica contiene autorizzazioni che non hanno alcun effetto. Vedi Provisioning automation per il modello con ambito di account.
La seconda istruzione concede `CallWithBearerToken` e `AssumeConsole` su tutte le risorse perché entrambe sono azioni senza percorso che non si legano a un ARN dell'area di

lavoro. Ometti la seconda istruzione se il ruolo utilizza solo SigV4 e non viene mai federato alla Claude Console.

Esempio: blocco delle funzionalità per uno spazio di lavoro ZDR-sensitive

Blocca l'elaborazione in batch e il caricamento di file su un'area di lavoro specifica lasciando disponibile l'inferenza sincrona. Utile quando un'area di lavoro gestisce dati Zero Data Retention (ZDR) che non devono persistere sul lato server. Allega questo criterio a un criterio Allow, come `AnthropicLimitedAccess` nell'esempio dello spazio di lavoro singolo riportato sopra; di per sé, un criterio non concede autorizzazioni: Deny-only

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "aws-external-anthropic:CreateBatchInference",
        "aws-external-anthropic:CreateFile"
      ],
      "Resource": "arn:aws:aws-external-anthropic:us-west-2:123456789012:workspace/
wrkspc_01AbCdEf23GhIj"
    }
  ]
}
```

> **Note**
>
> Questa negazione blocca solo la creazione. Le altre azioni relative ai file e ai batch non vengono negate a meno che non vengano elencate anche queste. Per un blocco completo in cui l'area di lavoro non deve mai contenere file o batch, utilizza anche denyaws-external-anthropic:GetFile,,,aws-external-anthropic:ListFiles, aws-external-anthropic:DeleteFileaws-external-anthropic:GetBatchInference, aws-external-anthropic:ListBatchInferences e. aws-external-anthropic:CancelBatchInference aws-external-anthropic:DeleteBatchInference

Esempio: automazione del provisioning

Assegna a un CI/CD ruolo le azioni necessarie per creare e gestire gli spazi di lavoro, senza alcuna autorizzazione di inferenza:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "aws-external-anthropic:CreateWorkspace",
        "aws-external-anthropic:GetWorkspace",
        "aws-external-anthropic:ListWorkspaces",
        "aws-external-anthropic:UpdateWorkspace",
        "aws-external-anthropic:ArchiveWorkspace"
      ],
      "Resource": "*"
    }
  ]
}
```

`CreateWorkspace`e `ListWorkspaces` sono operazioni relative all'account. Specificare un ARN dell'area di lavoro su queste azioni non ha alcun effetto; utilizzare. `Resource: "*"`

Policy gestite

AWS fornisce policy gestite per modelli di accesso comuni:

- **`AnthropicFullAccess`**: Sovvenzioni `aws-external-anthropic:*` per tutte le risorse.
- **`AnthropicReadOnlyAccess`**: Sovvenzioni `Get*` e `CallWithBearerToken` su tutte le risorse. `List*`
- **`AnthropicInferenceAccess`**: concede le ReadOnly azioni più le azioni di inferenza (`CreateInference`,,, `CreateBatchInference` `CancelBatchInferenceDeleteBatchInference`,`CountTokens`) su tutte le risorse.
- **`AnthropicLimitedAccess`**: Concede `AnthropicInferenceAccess` le azioni più tutte le azioni di Claude Managed Agents (agenti, sessioni, ambienti, vault, archivi di memoria) su tutte le risorse.

- **`AnthropicSelfHostedEnvironmentAccess`**: concede le autorizzazioni necessarie a un operatore di sandbox ospitato autonomamente per interrogare ed elaborare gli elementi di lavoro dell'ambiente, leggere l'ambiente, la sessione e le risorse relative alle competenze associate e aggiornare lo stato della sessione. Associa questa policy al ruolo IAM che assume il lavoratore dell'ambiente ospitato autonomamente. `AnthropicSelfHostedEnvironmentAccess`è l'impostazione predefinita consigliata per un solo ruolo; se esegui il work poller e la sandbox per sessione con principi IAM separati, puoi suddividere queste autorizzazioni in due policy personalizzate più ristrette per ottenere i privilegi minimi.

`AnthropicInferenceAccess`è la politica gestita più ristretta sufficiente per eseguire l'inferenza. Tramite `Get*` and `List*` wildcard, garantisce l'accesso in lettura a tutte le risorse API del namespace, incluso lo scaricamento del contenuto dei file e il contenuto della memoria. `GetFile` `GetMemoryStore` Non consente la creazione o l'eliminazione di file o competenze, la gestione dei profili utente, la modifica dell'area di lavoro o la federazione delle console.

> **Note**
>
> `AnthropicReadOnlyAccessAnthropicInferenceAccess`, e non `AnthropicLimitedAccess` concedono. `AssumeConsole` I mandanti che devono effettuare la federazione alla Claude Console richiedono una concessione separata per`aws-external-anthropic:AssumeConsole`, tramite `AnthropicFullAccess` o tramite una politica personalizzata. Vedi Federazione alla Claude Console.

> **Note**
>
> `CreateInferencee` `CreateBatchInference` sono azioni separate. Negare l'una non blocca l'altra. Se intendi impedire tutte le chiamate al modello, nega entrambe.

Connessione alla Claude Console

`aws-external-anthropic:AssumeConsole`consente a un preside IAM di federarsi nella Claude Console Anthropic-operated . L'accesso all'interno della console è ancora regolato da IAM per la maggior parte delle operazioni, ma un sottoinsieme di operazioni di amministrazione, principalmente viste di utilizzo che non hanno un'API IAM corrispondente, si basano sulla funzionalità con cui il principale si è federato.

Esistono due funzionalità:

- **`developer`**— consente le operazioni di cui uno sviluppatore Claude Platform on AWS ha bisogno per il lavoro quotidiano: esecuzione di inferenze dalla console, lettura dei dati dell'area di lavoro, visualizzazione dell'utilizzo personale.
- **`admin`**— consente inoltre le operazioni della console riservate agli amministratori, comprese le visualizzazioni di utilizzo a livello di account e le impostazioni amministrative che non vengono visualizzate tramite le azioni IAM.

Controlla quale funzionalità può richiedere un principale con il tasto condition sull'azione: `aws-external-anthropic:Capability` `AssumeConsole`

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "aws-external-anthropic:AssumeConsole",
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "aws-external-anthropic:Capability": "admin"
        }
      }
    }
  ]
}
```

`AnthropicFullAccess`sovvenzioni `AssumeConsole` senza restrizioni di capacità. Per qualsiasi concessione più limitata, ad esempio accesso alla console riservato agli sviluppatori o accesso riservato agli amministratori, allega una policy personalizzata che soddisfi la condizione illustrata sopra. `AssumeConsole` `aws-external-anthropic:Capability`

Chiamate con token al portatore

`aws-external-anthropic:CallWithBearerToken`autorizza il percorso di SigV4-free richiesta utilizzato quando una chiave API viene presentata come token portatore. Qualsiasi principale che si autentica con una chiave API necessita di questa azione nell'area di lavoro di destinazione. Ciò

vale sia che si utilizzi `ANTHROPIC_AWS_API_KEY` o si imposti direttamente l'`Authorization: Bearer`intestazione.

Ciò è necessario per i chiamanti delle chiavi API oltre alle azioni di inferenza (`CreateInferenceCreateBatchInference`, ecc.). `CallWithBearerToken`In caso contrario, le richieste di chiavi API vengono rifiutate prima di raggiungere il controllo di autorizzazione all'inferenza. I chiamanti SigV4 non hanno bisogno di questa azione.

`AnthropicReadOnlyAccess`, `AnthropicInferenceAccessAnthropicLimitedAccess`, e tutto include. `AnthropicFullAccess` `CallWithBearerToken` Se scrivi una policy personalizzata per l'accesso alle chiavi API, aggiungila esplicitamente.

Federazione delle identità Web in uscita (richiesta per l'accesso alla console)

La console Claude funziona nell'infrastruttura Anthropic, non in AWS. Quando un principale IAM chiama`AssumeConsole`, AWS STS emette un token di identità web destinato al pubblico di Anthropic; la console Anthropic accetta quindi quel token e stabilisce la sessione federata. Affinché ciò funzioni, l'account AWS deve consentire la federazione delle identità Web in uscita al `aws-external-anthropic` pubblico.

I principali che effettuano la chiamata `AssumeConsole` necessitano delle seguenti autorizzazioni STS oltre all'azione: `aws-external-anthropic:AssumeConsole`

- **`sts:GetWebIdentityToken`**— consente l'emissione del token di identità web utilizzato dalla console Anthropic.
- **`sts:TagGetWebIdentityToken`**— consente di allegare tag di sessione al token di identità web. Claude Platform su AWS utilizza questi tag per trasmettere alla console le funzionalità e il contesto dell'area di lavoro del principale.

Includili sia nella policy di fiducia di qualsiasi ruolo assunto dagli utenti della console, sia nella inline/managed policy allegata alle identità degli utenti che effettuano chiamate dirette. `AssumeConsole` `AnthropicFullAccess`include entrambe le azioni STS. Gli ambienti che negano `sts:*` a livello di SCP o di limite di autorizzazione devono consentire esplicitamente queste due azioni affinché la federazione delle console abbia successo.

Azioni IAM per Claude Platform su AWS

Riferimento all'azione IAM per controllare l'accesso alla piattaforma Claude su AWS tramite le policy AWS.

La piattaforma Claude su AWS utilizza AWS IAM per il controllo degli accessi. Ogni instradamento dell'API è mappato a un'azione IAM nel `aws-external-anthropic` namespace. Questa pagina elenca tutte le azioni, i percorsi autorizzati da ciascuna azione e le politiche gestite disponibili per i modelli di accesso comuni. Per la configurazione e l'autenticazione della piattaforma, consulta What is Claude Platform on AWS? .

Dettagli del servizio

prefisso del servizio IAM	`aws-external-anthropic`
Tipi di risorse	`workspace`
ARN dell'area di lavoro	`arn:aws:aws-external-anthropic:{region}:{account-id}:workspace/{workspace-id}`

La regione ARN è sempre popolata e corrisponde alla regione a cui è associato lo spazio di lavoro. Il segmento di risorse è l'ID dello spazio di lavoro con tag (`wrkspc_…`), lo stesso valore che passi nell'intestazione. `anthropic-workspace-id`

Azioni

Il servizio definisce 65 azioni in 15 gruppi. Le azioni seguono la `VerbNoun` convenzione AWS e utilizzano la disciplina verbale in modo che `Get*` i caratteri `List*` jolly producano un confine pulito di sola lettura.

Inferenza

Azione	Percorsi autorizzati
`CreateInference`	`POST /v1/messages`

Azione	Percorsi autorizzati
CountTokens	POST /v1/messages/count_tokens

Elaborazione in batch

Azione	Percorsi autorizzati
CreateBatchInference	POST /v1/messages/batches
GetBatchInference	GET /v1/messages/batches/{id} GET /v1/messages/batches/{id}/results
ListBatchInferences	GET /v1/messages/batches
CancelBatchInference	POST /v1/messages/batches/{id}/cancel
DeleteBatchInference	DELETE /v1/messages/batches/{id}

Modelli

Azione	Percorsi autorizzati
GetModel	GET /v1/models/{id}
ListModels	GET /v1/models

File

Azione	Percorsi autorizzati
CreateFile	POST /v1/files
GetFile	GET /v1/files/{id} GET /v1/files/{id}/content
ListFiles	GET /v1/files

Azione	Percorsi autorizzati
`DeleteFile`	`DELETE /v1/files/{id}`

> **Note**
>
> `GetFile`autorizza sia il download di metadati che di contenuti. Un principale con accesso in sola lettura può scaricare byte di file, non solo file di elenco.

Competenze

Azione	Percorsi autorizzati
`CreateSkill`	`POST /v1/skills`
`GetSkill`	`GET /v1/skills/{id}` `GET /v1/skills/{id}/versions` `GET /v1/skills/{id}/versions/{version}` `GET /v1/skills/{id}/versions/{version}/content`
`ListSkills`	`GET /v1/skills`
`UpdateSkill`	`POST /v1/skills/{id}/versions` `DELETE /v1/skills/{id}/versions/{version}`
`DeleteSkill`	`DELETE /v1/skills/{id}`

> **Note**
>
> L'eliminazione di una versione di una singola abilità corrisponde a`UpdateSkill`, no`DeleteSkill`. Una politica che nega consente `aws-external-anthropic:Delete*` comunque l'eliminazione delle versioni. Nega `UpdateSkill` e `CreateSkill` anche se hai bisogno di prevenire qualsiasi mutazione dell'abilità.

Profili utente

Azione	Percorsi autorizzati
`CreateUserProfile`	`POST /v1/user_profiles`
`GetUserProfile`	`GET /v1/user_profiles/{id}`
`ListUserProfiles`	`GET /v1/user_profiles`
`UpdateUserProfile`	`POST /v1/user_profiles/{id}`

> **Warning**
>
> L'action matching IAM non fa distinzione tra maiuscole e minuscole. Il carattere jolly `aws-external-anthropic:*File` corrisponde a `CreateFile`, `GetFile`, e`DeleteFile`, ma non corrisponde `ListFiles` (il che termina con «files», non «file»). Inoltre, è superiore alla media `CreateUserProfile`, `GetUserProfile`, e `UpdateUserProfile` perché «Profilo» termina con «file». Se intendi concedere o negare solo le azioni dell'API Files, enumerale esplicitamente (`,CreateFile`, `GetFile`, `ListFiles`,`DeleteFile`) anziché utilizzare uno schema di suffissi. `*File`

Agents (Agenti)

Definizioni degli agenti per Claude Managed Agents.

Azione	Percorsi autorizzati
`CreateAgent`	L'agente crea percorsi
`GetAgent`	L'agente legge i percorsi
`ListAgents`	Elenco dei percorsi degli agenti
`UpdateAgent`	Percorsi di aggiornamento degli agenti
`ArchiveAgent`	Percorsi di archiviazione degli agenti

Sessioni

Sessioni degli agenti e cronologia degli eventi.

Azione	Percorsi autorizzati
`CreateSession`	Sessione: crea percorsi
`GetSession`	Percorsi di lettura della sessione
`ListSessions`	Percorsi dell'elenco delle sessioni
`UpdateSession`	Percorsi di aggiornamento della sessione
`ArchiveSession`	Percorsi di archiviazione delle sessioni
`DeleteSession`	Percorsi di eliminazione della sessione

Ambienti

Configurazioni di contenitori cloud per le sessioni.

Azione	Percorsi autorizzati
`CreateEnvironment`	Ambiente: crea percorsi
`GetEnvironment`	Ambiente: leggi i percorsi
`ListEnvironments`	Elenco dei percorsi ambientali
`UpdateEnvironment`	Percorsi di aggiornamento dell'ambiente
`ArchiveEnvironment`	Percorsi di archiviazione ambientale
`DeleteEnvironment`	Ambiente: elimina percorsi
`ProcessEnvironmentWork`	Self-hosted percorsi degli addetti all'ambiente

Vault

archivi di credenziali per l'autenticazione della sessione.

Azione	Percorsi autorizzati
`CreateVault`	Vault crea percorsi
`GetVault`	Vault leggi i percorsi
`ListVaults`	Elenca i percorsi del Vault
`UpdateVault`	Percorsi di aggiornamento del Vault
`ArchiveVault`	Percorsi di archiviazione del Vault
`DeleteVault`	Vault elimina i percorsi

> **Note**
>
> Le operazioni del vault sono classificate come eventi di CloudTrail gestione (anziché eventi di dati) perché gli archivi contengono le credenziali e traggono vantaggio dalla registrazione di controllo predefinita. Altre operazioni di Claude Managed Agents (agenti, sessioni, ambienti, archivi di memoria) sono eventi Data.

Archivi di memoria

Memoria persistente dell'agente.

Azione	Percorsi autorizzati
`CreateMemoryStore`	Memory Store crea percorsi
`GetMemoryStore`	Memorizza percorsi di lettura
`ListMemoryStores`	Elenca i percorsi dell'archivio di memoria
`UpdateMemoryStore`	Percorsi di aggiornamento dell'archivio di memoria

Azione	Percorsi autorizzati
`ArchiveMemoryStore`	Percorsi di archiviazione dell'archivio di memoria
`DeleteMemoryStore`	Percorsi di eliminazione dell'archivio di memoria

> **Note**
>
> `GetMemoryStore`legge il contenuto della memoria. Il carattere `Get*` jolly in una politica gestita o personalizzata garantisce quindi l'accesso alla lettura della memoria. Ambita esplicitamente le politiche se i contenuti della memoria devono essere limitati.

Webhook

Notifiche degli eventi del ciclo di vita per le sessioni.

Azione	Percorsi autorizzati
`ListWebhooks`	`GET /v1/webhooks`
`GetWebhook`	`GET /v1/webhooks/{webhook_endpoint_id}`
`CreateWebhook`	`POST /v1/webhooks`
`UpdateWebhook`	`POST /v1/webhooks/{webhook_endpoint_id}`
`DeleteWebhook`	`DELETE /v1/webhooks/{webhook_endpoint_id}`
`RotateWebhookSecret`	`POST /v1/webhooks/{webhook_endpoint_id}/regenerate_signing_secret`

WorkSpace

Azione	Percorsi autorizzati
`CreateWorkspace`	`POST /v1/organizations/workspaces`

Azione	Percorsi autorizzati
`GetWorkspace`	`GET /v1/organizations/workspaces/{id}`
`ListWorkspaces`	`GET /v1/organizations/workspaces`
`UpdateWorkspace`	`POST /v1/organizations/workspaces/{id}`
`ArchiveWorkspace`	`POST /v1/organizations/workspaces/{id}/archive`

Al momento dell'iscrizione viene fornito uno spazio di lavoro predefinito; vedi Workspaces.

Autenticazione

Azione	Percorsi autorizzati
`CallWithBearerToken`	(nessuno)

`CallWithBearerToken`è un'autorizzazione di livello di autenticazione che autorizza un principale ad autenticarsi tramite una chiave API (token bearer) anziché AWS SigV4. Non esegue il mapping su un percorso. Concedilo insieme alle azioni mappate sul percorso che desideri che il titolare della chiave API esegua.

Accesso tramite la console

Azione	Percorsi autorizzati
`AssumeConsole`	(nessuno)

`AssumeConsole`autorizza un principale ad aprire la console Claude per una piattaforma Claude nello spazio di lavoro AWS tramite il flusso di federazione della console AWS. Non esegue la mappatura su un percorso. Concedilo ai responsabili che dovrebbero poter fare clic su Open Claude Console nella pagina del servizio Claude Platform on AWS nella Console AWS. Il tasto `aws-external-anthropic:Capability` condition sull'`AssumeConsole`azione controlla quale

funzionalità della console (sviluppatore o amministratore) può richiedere un principale. Consulta le politiche IAM per i dettagli e Utilizzo della console Claude per il flusso di accesso.

> **Warning**
>
> `aws-external-anthropic:AssumeConsole`emette una sessione di Claude Console che dura fino a 12 ore, indipendentemente dalla durata della sessione del chiamante. Un chiamante la cui sessione IAM è inferiore a 12 ore può comunque ottenere una sessione di console che dura più a lungo delle proprie credenziali di origine. Limita questa autorizzazione ai responsabili che richiedono l'accesso alla console Claude e revocala prontamente quando non è più necessaria.

> **Note**
>
> Le azioni eseguite all'interno della console Claude dopo la federazione non sono registrate in AWS CloudTrail o attribuibili tramite IAM. Ciò include attività globali relative all'area di lavoro, come la visualizzazione dei report di utilizzo. Se hai bisogno di un audit trail per l'attività della console, utilizza invece i log di controllo disponibili nella console Claude. CloudTrail

Route-to-action mappatura

La tabella seguente elenca ogni route su Claude Platform su AWS e l'azione IAM richiesta per chiamarla. L'azione IAM della route stabile autorizza anche le richieste che utilizzano l'`anthropic-beta`intestazione. CloudTrail classifica ogni azione come evento Data (operazioni sul piano dati ad alto volume) o come evento di gestione (operazioni sul piano di controllo).

Metodo	Route	Azione IAM	CloudTrail tipo di evento
`POST`	`/v1/messages`	`CreateInference`	Dati
`POST`	`/v1/messages/count_tokens`	`CountTokens`	Dati
`POST`	`/v1/messages/batches`	`CreateBatchInference`	Dati

Metodo	Route	Azione IAM	CloudTrail tipo di evento
GET	/v1/messages/batches	ListBatchInferences	Dati
GET	/v1/messages/batches/{id}	GetBatchInference	Dati
GET	/v1/messages/batches/{id}/results	GetBatchInference	Dati
POST	/v1/messages/batches/{id}/cancel	CancelBatchInference	Dati
DELETE	/v1/messages/batches/{id}	DeleteBatchInference	Dati
GET	/v1/models	ListModels	Dati
GET	/v1/models/{id}	GetModel	Dati
POST	/v1/files	CreateFile	Dati
GET	/v1/files	ListFiles	Dati
GET	/v1/files/{id}	GetFile	Dati
GET	/v1/files/{id}/content	GetFile	Dati
DELETE	/v1/files/{id}	DeleteFile	Dati
POST	/v1/skills	CreateSkill	Dati
GET	/v1/skills	ListSkills	Dati
GET	/v1/skills/{id}	GetSkill	Dati
DELETE	/v1/skills/{id}	DeleteSkill	Dati
POST	/v1/skills/{id}/versions	UpdateSkill	Dati

Metodo	Route	Azione IAM	CloudTrail tipo di evento
GET	/v1/skills/{id}/versions	GetSkill	Dati
GET	/v1/skills/{id}/versions/{version}	GetSkill	Dati
GET	/v1/skills/{id}/versions/{version}/content	GetSkill	Dati
DELETE	/v1/skills/{id}/versions/{version}	UpdateSkill	Dati
POST	/v1/user_profiles	CreateUserProfile	Dati
GET	/v1/user_profiles	ListUserProfiles	Dati
GET	/v1/user_profiles/{id}	GetUserProfile	Dati
POST	/v1/user_profiles/{id}	UpdateUserProfile	Dati
POST	/v1/organizations/workspaces	CreateWorkspace	Gestione
GET	/v1/organizations/workspaces	ListWorkspaces	Gestione
GET	/v1/organizations/workspaces/{id}	GetWorkspace	Gestione
POST	/v1/organizations/workspaces/{id}	UpdateWorkspace	Gestione
POST	/v1/organizations/workspaces/{id}/archive	ArchiveWorkspace	Gestione

I percorsi di Claude Managed Agents (agenti, sessioni, ambienti, vault, archivi di memoria) seguono lo stesso schema di route-to-action; per la mappatura completa per percorso, consulta il riferimento alle azioni di Anthropic IAM. Le route Vault sono eventi di gestione; le route agent, session, environment e memory store sono eventi Data.

Le rotte non su Claude Platform su AWS vengono negate al gateway per impostazione predefinita.

Consulta anche

- Politiche IAM, ad esempio politiche e politiche gestite
- Guida per l'utente di AWS IAM per la sintassi e la logica di valutazione delle policy IAM
- AWS CloudTrail User Guide per la configurazione dei log di audit

Cronologia dei documenti

La tabella seguente descrive importanti modifiche alla Claude Platform on AWS User Guide.

Modifica	Descrizione	Data
Pubblicazione iniziale	Ha pubblicato la Claude Platform on AWS User Guide.	7 maggio 2026

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.