Benutzer-Leitfaden

Claude-Plattform auf AWS

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Claude-Plattform auf AWS: Benutzer-Leitfaden

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Die Marken und Handelsmarken von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, die geeignet ist, Kunden irrezuführen oder Amazon in irgendeiner Weise herabzusetzen oder zu diskreditieren. Alle anderen Marken, die nicht im Besitz von Amazon sind, gehören den jeweiligen Besitzern, die möglicherweise mit Amazon verbunden sind oder von Amazon gesponsert werden.

Table of Contents

Was ist Claude Platform auf AWS? 1
Wie funktioniert die Plattformintegration 1
Funktionen der Claude-Plattform auf AWS 2
Wie ist dieses Handbuch aufgebaut 2
Claude Platform auf AWS im Vergleich zu Amazon Bedrock 3
Einrichten Ihres Kontos 5
Schritt 1: Melden Sie sich in der AWS-Konsole an 5
Schritt 2: Richten Sie Ihre Anthropic-Organisation ein 6
Schritt 3: Notieren Sie sich Ihre Workspace-ID 6
Schritt 4: Melden Sie sich bei der Claude Console an 7
Problembehebung bei der Kontoeinrichtung 7
Voraussetzungen 8
Aktiviere den Identitätsverbund für ausgehende Webidentitäten 8
Besorgen Sie sich Ihre Workspace-ID 8
Authentifizierung 10
SigV4-Authentifizierung 10
API-Schlüsselauthentifizierung 11
Short-term API-Schlüssel 11
Vorrang der Anmeldeinformationen und Auflösung der Region 12
Arbeitsbereich-ID 12
Installieren eines -SDK 13
Verfügbare Modelle 15
Modell-IDs 15
Senden von Anforderungen 16
Verwenden Sie den Anthropic-Basisclient 19
Feature-Unterstützung 21
Von Claude verwaltete Agenten 21
Compliance 22
Wie wird die Workspace-Mitgliedschaft modelliert 22
Funktionen, die derzeit nicht verfügbar sind 23
Datenresidenz 24
Arbeitsbereiche 28
Geltungsbereich des Arbeitsbereichs 28
Arbeitsbereiche erstellen 28

Einstellung der Workspace-ID in Ihrem Client ... 29
Markieren von Workspaces ... 29
Tag-based Zugriffskontrolle ... 29
Verwendung der Claude-Konsole ... 31
Anmelden ... 31
Verfügbare Seiten ... 31
Organisation wechseln ... 33
Ratenbegrenzungen und Kontingente ... 34
Standardlimits ... 34
Überschriften zur Ratenbegrenzung ... 34
Höhere Grenzwerte anfordern ... 35
Fehler bei der Ratenbegrenzung ... 35
Fakturierung ... 36
Überwachung und Protokollierung ... 37
IDs anfordern ... 37
Nutzungsmetriken und Dashboards ... 41
Kostenzuweisung und Überwachung ... 41
Migration von Amazon Bedrock ... 43
Was ändert sich ... 43
Was gewinnen Sie ... 44
Was bleibt gleich ... 44
Fallstricke bei der Migration ... 44
Kommerzielle Überlegungen ... 45
IAM-Richtlinien ... 46
Beispiel: Batch-Inferenz verweigern ... 46
Beispiel: synchrone Inferenz für einen einzelnen Workspace ... 47
Beispiel: Workspace-Isolierung pro Kunde ... 48
Beispiel: Sperrung von Funktionen für einen Workspace ZDR-sensitive ... 49
Beispiel: Automatisierung der Bereitstellung ... 50
Verwaltete Richtlinien ... 50
Verbindung zur Claude Console herstellen ... 52
Anrufe mit Inhaber-Tokens ... 53
Verbund ausgehender Web-Identitätsverbund (für Konsolenzugriff erforderlich) ... 53
IAM-Aktionen für Claude Platform auf AWS ... 55
Service-Details ... 55
Aktionen ... 55

Inferenz 55
Batch-Verarbeitung 56
Modelle 56
Dateien 56
Skills 57
Benutzerprofile 58
Kundendienstmitarbeiters (Kundendienstmitarbeiter) 58
Sitzungen 59
Umgebungen 59
Tresore 60
Speicherspeicher 61
Webhooks 61
Arbeitsbereiche 62
Authentifizierung 62
Konsolenzugriff 63
Route-to-action Zuordnung 64
Weitere Informationen finden Sie auch unter 66
Dokumentverlauf 67
..... lxviii

Was ist Claude Platform auf AWS?

Greifen Sie über AWS mit Anthropic-managed Infrastruktur auf alle Plattformfunktionen von Claude zu.

Die Claude-Plattform auf AWS bietet Ihnen das gesamte Plattformerlebnis von Anthropic, einschließlich der Nachrichten-API, der Agentenfähigkeiten, der Codeausführung und der Beta-Funktionen, auf die Sie über Ihr AWS-Konto zugreifen können. Im Gegensatz zu Amazon Bedrock, wo AWS den Inferenzstapel betreibt, betreibt Anthropic die Claude-Plattform auf AWS. AWS stellt die Authentifizierungsebene (SigV4 oder API-Schlüssel), die IAM-based Zugriffskontrolle und die Abrechnungsintegration über AWS Marketplace bereit.

> **Note**
>
> Die Anthropic SDKs unterstützen die Claude-Plattform auf AWS. Informationen zur Verfügbarkeit der Clients pro Sprache finden Sie in der Dokumentation zu den Anthropic Client SDKs.

Wie funktioniert die Plattformintegration

Claude-Modelle laufen auf Anthropic-managed Infrastruktur. Dies ist eine kommerzielle Integration für Abrechnung und Zugriff über AWS. Sowohl Anthropic als auch AWS agieren als unabhängige Datenverarbeiter. Die AWS-Servicebedingungen regeln die Claude-Plattform auf AWS. Die kommerziellen Nutzungsbedingungen, der Zusatz zur Datenverarbeitung und die Nutzungsrichtlinie von Anthropic gelten ebenfalls.

Beachten Sie die folgenden Betriebsmerkmale. Daten befinden sich möglicherweise nicht in AWS. Inference kann an die primäre Cloud von Anthropic weitergeleitet werden. Unterdienste können sich ohne vorherige Ankündigung ändern. Stellen Sie den `inference_geo` Parameter pro Anfrage ein, um Rückschlüsse auf eine bestimmte Region zu ziehen. Einzelheiten finden Sie unter Datenresidenz.

Die Claude-Plattform auf AWS folgt derselben Datenaufbewahrungsrichtlinie wie die Claude-API von Erstanbietern. Zero Data Retention (ZDR) ist auf Anfrage erhältlich. Wenden Sie sich an Ihren Anthropic-Kundenbetreuer, um es für Ihr Konto zu aktivieren.

Funktionen der Claude-Plattform auf AWS

Die Claude-Plattform auf AWS bietet die folgenden Funktionen:

- Same-day Zugriff auf Modelle und Funktionen — Neue Claude-Modelle und API-Funktionen sind am selben Tag verfügbar, an dem sie auf der Claude-API von Erstanbietern eingeführt werden.
- Fähigkeiten für Agenten — Verwenden Sie vorgefertigte Skills für die Dokumentgenerierung (ExcelPowerPoint, Word, PDF) und erstellen Sie benutzerdefinierte Skills.
- Codeausführung — Führen Sie Code in der verwalteten Sandbox von Anthropic aus.
- Erweitertes Denken — Ermöglichen Sie erweitertes Denken für komplexe Denkaufgaben.
- Streaming und Batch-Verarbeitung — Verwenden Sie SSE-Streaming in Echtzeit oder senden Sie Batch-Anfragen für Workloads mit hohem Durchsatz.
- Promptes Caching — Cache-Tools, Systemaufforderungen und Nachrichtenverlauf, um Latenz und Kosten zu reduzieren.
- Datei-API — Dateien können bei allen Anfragen hochgeladen und referenziert werden.
- AWS-IAM-Integration — Steuern Sie den Zugriff mit Standard-IAM-Richtlinien und SigV4-Authentifizierung.
- Einheitliche AWS-Abrechnung — Zahlen Sie über Ihr bestehendes AWS-Konto mit verbrauchsbasierter Messung (Marketplace als Abrechnungs-Backend).

Die vollständige Liste der unterstützten Funktionen finden Sie unter Feature-Support.

Wie ist dieses Handbuch aufgebaut

Dieser Leitfaden behandelt die folgenden Themen:

- Erste Schritte — Kontoeinrichtung, Voraussetzungen, Authentifizierung und SDK-Installation.
- Verwenden der API — Verfügbare Modelle, Anfragen stellen und Funktionsunterstützung.
- Verwaltung Ihrer Umgebung — Datenresidenz, Workspaces, Claude Console, Tarifgrenzen und Abrechnung.
- Überwachung und Betrieb — CloudTrail Protokollierung und Nachverfolgung von Anfrage-IDs.
- Migration — Umstellung von Amazon Bedrock auf Claude Platform auf AWS.
- Sicherheit und Zugriffskontrolle — Referenz zu IAM-Richtlinien und -Aktionen.

Claude Platform auf AWS im Vergleich zu Amazon Bedrock

Mit beiden Angeboten können Sie Claude über AWS verwenden, sie unterscheiden sich jedoch erheblich in Architektur, API-Oberfläche und Funktionsverfügbarkeit.

	Claude-Plattform auf AWS	Amazon Bedrock
Wer betreibt den Stack	Anthropic	AWS
API-Oberfläche	API für anthropische Nachrichten () `/v1/messages`	Bedrock-API (Converse/) InvokeModel
Verfügbarkeit von Funktionen	Same-day als Claude API	Hängt vom Zeitplan der AWS-Integration ab
Fähigkeiten der Agenten	Available (Verfügbar)	Nicht verfügbar (erfordert Codeausführung)
Beta-Funktionen	Mit `anthropic-beta` Headern weiterleiten (siehe Feature-Unterstützung)	AWS-Support erforderlich
Authentifizierung	AWS IAM/SigV4 oder API-Schlüssel	AWS IAM/SigV4 oder Bearer-Token (nur C#-, Go- und Java-SDKs)
Basis-URL	`aws-external-anthropic.{region}.api.aws`	`bedrock-runtime.{region}.amazonaws.com`
SDK-Client	Platform-specific Client-Klasse (zum Beispiel `AnthropicAWS` in Python), in der Betaversion	`AnthropicBedrock` /Bedrock SDK
Konsole	Claude Console (`platform.claude.com` , Zugriff über die AWS-Konsole)	Bedrock-Konsole

	Claude-Plattform auf AWS	Amazon Bedrock
Ratenlimits und Kontingente	Verwaltet von Anthropic	Von AWS verwaltet
Datenverarbeiter	Anthropic und AWS	AWS

Wenn Sie AWS-operated Claude mit der Bedrock-API benötigen, finden Sie weitere Informationen unter Amazon Bedrock.

> **Important**
>
> Anthropic bietet Claude Platform auf AWS als Drittanbieterangebot an. Es ist nicht durch die standardmäßigen AWS-Compliance-Programme, Zertifizierungen oder Prüfberichte (wie die SOC-, ISO- oder HIPAA-Eignung) abgedeckt. Kunden sind allein dafür verantwortlich, ihre eigene Sorgfaltspflicht einzuhalten, um sicherzustellen, dass dieses Drittanbieterangebot ihren regulatorischen, rechtlichen und Compliance-Anforderungen entspricht.

Wann sollten Sie sich für Bedrock entscheiden: Wählen Sie Amazon Bedrock, wenn Ihr Unternehmen AWS-operated Inferenz, AWS als alleinigen Datenverarbeiter oder Schutz im Rahmen von AWS-Compliance-Programmen (einschließlich FedRAMP High, IL4, IL5, SOC, ISO und HIPAA-Eignung) benötigt. Bedrock läuft vollständig auf der AWS-controlled Infrastruktur mit AWS als Betriebspartei.

Einrichten Ihres Kontos

Die Einrichtung der Claude-Plattform auf AWS besteht aus vier Phasen: Melden Sie sich in der AWS-Konsole an, richten Sie Ihre Anthropic-Organisation ein, notieren Sie sich Ihre Workspace-ID und melden Sie sich bei der Claude-Konsole an.

> **Note**
>
> Wenn Sie sich über die AWS-Konsole registrieren, erhalten Sie eine neue anthropische Organisation, die mit Ihrem AWS-Konto verknüpft ist. Diese Organisation ist unabhängig von allen bestehenden Organisationen, die Ihr Unternehmen bei Anthropic unterhält, einschließlich Claude Enterprise-Organisationen, die über AWS Marketplace erworben wurden. API-Schlüssel, Workspaces und Claude Console-Einstellungen von einer Anthropic-Erstanbieter-Organisation werden nicht übernommen.
> Wenn Sie bereits über ein Privatangebot von Amazon Bedrock verfügen, wenden Sie sich vor der Registrierung an Ihren Kundenbetreuer bei Anthropic oder AWS, damit Ihr discount ab Ihrer ersten Anfrage gilt. Rabatte können nicht rückwirkend auf die Nutzung angewendet werden, die vor der Annahme Ihres privaten Angebots erfolgt ist. Siehe Private Angebote.

Schritt 1: Melden Sie sich in der AWS-Konsole an

1. Öffnen Sie die AWS-Konsole und navigieren Sie zur Serviceseite Claude Platform on AWS.
2. Wählen Sie „Registrieren“.
3. Klicken Sie auf Weiter.

Auf der Seite wird ein Banner „Sign-up In Bearbeitung“ angezeigt. Bleib auf der Seite. Sign-up dauert ein paar Minuten, bis AWS das AWS Marketplace Marketplace-Abonnement für Sie abwickelt, und leitet Sie dann automatisch weiter.

Wenn Ihre Organisation über ein privates Angebot von Anthropic verfügt, sucht die Konsole danach und fordert Sie auf, es im AWS Marketplace anzunehmen. Einzelheiten finden Sie unter Private Angebote.

> **Note**
>
> Wenn Sie Claude Platform auf AWS verwenden, werden Ihre Inhalte (wie Eingabeaufforderungen und Vervollständigungen) von Anthropic außerhalb von AWS verarbeitet. Einzelheiten zur Verarbeitung und Speicherung von Inhalten und Metadaten finden Sie in den Datennutzungsrichtlinien von Anthropic.

Schritt 2: Richten Sie Ihre Anthropic-Organisation ein

Nach Abschluss der Registrierung werden Sie weitergeleitet zu. `platform.claude.com/partner-signup`

1. Geben Sie die E-Mail-Adresse des Inhabers Ihrer Organisation ein und wählen Sie Erste Schritte aus.
2. Suchen Sie in diesem E-Mail-Posteingang nach einem Link zur Einrichtung und folgen Sie ihm. Wenn Ihr Browser die Seite „Als anderes Konto angemeldet“ anzeigt, wählen Sie Abmelden und fortfahren.
3. Füllen Sie das Formular mit den Organisationsdetails aus (Name der Organisation, Entitätstyp, Land, Verwendungszweck) und wählen Sie Einrichtung abschließen aus.

Wenn Sie die Einrichtung abschließen, wird Ihre Anthropic-Organisation erstellt. Die AWS-Servicebedingungen regeln die Claude-Plattform auf AWS. Die kommerziellen Nutzungsbedingungen, der Zusatz zur Datenverarbeitung und die Nutzungsrichtlinie von Anthropic gelten ebenfalls. Auf der Serviceseite der AWS-Konsole wird jetzt eine linke Navigation mit Home, API-Schlüsseln, Schnellstart und Workspaces angezeigt.

Schritt 3: Notieren Sie sich Ihre Workspace-ID

Ein Standard-Workspace wird automatisch bereitgestellt, wenn du dich registrierst. Pro Region können zusätzliche Workspaces erstellt werden. Weitere Informationen zur Regionsbindung, Workspace-Verwaltung und IAM-Ressourceneinteilung finden Sie unter Workspaces.

Suchen Sie die Workspace-ID unter Workspaces auf der AWS-Serviceseite der AWS-Konsole Claude Platform oder in der Claude-Konsole (siehe Verwenden der Claude-Konsole). Workspace-IDs verwenden das Format, `wrkspc_` gefolgt von einer alphanumerischen Kennung.

Schritt 4: Melden Sie sich bei der Claude Console an

Der Zugriff auf die Claude Console ist über AWS IAM gebündelt:

1. Nehmen Sie eine IAM-Rolle mit der entsprechenden Genehmigung an. `aws-external-anthropic:AssumeConsole` Siehe IAM-Aktionen.
2. Wählen Sie auf der Serviceseite Claude Platform on AWS die Option Anmelden aus. Die AWS-Konsole gibt ein JWT aus und leitet Sie weiter zu`platform.claude.com`.
3. Bei der ersten Anmeldung werden Sie zur Eingabe einer E-Mail-Adresse aufgefordert. Geben Sie Ihre geschäftliche E-Mail-Adresse ein. Die Plattform stellt Ihren Claude Console-Benutzer just-in-Time bereit.

Wenn Sie über die AWS-Konsole angemeldet sind, gilt die Claude-Konsole für Ihre Claude-Plattform auf AWS-Organisation. In der Seitenleiste von Claude Console wird eine Anzeige „Von AWS verwaltetes Konto“ angezeigt.

Problembehebung bei der Kontoeinrichtung

- "Sign-up fehlgeschlagen: Fehler beim Aktivieren OutboundWebIdentityFederation „: Wenn Sie beim ersten Absenden dieses rote Banner sehen, wählen Sie erneut Weiter. Es kann einen Moment dauern, bis die IAM-Aktivierung wirksam wird.
- Keine Fortschrittsanzeige bei der Anmeldung: Sign-up dauert ein paar Minuten. Auf der Seite wird ein statisches „Sign-up In Bearbeitung“ -Banner ohne Fortschrittsbalken angezeigt, während AWS Ihr Konto bereitstellt.
- „Mit einem anderen Konto angemeldet“, nachdem Sie dem Link zur Einrichtung gefolgt sind: Wählen Sie Abmelden und fortfahren. Auf der Seite werden Sie erneut mit der von Ihnen eingegebenen E-Mail-Adresse authentifiziert.
- Meldung „Nicht gefunden“ bei der Anmeldung: Diese Meldung kann während der Weiterleitung kurz erscheinen. Sie können sie verwerfen.
- Auf der Nutzungsseite werden nach Ihrem ersten API-Aufruf keine Daten angezeigt: Es kann einige Minuten dauern, bis Nutzungsdaten in der Claude-Konsole angezeigt werden.

Voraussetzungen

Bevor Sie API-Aufrufe tätigen, stellen Sie sicher, dass Sie über Folgendes verfügen:

1. Ein aktives AWS-Konto mit einem Abonnement für Claude Platform auf AWS (siehe Konto einrichten).
2. Die AWS-CLI wurde installiert und konfiguriert.
3. Ausgehender Web-Identitätsverbund ist auf Ihrem AWS-Konto aktiviert (ein einmaliger Einrichtungsschritt; siehe Ausgehenden Web-Identitätsverbund aktivieren).
4. Ihre Workspace-ID (siehe Beziehen Sie Ihre Workspace-ID).

Aktiviere den Identitätsverbund für ausgehende Webidentitäten

Die Claude-Plattform auf dem AWS-Gateway ruft `sts:GetWebIdentityToken` serverseitig auf, um ein JWT zu prägen, das es an Anthropic weiterleitet. Diese STS-Funktion ist standardmäßig für jedes AWS-Konto deaktiviert. Aktivieren Sie sie einmal pro Konto:

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

Wenn die Antwort lautet`[ERROR] (FeatureEnabled) … already enabled` , ist die Einstellung für Ihr Konto bereits aktiviert und Sie können weitermachen. Überprüfen Sie die Aussteller-URL Ihres Kontos und rufen Sie sie ab:

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

> **Warning**
>
> Ohne diesen Schritt wird jede Anfrage zurückgegeben`"Outbound web identity federation is disabled for your account"`. Dies ist der häufigste Einrichtungsfehler.

Besorgen Sie sich Ihre Workspace-ID

Wenn du dich registrierst, wird in jeder Region automatisch ein Standard-Workspace bereitgestellt (siehe Konto einrichten). Workspaces sind an eine einzige AWS-Region gebunden. Sie finden die

Workspace-ID in der Claude-Konsole unter Workspaces oder im Bereich Workspaces der AWS-Konsolen-Serviceseite. Informationen zur Regionsbindung und zum IAM-Ressourcen-Scoping finden Sie unter Workspaces.

Stellen Sie die Umgebungsvariablen `ANTHROPIC_AWS_WORKSPACE_ID` und die `AWS_REGION` Umgebungsvariablen so ein, dass die SDK-Clients sie automatisch lesen:

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

Die Region ist erforderlich. Der SDK-Client wird ausgelöst, wenn keine Region festgelegt ist. `aws_region`Übergibt/`awsRegion`an den Konstruktor oder set `AWS_REGION` (oder`AWS_DEFAULT_REGION`). Alle kommerziellen AWS-Regionen werden unterstützt. Für Opt-in-Regionen müssen Sie das Konto zunächst für die Region anmelden.

Authentifizierung

Die Claude-Plattform auf AWS unterstützt zwei Authentifizierungsmethoden: AWS IAM mit SigV4-Anforderungssignierung (primär) und API-Schlüsselauthentifizierung. Beide verwenden dieselbe Basis-URL und dasselbe Anforderungsformat.

Die Anthropic-SDKs (siehe SDK installieren) implementieren den unten beschriebenen Authentifizierungsablauf und die Auflösung von Anmeldeinformationen. Wenn Sie kein Anthropic-SDK verwenden, müssen Sie SigV4-signed Anfragen für den regionalen Endpunkt erstellen (oder Ihren API-Schlüssel als Bearer-Token präsentieren). `https://aws-external-anthropic.<region>.api.aws` Informationen zur SDK-specific Client-Konfiguration und zur verbindlichen Reihenfolge zur Auflösung von Anmeldeinformationen finden Sie im Claude on AWS-Einrichtungsleitfaden in der Anthropic-Dokumentation.

SigV4-Authentifizierung

SigV4 ist der unternehmensnative Pfad und lässt sich in Ihre bestehenden AWS IAM-Richtlinien, -Rollen und -Audits integrieren. Konfigurieren Sie AWS-Anmeldeinformationen mit einer beliebigen Methode, die von der AWS-Standardanbieterkette für Anmeldeinformationen unterstützt wird:

- Umgebungsvariablen (`AWS_ACCESS_KEY_ID`,`AWS_SECRET_ACCESS_KEY`,`AWS_SESSION_TOKEN`)
- Datei mit gemeinsam genutzten Anmeldeinformationen (`~/.aws/credentials`)
- Gemeinsam genutzte Konfigurationsdatei (`~/.aws/config`) einschließlich SSO und `credential_process`
- Webidentität (`AWS_WEB_IDENTITY_TOKEN_FILE`und`AWS_ROLE_ARN`) für IRSA und Aktionen GitHub
- Anmeldeinformationen für den ECS-Container
- Metadatendienst für EC2-Instanzen (IMDS)

Um zu überprüfen, ob das Anthropic SDK Ihre Anmeldeinformationen abruft und zum richtigen regionalen Endpunkt weiterleitet, stellen Sie eine Testanfrage gemäß den Beispielen unter Anfragen stellen. Eine erfolgreiche Antwort bestätigt, dass Anmeldeinformationen, Region, Basis-URL und Workspace-ID alle korrekt konfiguriert sind.

API-Schlüsselauthentifizierung

Für einfachere Integrationspfade (lokale Entwicklung und Skripte) können Sie sich mit einem API-Schlüssel statt mit SigV4 authentifizieren. Legen Sie die `ANTHROPIC_AWS_API_KEY` Umgebungsvariable fest oder übergeben Sie `apiKey` sie an den SDK-Konstruktor. Das Anthropic SDK sendet den Schlüssel als Bearer-Token an die Claude-Plattform auf dem AWS-Endpunkt. IAM autorisiert die Anfrage über die `aws-external-anthropic:CallWithBearerToken` Aktion (siehe IAM-Richtlinien).

Generieren Sie API-Schlüssel in der AWS-Konsole unter Claude Platform auf AWS → API-Schlüssel. Wählen Sie Schlüssel generieren und kopieren Sie dann den Schlüsselwert. Erteilen Sie die `aws-external-anthropic:CallWithBearerToken` IAM-Aktion den Prinzipalen, die die API-Schlüsselauthentifizierung verwenden dürfen sollen.

> **Note**
>
> Nur API-Schlüssel, die in der AWS-Konsole unter Claude Platform auf AWS erstellt wurden, funktionieren mit diesem Service.
>
> - Schlüssel, die in der Standard-Claude-Konsole für den API-Zugriff von Erstanbietern erstellt wurden, funktionieren nicht für die Claude-Plattform auf dem AWS-Endpunkt.
> - Amazon Bedrock API-Schlüssel funktionieren auch nicht — Bedrock verwendet einen separaten Endpunkt, Authentifizierungsablauf und IAM-Namespace.

Short-term API-Schlüssel

Für Fälle, in denen Sie eine API-Schlüsselauthentifizierung wünschen, aber nicht die Lebensdauer eines langlebigen Schlüssels haben, veröffentlicht Anthropic Token-Generator-Bibliotheken, die kurzfristige API-Schlüssel aus Ihren AWS IAM-Anmeldeinformationen ableiten. Token haben standardmäßig eine Gültigkeitsdauer von 12 Stunden und können auf einen bestimmten Workspace und die Aktion beschränkt werden. `aws-external-anthropic:CallWithBearerToken` Installieren Sie den Generator für Ihre Sprache, tauschen Sie die IAM-Anmeldeinformationen gegen einen API-Schlüssel aus und übergeben Sie den Schlüssel an das Anthropic SDK.

- Python: aws/token-Generator-für-AWS-external-anthropic-Python
- JavaScript / TypeScript: aws/token -generator-fuer-aws-external-anthropic-js

- Java: aws/token-Generator-für-AWS-External-Anthropic-Java

Short-term API-Schlüssel erfordern weiterhin, dass der IAM-Principal des Aufrufers auf dem Ziel-Workspace bleibt. `aws-external-anthropic:CallWithBearerToken` Sie laufen von selbst ab und müssen nicht explizit rotiert oder gesperrt werden.

Vorrang der Anmeldeinformationen und Auflösung der Region

Die Claude-Plattform auf dem AWS-Client des Anthropic SDK löst Anmeldeinformationen und Region anhand einer definierten Prioritätsreihenfolge auf. Die Namen der Argumente folgen den Konventionen der einzelnen Sprachen: camelCase für TypeScript und PHP, snake_case für Python und Ruby, für Go und Property- oder Builder-Muster PascalCase für C# und Java.

Die autoritative Rangfolge, die unterstützten Umgebungsvariablen und die Konstruktorargumente für jedes SDK finden Sie in der Anthropic-Dokumentation unter Claude on AWS setup.
Im Allgemeinen haben explizite Konstruktorargumente Vorrang vor Umgebungsvariablen und `ANTHROPIC_AWS_API_KEY` haben Vorrang vor der standardmäßigen AWS-Anmeldeinformationsanbieterkette. Region ist erforderlich; im Gegensatz zu `AnthropicBedrock` (was auf zurückgreiftus`-east-1`) wird der Claude on AWS-Client ausgelöst, wenn keine Region vom Konstruktor oder von`AWS_REGION`/angegeben wird. `AWS_DEFAULT_REGION`

Arbeitsbereich-ID

Jede Anfrage auf Datenebene muss die Workspace-ID im Header enthalten. `anthropic-workspace-id` Die Anthropic-SDKs lesen dies standardmäßig aus der `ANTHROPIC_AWS_WORKSPACE_ID` Umgebungsvariablen, oder Sie können`workspaceId`/`workspace_id`an den Client-Konstruktor übergeben. Wenn Sie den `Anthropic` Basisclient (nicht den Claude-on-AWS Client) verwenden, übergeben Sie den Header explizit. Informationen dazu, wie du deine Workspace-ID findest, findest du unter Anfragen für Beispiele pro Sprache stellen und Workspaces.

Installieren eines -SDK

Die Kunden-SDKs von Anthropic unterstützen die Claude-Plattform auf AWS. Alle sieben SDKs bieten eine plattformspezifische Clientklasse. Alternativ können Sie den Basisclient mit der Basis-URL der Claude-Plattform auf AWS konfigurieren.

Python

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

> **Tip**
>
> Auf macOS mit Homebrew Python oder anderen extern verwalteten Python-Umgebungen `pip install` kann es mit einem PEP 668-Fehler fehlschlagen. `externally-managed-environment` Erstellen und aktivieren Sie zuerst eine virtuelle Umgebung:. `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**
>
> SDK-Clients für Claude Platform auf AWS befinden sich in der Beta-Phase.

Verfügbare Modelle

Die Claude-Plattform auf AWS bietet dieselben Claude-Modelle wie die Claude-API von Erstanbietern.

Die aktuelle Liste der Modelle, Modell-IDs, Größen und Funktionen von Kontextfenstern finden Sie in der Anthropic-Dokumentation unter Modellübersicht.

Modell-IDs

Modell-IDs auf der Claude-Plattform auf AWS sind identisch mit denen, die auf der Claude-API von Erstanbietern verwendet werden (z. B.`claude-opus-4-6`,`claude-sonnet-4-6`,`claude-haiku-4-5`). Es gibt keine Bedrock-style ARNs oder `anthropic.` Präfixe — verwenden Sie die Modell-ID genau so, wie Anthropic sie veröffentlicht.

Senden von Anforderungen

Die Claude-Plattform auf AWS verwendet die Anthropic Messages API (`/v1/messages`), dieselbe API-Oberfläche wie die Claude-Erstanbieterplattform. Die Unterschiede bestehen in der Basis-URL, der Authentifizierungsmethode und einem erforderlichen `anthropic-workspace-id` Header, der angibt, auf welchen Workspace die Anfrage abzielt.

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
```

Der Client liest `AWS_REGION` (oder`AWS_DEFAULT_REGION`) `ANTHROPIC_AWS_WORKSPACE_ID` aus der Umgebung. Sie können entweder überschreiben, indem Sie`aws_region`/`awsRegion`oder`workspace_id`/`workspaceId`an den Konstruktor übergeben. Sowohl die Regions- als auch die Workspace-ID sind erforderlich. Der Konstruktor gibt aus, wenn beide Werte nicht aus diesen Quellen aufgelöst werden können.

> **Note**
>
> Der `x-amz-security-token` Header (curl) ist nur für temporäre Anmeldeinformationen wie IAM-Rollen, SSO oder STS erforderlich. Lassen Sie ihn weg, wenn Sie langfristige IAM-Benutzeranmeldeinformationen verwenden. Die SDK-Clients verarbeiten dies automatisch auf der Grundlage der Quelle der Anmeldeinformationen.

Der `--aws-sigv4` Wert folgt dem Format`aws:amz:<region>:<service>`. Der SigV4-Dienstname lautet`aws-external-anthropic`, und die Region muss mit der Region in Ihrer Endpunkt-URL übereinstimmen. Eine Nichtübereinstimmung in beiden Fällen führt eher zu einem generischen Fehler bei der Ablehnung der Signatur als zu einer spezifischen Diagnose.

Verwenden Sie den Anthropic-Basisclient

Wenn Sie die AWS-SDK-Abhängigkeit nicht hinzufügen möchten, können Sie den `Anthropic` Basisclient verwenden. Dieser Pfad verwendet die Namen der Vanille-SDK-Umgebungsvariablen und nicht die `ANTHROPIC_AWS_*` Namen, die der dedizierte Client liest:

```
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'
```

Die Python- TypeScript, und Go-SDKs werden gelesen `ANTHROPIC_API_KEY` und `ANTHROPIC_BASE_URL` automatisch; für die anderen Sprachen werden sie an den Konstruktor übergeben. Übergeben Sie in jeder Sprache die Workspace-ID in der `anthropic-workspace-id` Kopfzeile.

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);
```

Feature-Unterstützung

Die Claude-Plattform auf AWS verwendet die Anthropic Messages API direkt. Sie erhalten die volle Funktionsparität mit der Claude-API von Erstanbietern, außer in den Fällen, in denen dies unter Derzeit nicht verfügbare Funktionen angegeben ist:

- Beta-Funktionen: Übergeben Sie den `anthropic-beta` Standard-Header, um auf Beta-Funktionen zuzugreifen, genau wie bei der Claude-API.
- Agentenfähigkeiten: Verwenden Sie vorgefertigte und benutzerdefinierte Agentenfähigkeiten mit denselben `container.skills` Parametern und Beta-Headern wie bei der Claude-API. Alle vorgefertigten Skills (ExcelPowerPoint, Word, PDF) sind sofort einsatzbereit.
- Codeausführung: Führen Sie mithilfe des Tools zur Codeausführung Code in der verwalteten Sandbox von Anthropic aus.
- Verwendung von Tools: Computernutzung und alle anderen Funktionen zur Verwendung von Tools sind verfügbar.
- Erweitertes Denken: Ermöglichen Sie erweitertes Denken mit denselben Parametern wie die Claude-API.
- Streaming: Vollständige SSE-Streaming-Unterstützung für Antworten in Echtzeit.
- Stapelverarbeitung: Senden Sie Batch-Anfragen für Workloads mit hohem Durchsatz.
- Promptes Caching: Cache-Tools, Systemaufforderungen und Nachrichtenverlauf, um Latenz und Kosten zu reduzieren. Alle Funktionen zum Zwischenspeichern von Eingabeaufforderungen (TTL von 5 Minuten, TTL von 1 Stunde und automatisches Caching) sind verfügbar.
- Datei-API: Dateien bei Anfragen hochladen und referenzieren.
- Workspace-Tags mit IAM-Integration: Workspaces können mit Tags versehen werden, und die Tags werden an IAM für die attributebasierte Zugriffskontrolle und an AWS-Kosten- und Nutzungsberichte für die Kostenzuweisung weitergeleitet. Workspace-Tagging ist eine allgemeine Funktion auf der Claude-Plattform auf AWS (es handelt sich um eine Beta-Funktion der Claude-API von Erstanbietern).

Von Claude verwaltete Agenten

Claude Managed Agents sind auf der Claude-Plattform auf AWS verfügbar. Agenten, Sitzungen, Umgebungen, Anmeldeinformationstresore und Speicherspeicher sind erstklassige IAM-Ressourcen. Die Aktionsreferenz finden Sie unter IAM-Aktionen und die entsprechenden Seiten unter Verwenden

der Claude-Konsole. Das Anthropic Agent SDK funktioniert mit Claude Platform auf AWS ohne zusätzliche Integration. Zeigen Sie auf die Basis-URL der Claude-Plattform auf AWS und authentifizieren Sie sich mit Sigv4 oder einem API-Schlüssel.

Autonome Sitzungen auf der Claude-Plattform auf AWS müssen alle 6 Stunden erneut authentifiziert werden. Long-running Agentenläufe müssen ihre SigV4-Anmeldeinformationen oder ihren API-Schlüssel innerhalb dieses Fensters aktualisieren, andernfalls endet die Sitzung.

Die folgenden Funktionen von Claude Managed Agents sind derzeit auf der Claude-Plattform auf AWS nicht verfügbar:

- Ergebnisse: Die Ergebnisverfolgung für Agentensitzungen ist nicht verfügbar.
- Multi-agent Sitzungen: Sitzungen mit mehreren interagierenden Agenten sind nicht verfügbar.

Compliance

> **Important**
>
> Anthropic bietet diesen Service als Angebot eines Drittanbieters an. Es ist nicht durch die standardmäßigen AWS-Compliance-Programme, Zertifizierungen oder Prüfberichte (wie die SOC-, ISO- oder HIPAA-Eignung) abgedeckt. Kunden sind allein dafür verantwortlich, ihre eigene Sorgfaltspflicht einzuhalten, um sicherzustellen, dass dieses Drittanbieterangebot ihren regulatorischen, rechtlichen und Compliance-Anforderungen entspricht. Bevor Sie sensible oder regulierte Daten verarbeiten, sollten Sie die offizielle Compliance-Dokumentation von Anthropic im Anthropic Trust Center überprüfen. Weitere Informationen finden Sie in den AWS-Servicebedingungen.

Wie wird die Workspace-Mitgliedschaft modelliert

Bei der Claude-API von Erstanbietern wird der Zugriff durch die Mitgliedschaft von Benutzern zu Workspaces geregelt — ein Benutzer wird zu einem Workspace hinzugefügt und erbt die Zugriffsrechte des Workspace. Claude Platform auf AWS ersetzt dieses Modell durch die IAM-Autorisierung: Es gibt keine Benutzerliste pro Workspace. Stattdessen gelten für IAM-Prinzipale (Benutzer oder Rollen) Richtlinien, die `aws-external-anthropic` Aktionen gegen bestimmte Workspace-ARNs zulassen. Die Bewertung der IAM-Richtlinien bestimmt, was jeder Prinzipal in jedem Workspace tun kann.

Gewähren und entziehen Sie den Workspace-Zugriff, indem Sie die IAM-Richtlinien (SCPs, Berechtigungsgrenzen, identitätsbezogene Richtlinien oder Bedingungen auf Ressourcenebene) ändern, anstatt die Mitgliedschaft pro Workspace in der Claude-Konsole zu verwalten. Beispiele finden Sie unter IAM-Richtlinien.

Funktionen, die derzeit nicht verfügbar sind

Die folgenden Funktionen sind derzeit auf der Claude-Plattform auf AWS nicht verfügbar:

- HIPAA-Bereitschaft: Das HIPAA-ready Programm von Anthropic ist auf der Claude-Plattform auf AWS nicht verfügbar. Kunden mit HIPAA-Anforderungen sollten Claude stattdessen in Amazon Bedrock testen.
- Servicestufen, die über Standard und Batch hinausgehen: Priority Tier ist nicht verfügbar.
- Admin-API — Endpunkte für Organisationsmitglied, Workspace-Mitglied, Einladung, API-Schlüssel, Nutzungsbericht, Kostenbericht und Ratenlimitbericht: Diese Admin-API-Endpunkte sind derzeit nicht verfügbar. Workspace CRUD- und Rename-Endpunkte (`/v1/organizations/workspaces`) sind über den `aws-external-anthropic` Namespace verfügbar; siehe IAM-Aktionen. Die Mitgliedschaft wird anhand von AWS IAM und nicht anhand von Workspace-Mitgliedslisten modelliert (siehe vorhergehender Abschnitt). Der API-Schlüssellebenszyklus wird in der AWS-Konsole unter Claude Platform auf AWS → API-Schlüssel verwaltet. Daten zur Nutzung, zu den Kosten und zum Ratenlimit finden Sie in den Ansichten der Claude-Konsole (siehe Überwachung und Protokollierung) oder in der API „Anthropic Usage and Cost“.
- Ausgabenlimits: Nicht verfügbar. Verlassen Sie sich stattdessen auf AWS-Abrechnungskontrollen.
- Claude Code-Workspace und Analytics-API: Der Claude Code-Workspace mit automatischen Ratenlimits ist nicht verfügbar. Die Verwendung von Claude Code wird in der allgemeinen Nutzungsansicht und nicht in einem speziellen Bildschirm angezeigt.
- OAuth-Authentifizierung: Wird nicht unterstützt. Verwenden Sie die Sigv4- oder API-Schlüsselauthentifizierung.
- OpenAI-compatible API-Endpunkte: Nicht verfügbar auf der Claude-Plattform auf AWS.
- Workspace-level Geografische Inferenzkontrollen: `allowed_inference_geos` und `default_inference_geo` sind nicht verfügbar. Wird stattdessen `inference_geo` bei jeder Anfrage festgelegt.

Datenresidenz

Die Claude-Plattform auf AWS unterstützt die folgenden Inferenzregionen:

- USA: Inference verbleibt in US-Rechenzentren. Es gilt ein 1,1-facher Preismultiplikator.
- Global: Inference kann an jedes Anthropic-operated Rechenzentrum weltweit weitergeleitet werden. Es gelten die Standardpreise.

Legen Sie die Geografie der Inferenz pro Anfrage mit dem folgenden `inference_geo` Parameter fest:

> **Note**
>
> Der `inference_geo` Parameter wird von Claude Opus 4.6, Claude Sonnet 4.6 und späteren Modellen unterstützt. Anfragen `inference_geo` auf Claude Opus 4.5, Claude Sonnet 4.5 oder Claude Haiku 4.5 geben einen 400-Fehler zurück. Einzelheiten zur Verfügbarkeit von Modellen finden Sie unter Datenresidenz.

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
```

Wenn Sie es weglassen`inference_geo`, ist die Anforderung standardmäßig auf. `global`

Workspace-level Inference Geography Controls (`allowed_inference_geos`und`default_inference_geo`) sind auf der Claude-Plattform auf AWS nicht verfügbar. Wird stattdessen `inference_geo` bei jeder Anfrage festgelegt.

Arbeitsbereiche

Inferenz- und Ressourcenanfragen auf der Claude-Plattform auf AWS zielen auf einen Workspace ab. Bei diesen API-Aufrufen übergeben Sie die Workspace-ID im `anthropic-workspace-id` Header. Workspace-IDs verwenden das Tag-Format, `wrkspc_` gefolgt von einer alphanumerischen Kennung (z. B.`wrkspc_01AbCdEf23GhIj`). Weitere Informationen findest du unter Ermitteln deiner Workspace-ID, falls du sie noch nicht hast.

Geltungsbereich des Arbeitsbereichs

Workspaces sind an eine einzige AWS-Region gebunden. Auf einen Workspace, der in erstellt wurde, `us-west-2` kann nur über den `us-west-2` Endpunkt zugegriffen werden. Nutzung, Kontingente, Kosten, Dateien, Batches und Fähigkeiten werden alle pro Workspace zusammengefasst, sodass Sie in der Claude-Konsole nach Regionen aufgeschlüsselt sind.

Workspaces dienen auch als primäre IAM-Ressource für Claude Platform auf AWS. Sie gewähren oder verweigern den Zugriff auf bestimmte Workspaces über AWS IAM-Richtlinien mithilfe des Workspace-ARN. Das Ressourcensegment des ARN hat dieselbe ID mit dem `wrkspc_` Präfix, die Sie im Header übergeben: `anthropic-workspace-id`

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

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

Richtlinienbeispiele finden Sie unter IAM-Richtlinien.

Arbeitsbereiche erstellen

Ein Standard-Workspace wird bei der Registrierung automatisch bereitgestellt. Zusätzliche Workspaces können über die `/v1/organizations/workspaces` Endpoints erstellt, aktualisiert, umbenannt und archiviert werden, die durch die entsprechenden `aws-external-anthropic` Aktionen (, `CreateWorkspaceUpdateWorkspace,ArchiveWorkspace`) autorisiert werden; siehe IAM-Aktionen.

Einstellung der Workspace-ID in Ihrem Client

Jede Anfrage auf Datenebene muss die Workspace-ID in der `anthropic-workspace-id` Kopfzeile enthalten. Wenn Sie den Claude-on-AWS Client eines Anthropic SDK verwenden, gibt es drei Möglichkeiten, ihn bereitzustellen:

1. Umgebungsvariable — gesetzt`ANTHROPIC_AWS_WORKSPACE_ID`, und der Client liest sie automatisch.

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

2. Konstruktorargument — übergeben Sie`workspaceId`/`workspace_id`(Name folgt der Sprachkonvention des SDK), wenn Sie den Client instanziieren.
3. Per-request override — Übergeben Sie den Header direkt bei einer einzelnen Anfrage, wenn Sie mehrere Arbeitsbereiche vom selben Client aus adressieren müssen.

Wenn Sie den `Anthropic` Basisclient (ohne das AWS-Backend) verwenden, legen Sie den `anthropic-workspace-id` Header bei jeder Anfrage explizit fest — Beispiele pro Sprache finden Sie unter Anfragen stellen. Die Namen der SDK-specific Argumente finden Sie in der Dokumentation zu den Anthropic Client SDKs.

Markieren von Workspaces

Workspace-Tags sind Schlüssel-Wert-Paare, die an einen Workspace angehängt sind. Sie lassen sich in AWS IAM für die attributbasierte Zugriffskontrolle und in AWS-Kosten- und Nutzungsberichte für die Kostenzuweisung integrieren (Informationen zu CUR finden Sie unter Überwachung und Protokollierung). Tagging ist auf Claude Platform auf AWS allgemein verfügbar.

Aktualisieren Sie die Tags im vorhandenen Workspace mit `TagResource` und `UntagResource` im `aws-external-anthropic` Namespace. Dieselben Tag-Schlüssel können die IAM-Bedingungen und die Kostenzuweisung ohne zusätzliche Konfiguration beeinflussen.

Tag-based Zugriffskontrolle

Mithilfe des `aws:ResourceTag/<key>` Bedingungskontextschlüssels können Sie `aws-external-anthropic` Aktionen für Workspace-Tag-Werte steuern. Die folgende Richtlinie erlaubt nur Rückschlüsse auf markierte Workspaces: `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"
        }
      }
    }
  ]
}
```

Verwenden Sie die Tasten `aws:RequestTag/<key>` und geben Sie die `aws:TagKeys` Bedingung ein`TagResource`, um Tag-Richtlinien durchzusetzen (z. B. die Anforderung, dass der Workspace Tags enthalten `CostCenter` muss). `Owner` Weitere Beispiele finden Sie unter IAM-Richtlinien.

Verwendung der Claude-Konsole

Claude Platform auf AWS verwendet die Standard-Claude-Konsole unter platform.claude.com. Wenn Sie sich von der AWS-Konsole aus anmelden, wird in der Seitenleiste der Claude Console die Anzeige Account managed by AWS angezeigt. Die Konsole gilt für Ihre Claude-Plattform auf AWS-Organisation. Sie bietet Nutzungsanalysen, Kostenaufschlüsselungen, Sichtbarkeit von Ratenlimits, Sichtbarkeit von Arbeitsbereichen und Seiten für die Verwaltung von Dateien, Agentenfähigkeiten, Batch-Jobs, Agenten, Sitzungen, Umgebungen, Anmeldeinformationstresoren und Speicherspeichern.

Anmelden

Der Zugriff auf die Claude-Konsole wird über AWS IAM gebündelt. Den vollständigen Ablauf der Erstanmeldung finden Sie unter Konto einrichten. Kurz gesagt:

1. Nehmen Sie eine IAM-Rolle mit der entsprechenden `aws-external-anthropic:AssumeConsole` Genehmigung an. Siehe IAM-Aktionen.
2. Navigieren Sie in der AWS-Konsole zur Seite Claude Platform on AWS.
3. Wählen Sie Open Claude Console. Die AWS-Konsole gibt ein JWT aus und leitet Sie weiter zu `platform.claude.com`.
4. Bei der ersten Anmeldung werden Sie zur Eingabe einer E-Mail-Adresse aufgefordert. Geben Sie Ihre geschäftliche E-Mail-Adresse ein. Die Plattform stellt Ihren Claude Console-Benutzer just-in-Time bereit.

Es sind zwei Claude Console-Rollen verfügbar: Administrator und Entwickler. Die Admin-Rolle gewährt Zugriff auf alle Claude Console-Seiten und Einstellungen, die für die Claude Platform auf AWS verfügbar sind. Die Entwicklerrolle gewährt Lesezugriff auf Nutzungs-, Kosten-, Ratenlimit- und Workspace-Informationen. Wenden Sie sich an Ihren Kundenbetreuer bei Anthropic, um einem Principal die Admin- oder Entwicklerrolle zuzuweisen.

Verfügbare Seiten

Die Spalte Via AWS-Gateway gibt an, ob die Seite Daten über das AWS-Gateway liest und schreibt (und daher durch IAM-Aktionen gesteuert wird). Seiten, die mit Nein markiert sind, lesen Metadaten auf Organisationsebene direkt über Anthropic-APIs und umgehen IAM-Aktionsprüfungen.

Seite	Available (Verfügbar)	Über das AWS-Gateway	Hinweise
Usage	Ja	Nein	Zeigen Sie die Token-Nutzung nach Modell, Arbeitsbereich und Dimension an. Es kann einige Minuten dauern, bis Daten nach einer Anfrage angezeigt werden.
Kosten	Ja	Nein	Sehen Sie sich Kostenaufschlüsselungen nach Modell und Arbeitsbereich an. AWS Cost Explorer zeigt den aggregierten CCU-Einzelposten an.
Beschränkungen	Ja	Nein	Ratenlimits anzeigen (schreibgeschützt).
Arbeitsbereiche	Ja	Nein	Arbeitsbereiche pro Region anzeigen (schreibgeschützt).
Datei-	Ja	Ja	Hochgeladene Dateien anzeigen und verwalten.
Fähigkeiten	Ja	Ja	Agentenfähigkeiten anzeigen und verwalten.
Batches	Ja	Ja	Batch-Verarbeitungsaufträge anzeigen und verwalten.
Kundendienstmitarbeiters (Kundendienstmitarbeiter)	Ja	Ja	Agentendefinitionen anzeigen und verwalten.
Sitzungen	Ja	Ja	Agentensitzungen und den Ereignisverlauf anzeigen.

Seite	Available (Verfügbar)	Über das AWS-Gateway	Hinweise
Umgebungen	Ja	Ja	Cloud-Container-Konfigurationen für Sitzungen anzeigen und verwalten.
Tresore für Anmeldeinformationen	Ja	Ja	Anmeldeinformationstresore für die Sitzungsauthentifizierung anzeigen und verwalten.
Speichert Speicher	Ja	Ja	Persistenten Agentenspeicher anzeigen und verwalten.
API-Schlüssel	Nein	N/A	API-Schlüssel in der AWS-Konsole verwalten (Claude-Plattform auf AWS → API-Schlüssel). Siehe Authentifizierung.
Mitglieder	Nein	N/A	Nicht zutreffend. AWS IAM verwaltet den Zugriff.
Fakturierung	Nein	N/A	Nicht zutreffend. AWS Marketplace verwaltet die Abrechnung und Rechnungsstellung. Kostenaufschlüsselungen finden Sie auf der Kostenseite.
Claude Code	Nein	N/A	Sehen Sie sich die Verwendung von Claude Code auf der Nutzungsseite an.

Organisation wechseln

Die Claude-Konsole unterstützt keinen Organisationswechsel zur Claude-Plattform auf AWS. Um auf eine andere Organisation zuzugreifen, melden Sie sich über die AWS-Konsole ab und authentifizieren Sie sich erneut, indem Sie die IAM-Rolle für das AWS-Konto dieser Organisation verwenden.

Ratenbegrenzungen und Kontingente

Die Claude-Plattform auf AWS weist Tier-1-Ratenlimits zu, wenn Sie ein Abonnement abschließen. Anthropic verwaltet Ratenlimits direkt, nicht über AWS-Quotensysteme.

Standardlimits

Die Claude-Plattform auf AWS verwendet den Standard-Tierplan von Anthropic, der mit der Claude-API von Erstanbietern identisch ist. Tier-1-Limits gelten pro Workspace. Die Grenzwerte werden nach Modellfamilien zusammengefasst. Beispielsweise teilen sich alle Opus-Modelle ein kombiniertes Limit, alle Sonnet-Modelle haben ein anderes und alle Haiku-Modelle teilen sich ein drittes.

Aktuelle Tier-1-Werte (RPM, ITPM, OTPM) und Schwellenwerte höherer Stufen finden Sie auf der Anthropic-Dokumentationswebsite unter Ratenbegrenzungen. Die Seite Anthropic ist die Quelle der Wahrheit und wird aktualisiert, wenn sich die Grenzwerte ändern.

Überschriften zur Ratenbegrenzung

Jede Antwort enthält Kopfzeilen, die Ihren aktuellen Status der Ratenbegrenzung angeben. Wichtige Überschriften:

- `anthropic-ratelimit-requests-limit`— Maximale Anzahl von Anfragen pro Minute
- `anthropic-ratelimit-requests-remaining`— Anfragen, die im aktuellen Fenster verbleiben
- `anthropic-ratelimit-requests-reset`— Zeitpunkt, zu dem das Anforderungslimit zurückgesetzt wird (RFC 3339)
- `anthropic-ratelimit-tokens-limit`— Maximale kombinierte Tokens (Eingabe + Ausgabe) pro Minute
- `anthropic-ratelimit-tokens-remaining`— Kombinierte Token, die im aktuellen Fenster verbleiben
- `anthropic-ratelimit-tokens-reset`— Zeitpunkt, zu dem das kombinierte Token-Limit zurückgesetzt wird (RFC 3339)
- `anthropic-ratelimit-input-tokens-limit`//`-remaining`— Header `-reset` Input-token-specific

- `anthropic-ratelimit-output-tokens-limit/-remaining/-reset`— Überschriften Output-token-specific
- `retry-after`— Bei einer 429-Antwort die Anzahl der Sekunden, die gewartet werden muss, bevor es erneut versucht wird

Den vollständigen Satz finden Sie in den Antwort-Headern auf der Anthropic-Dokumentationswebsite.

Höhere Grenzwerte anfordern

Im Gegensatz zur Claude-API von Erstanbietern gilt die automatische Erhöhung der Stufe nicht für die Claude-Plattform auf AWS. Um höhere Limits zu beantragen, wenden Sie sich mit Ihrer Workspace-ID und dem gewünschten Durchsatz an Ihren Anthropic-Kundenbetreuer. Die Schwellenwerte für Stufen und weitere Informationen findest du unter Ratenlimits auf der Anthropic-Dokumentationswebsite.

Fehler bei der Ratenbegrenzung

Wenn Sie ein Ratenlimit überschreiten, gibt die API HTTP 429 mit einem `rate_limit_error` Typ zurück. Implementieren Sie exponentiellen Backoff mit Jitter in Ihrer Wiederholungslogik. Der `retry-after` Header gibt an, wie viele Sekunden gewartet werden muss, bevor es erneut versucht wird.

Fakturierung

Claude Platform auf AWS verwendet AWS Marketplace als Abrechnungs-Backend für die verbrauchsabhängige Messung. Anthropic misst die Nutzung stündlich über AWS Marketplace, und AWS stellt monatliche Rechnungen auf Ihre bestehende AWS-Rechnung aus.

Die Abrechnung erfolgt nur rückwirkend (Sie zahlen für das, was Sie genutzt haben) und vor Steuern (AWS wendet die Steuereinstellungen Ihres Kontos an).

Die Nutzung wird in Claude-Verbrauchseinheiten (CCU) zu 0,01 USD pro CCU angegeben. Der CCU-Preis ist fest und wird nicht rabattiert. Anthropic bewertet Ihre Token-Nutzung in USD zu Standardpreisen pro Modell und Funktion, wendet alle ausgehandelten discount an und rechnet das Ergebnis dann in CCU zu 0,01 USD pro CCU um. Rabatte führen dazu, dass weniger CCUs gemessen werden, und nicht zu einem niedrigeren CCU-Preis. Bei CCUs handelt es sich nicht um Prepaid-Credits; es gibt kein Guthaben oder eine Verpflichtung für die CCUs. Die CCU-Definition und die Token-Tarife pro Modell finden Sie unter Preise.

Überwachung und Protokollierung

AWS CloudTrail kann alle Anfragen an Claude Platform auf AWS erfassen. Workspace- und Tresorvorgänge werden standardmäßig als Management-Ereignisse protokolliert. Alle anderen Operationen (Inferenz, Batch, Datei, Fähigkeit, Modell, Benutzerprofil und Claude Managed Agents außer Tresoren) sind Datenereignisse. Ihre Protokollierung erfordert eine ausdrückliche Konfiguration und verursacht zusätzliche Gebühren. CloudTrail Die vollständige Klassifizierung der Ereignistypen finden Sie in den IAM-Aktionen und in der CloudTrail AWS-Dokumentation finden Sie weitere Informationen zur Konfiguration.

Wählen Sie bei der Konfiguration der CloudTrail Datenereignisprotokollierung den Ressourcentyp `AWS::AWSExternalAnthropic::Workspace` aus. Dies ist der CloudTrail Ressourcentyp für Claude Platform auf AWS-Workspaces und ist erforderlich, um Ereignisse auf Datenebene (Inferenz-, Batch-, Datei- und andere Workspace-Operationen) zu erfassen. Richten Sie die Auswahl von Datenereignissen auf bestimmte Workspace-ARNs ein, wenn Sie Aktivitäten für eine Teilmenge von Workspaces und nicht für das gesamte Konto protokollieren möchten.

IDs anfordern

Jede Antwort enthält zwei Anfrage-IDs in den Antwort-Headern:

- AWS-Anforderungs-ID (**`x-amzn-requestid`**): Die primäre ID, indexiert in CloudTrail. Verwenden Sie dies, wenn Sie Anfragen über AWS-Tools untersuchen oder den AWS-Support kontaktieren.
- Anthropische Anfrage-ID (**`request-id`**): Die sekundäre ID. Verwenden Sie diese, wenn Sie den Support von Anthropic kontaktieren.

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

Das Ruby-SDK stellt derzeit keinen Raw-Response-Accessor zur Verfügung, sodass Anfrage-IDs nicht aus den Antwort-Headern in Ruby gelesen werden können. Wenn Sie eine Anforderungs-ID-Protokollierung benötigen, verwenden Sie eine der anderen oben genannten Sprachen oder erfassen Sie die IDs auf der HTTP-Proxyebene.

Anthropic empfiehlt, Ihre Aktivitäten mindestens 30 Tage fortlaufend zu protokollieren, um die Nutzungsmuster zu verstehen und mögliche Probleme zu untersuchen.

> **Note**
>
> AWS CloudTrail ist in Ihrem AWS-Konto konfiguriert. Durch die Aktivierung der Protokollierung erhalten AWS oder Anthropic keinen Zugriff auf Ihre Inhalte, der über das hinausgeht, was für die Abrechnung und den Servicebetrieb erforderlich ist.

Nutzungsmetriken und Dashboards

Claude Platform on AWS veröffentlicht keine Nutzungsmetriken pro Workspace auf Amazon CloudWatch. Die Token-Anzahl, das Anforderungsvolumen und die Nutzung pro Modell sind stattdessen über die Nutzungsoberflächen von Anthropic verfügbar:

- Nutzungsansichten der Claude Console — Wenn sich ein Principal mit der Claude Console verbindet `aws-external-anthropic:AssumeConsole` (siehe IAM-Richtlinien), zeigen die Nutzungs-Dashboards das Anforderungsvolumen, die Token-Anzahl und die Aufschlüsselung pro Modell für zugängliche Workspaces. Account-wide Nutzungsansichten erfordern Funktionen der Admin-Konsole.
- Anthropische Nutzungs- und Kosten-API — Informationen zum programmatischen Zugriff auf Nutzungs- und Kostendaten, einschließlich der Aggregation pro Workspace und pro Modell, finden Sie unter Usage and Cost API in der Anthropic-Dokumentation.

Verwenden Sie für die Betriebsüberwachung (Fehlerraten, Latenz, Ratenlimit-Header) die bei jeder Anfrage zurückgegebenen Antwort-Header (siehe Ratenlimits und Kontingente) zusammen mit den erfassten AWS-Anforderungs-IDs. CloudTrail

Kostenzuweisung und Überwachung

Die Gebühren für die Claude-Plattform auf AWS werden auf Ihrer AWS-Rechnung unter dem Service Claude Platform on AWS aufgeführt, mit Nutzungsdimensionen pro Modell. Verwenden Sie den AWS-Kosten- und Nutzungsbericht (CUR) in Kombination mit Workspace-Tags für eine detaillierte Kostenzuweisung:

1. Kennzeichnen Sie Ihre Workspaces mit den Zuordnungsdimensionen, die Ihnen wichtig sind (Team, Anwendung, Umgebung, Kostenstelle). Einzelheiten zur Kennzeichnung finden Sie unter Arbeitsbereiche.

2. Aktivieren Sie in der AWS-Abrechnungskonsole jeden Tag-Schlüssel als Kostenzuweisungs-Tag. Nach der Aktivierung wird die Nutzung, die am oder nach dem Aktivierungsdatum erfolgt ist, in der CUR nach Tags aufgeteilt.
3. Gruppieren Sie in CUR oder AWS Cost Explorer nach der `resourceTags/user:<tag-key>` Spalte, um die Ausgaben nach Workspace-Tag aufzuschlüsseln.

Workspace-Tags fließen in CUR-Einzelposten für die gesamte Nutzung der Claude-Plattform auf AWS ein. Dieselben Tag-Schlüssel steuern sowohl die IAM-based Zugriffskontrolle als auch die Kostenzuweisung. Informationen zu Einrichtungsschritten und Verzögerungen bei der Aktivierung finden Sie in den Tags zur Kostenzuweisung in der AWS-Rechnungsdokumentation.

Migration von Amazon Bedrock

Wenn Sie derzeit Claude auf Bedrock verwenden, erfordert die Migration zu Claude Platform auf AWS Änderungen während Ihrer gesamten Integration. Die Sigv4-Signatur wird weiterhin unterstützt, aber der Signaturkontext, die Basis-URL, das API-Format, die Modell-IDs, der SDK-Client und das Paket, das Streaming-Format, die Anforderungsheader und die regionale Verfügbarkeit ändern sich. In der folgenden Tabelle sind die Unterschiede zusammengefasst.

Was ändert sich

	Amazon Bedrock	Claude-Plattform auf AWS
Basis-URL	`bedrock-runtime.{region}.amazonaws.com`	`aws-external-anthropic.{region}.api.aws`
API-Format	Grundgestein Converse/ InvokeModel	API für anthropische Nachrichten () `/v1/messages`
Modell-IDs	`anthropic.claude-opus-4-7-v1`	`claude-opus-4-7`
SDK-Client	`AnthropicBedrock` /Bedrock SDK	Platform-specific Client (`AnthropicAWS` ,`AnthropicAws` , usw.)`AnthropicAwsClient` , in der Betaversion
SDK-Paket	`anthropic[bedrock]` `@anthropic-ai/bedrock-sdk`usw.	`anthropic[aws]` `@anthropic-ai/aws-sdk` , usw. (siehe Installieren eines SDK)
SigV4-Dienstname	`bedrock`	`aws-external-anthropic`
Streaming-Format	AWS EventStream	SSE (wie Claude API)

	Amazon Bedrock	Claude-Plattform auf AWS
Workspace-Header	Nicht zutreffend	`anthropic-workspace-id` erforderlich
Verfügbarkeit in Regionen	Mehrere Handelsregionen	Alle kommerziellen AWS-Regionen (Opt-in-Regionen erfordern Konto-Opt-In)

Was gewinnen Sie

- In der Regel am selben Tag Zugriff auf neue Modelle und Funktionen, ohne dass ein separater Schritt zur Partnerintegration erforderlich ist
- Fähigkeiten von Mitarbeitern für die Dokumentgenerierung (PowerPoint, Excel, Word, PDF)
- Codeausführung in der verwalteten Sandbox von Anthropic
- Beta-Funktionen über den `anthropic-beta` Header (siehe Funktionen, die derzeit nicht verfügbar sind)
- Claude Console für die Sichtbarkeit von Kontingenten und Nutzungsanalysen
- Direkte anthropische Unterstützung
- API-Schlüsselauthentifizierung als Alternative zu Sigv4 (siehe Authentifizierung)

Was bleibt gleich

- AWS IAM-Authentifizierung (SigV4)
- Abrechnung über Ihr AWS-Konto
- AWS-Verpflichtung, Ruhestand

Fallstricke bei der Migration

> **Warning**
>
> Aktivieren Sie zuerst den ausgehenden Web-Identitätsverbund. Wenn Ihr AWS-Konto die Claude-Plattform noch nicht auf AWS verwendet hat, müssen Sie den ausgehenden Web-

Identitätsverbund einmal pro Konto aktivieren, bevor Sie Anfragen stellen können. Ohne diesen Schritt schlagen alle Anfragen mit einem Verbundfehler fehl. Dieser Schritt ist für Bedrock nicht erforderlich.

⚠ Warning

Zero Data Retention (ZDR) ist auf der Claude-Plattform auf AWS optional. Auf Bedrock ist AWS der Datenverarbeiter und Anthropic speichert keine Inferenzeingaben oder -ausgaben; das ZDR-Programm von Anthropic gilt dort nicht. Auf der Claude-Plattform auf AWS ist Anthropic der Datenverarbeiter, und ZDR folgt dem Claude-API-Modell von Erstanbietern: Es ist auf Anfrage über Ihren Anthropic-Kundenbetreuer erhältlich. Bestätigen Sie die ZDR-Registrierung, bevor Sie Produktionsworkloads migrieren, für die Datenaufbewahrungsgarantien erforderlich sind.

Kommerzielle Überlegungen

- Nutzungsbedingungen: Die AWS-Servicebedingungen regeln die Claude-Plattform auf AWS. Die kommerziellen Nutzungsbedingungen, der Zusatz zur Datenverarbeitung und die Nutzungsrichtlinie von Anthropic gelten ebenfalls.
- Rabatte und private Angebote: Ausgehandelte Rabatte und private Angebote von AWS Marketplace werden nicht automatisch zwischen Bedrock und Claude Platform auf AWS übertragen. Arbeiten Sie mit Ihrem Anthropic-Kundenbetreuer zusammen, um die Handelsbedingungen für die Claude-Plattform auf AWS festzulegen.

IAM-Richtlinien

Die Claude-Plattform auf AWS lässt sich für die Zugriffskontrolle in AWS IAM integrieren. Sie gewähren oder verweigern den Zugriff auf bestimmte API-Aktionen in bestimmten Workspaces mithilfe der standardmäßigen IAM-Richtliniensyntax.

Der SigV4-Dienstname und der IAM-Aktionsnamespace sind. `aws-external-anthropic` Aktionen folgen dem Muster `aws-external-anthropic:<Action>` (z. B.). `aws-external-anthropic:CreateInference`

Beispiel: Batch-Inferenz verweigern

Die folgende Richtlinie ermöglicht Inferenzen in Echtzeit und blockiert gleichzeitig die Stapelverarbeitung, eine häufige Anforderung für ZDR-sensitive Workloads:

```
{
  "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": "*"
    }
  ]
```

```
}
```

Die `GetBatchInference` Aktion autorisiert sowohl die Batch-Metadaten-Route als auch die Batch-Ergebnisroute. Wenn sie verweigert wird, werden außerdem sowohl Lesevorgänge als auch die Batch-Aufzählung blockiert. `ListBatchInferences`

Die `Allow` Anweisung zählt bestimmte `List*` Aktionen auf, anstatt `Get*` Platzhalter zu verwenden. Platzhalter würden Lesevorgänge gewähren `GetFile` (wodurch Datei-Bytes heruntergeladen werden) und andere Lesevorgänge, die Sie vielleicht nicht beabsichtigen, `Deny` `Allow` überschreiben, aber die explizite Form ist das sicherere Muster, das modelliert werden kann.

Beispiel: synchrone Inferenz für einen einzelnen Workspace

Erteilt die Mindestberechtigungen für einen IAM-Prinzipal, der Inferenzen für einen Produktions-Workspace ausführt:

```
{
  "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**
>
> Der `List*` Platzhalter in dieser Richtlinie entspricht ebenfalls`ListWorkspaces`, was sich auf den Kontobereich bezieht. Die ARN-Einschränkung für den Workspace filtert es im Hintergrund heraus, sodass diese Richtlinie das Auflisten von Workspaces nicht autorisiert. Wenn Ihr Dienstkonto Workspaces auflisten muss, fügen Sie eine separate Anweisung für with hinzu. `Allow` `ListWorkspaces` `Resource: "*"`

Diese Richtlinie setzt die AWS-SigV4-Authentifizierung voraus. Wenn sich der Principal mit einem API-Schlüssel authentifiziert, gewähren Sie diesen ebenfalls `aws-external-anthropic:CallWithBearerToken` (siehe Authentifizierung).

Beispiel: Workspace-Isolierung pro Kunde

Beschränkt eine Rolle auf einen einzelnen Workspace:

```
{
  "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

Der `aws-external-anthropic:*` Platzhalter in der ersten Anweisung beinhaltet kontobezogene Aktionen (`CreateWorkspace`,`ListWorkspaces`), die die Workspace-ARN-Einschränkung im Hintergrund herausfiltert. Dies entspricht der Absicht der „Isolierung“ — die Rolle kann keine Arbeitsbereiche erstellen oder aufzählen —, aber die Richtlinie enthält Berechtigungen, die keine Wirkung haben. Informationen zum Muster für den Kontobereich finden Sie unter Automatisierung der Bereitstellung.

Die zweite Anweisung gewährt `CallWithBearerToken` und für alle Ressourcen, da es sich `AssumeConsole` bei beiden um routenlose Aktionen handelt, die nicht an einen Workspace-

ARN gebunden sind. Lassen Sie die zweite Anweisung weg, wenn die Rolle nur Sigv4 verwendet und niemals eine Verbindung zur Claude-Konsole herstellt.

Beispiel: Sperrung von Funktionen für einen Workspace ZDR-sensitive

Blockiert die Stapelverarbeitung und das Hochladen von Dateien in einem bestimmten Workspace, während synchrone Inferenz verfügbar bleibt. Nützlich, wenn ein Workspace ZDR-Daten (Zero Data Retention) verarbeitet, die nicht serverseitig gespeichert werden dürfen. Fügen Sie diese Richtlinie zusammen mit einer Zulassungsrichtlinie wie `AnthropicLimitedAccess` oder dem obigen Beispiel für einen einzelnen Arbeitsbereich hinzu. Eine Deny-only Richtlinie allein gewährt keine Berechtigungen:

```
{
  "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

Diese Ablehnung blockiert nur die Erstellung. Andere Datei- und Batch-Aktionen werden nur verweigert, wenn Sie sie ebenfalls auflisten. Für einen vollständigen Lockdown, bei dem der Workspace niemals Dateien oder Stapel speichern darf, sollten Sie auch`aws-external-anthropic:GetFile`,`aws-external-anthropic:ListFiles`,`aws-external-anthropic:DeleteFile`,`aws-external-anthropic:GetBatchInference`, `aws-external-anthropic:ListBatchInferencesaws-external-`

`anthropic:CancelBatchInference`, und `aws-external-anthropic:DeleteBatchInference` verweigern.

Beispiel: Automatisierung der Bereitstellung

Erteilt einer CI/CD Rolle die Aktionen, die zum Erstellen und Verwalten von Workspaces erforderlich sind, ohne dass irgendwelche Inferenzberechtigungen erforderlich sind:

```
{
  "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`und `ListWorkspaces` sind kontobezogene Operationen. Die Angabe eines Workspace-ARN für diese Aktionen hat keine Auswirkung; verwenden Sie`Resource: "*"`.

Verwaltete Richtlinien

AWS bietet verwaltete Richtlinien für gängige Zugriffsmuster:

- **`AnthropicFullAccess`**: `aws-external-anthropic:*` Zuschüsse für alle Ressourcen.
- **`AnthropicReadOnlyAccess`**: `Get*` Zuschüsse `List*` und `CallWithBearerToken` für alle Ressourcen.
- **`AnthropicInferenceAccess`**: Gewährt die ReadOnly Aktionen sowie die Inferenzaktionen (`CreateInference`,`CreateBatchInference`,`CancelBatchInference`,`DeleteBatchInference`,C für alle Ressourcen.

- **`AnthropicLimitedAccess`**: Gewährt die `AnthropicInferenceAccess` Aktionen sowie alle Aktionen von Claude Managed Agents (Agenten, Sitzungen, Umgebungen, Tresore, Speicherspeicher) für alle Ressourcen.
- **`AnthropicSelfHostedEnvironmentAccess`**: Erteilt die Berechtigungen, die ein selbst gehosteter Sandbox-Worker benötigt, um Arbeitselemente der Umgebung abzufragen und zu verarbeiten, die zugehörigen Umgebungs-, Sitzungs- und Qualifikationsressourcen zu lesen und den Sitzungsstatus zu aktualisieren. Ordnen Sie diese Richtlinie der IAM-Rolle zu, die Ihr selbst gehosteter Umgebungsmitarbeiter annimmt. `AnthropicSelfHostedEnvironmentAccess`ist der empfohlene Standard für eine einzelne Rolle. Wenn Sie den Work Poller und die Sandbox pro Sitzung unter separaten IAM-Prinzipalen ausführen, können Sie diese Berechtigungen auf zwei engere benutzerdefinierte Richtlinien für die geringste Anzahl von Rechten aufteilen.

`AnthropicInferenceAccess`ist die engste verwaltete Richtlinie ausreichend, um Inferenzen durchzuführen. Über die `List*` Platzhalter `Get*` und gewährt sie Lesezugriff auf jede API-Ressource im Namespace, einschließlich des Herunterladens von Dateiinhalten `GetFile` und des Speicherinhalts durch. `GetMemoryStore` Das Erstellen oder Löschen von Dateien oder Fertigkeiten, die Verwaltung von Benutzerprofilen, die Änderung von Arbeitsbereichen oder ein Konsolenverbund sind nicht möglich.

> **Note**
>
> `AnthropicReadOnlyAccessAnthropicInferenceAccess`, und gewähren `AnthropicLimitedAccess` nicht`AssumeConsole`. Principals, die sich mit der Claude Console verbinden müssen, benötigen einen separaten Zuschuss `aws-external-anthropic:AssumeConsole` — entweder durch eine `AnthropicFullAccess` oder eine benutzerdefinierte Richtlinie. Weitere Informationen finden Sie unter Verbindung mit der Claude Console.

> **Note**
>
> `CreateInference`und `CreateBatchInference` sind separate Aktionen. Das eine zu leugnen blockiert das andere nicht. Wenn Sie alle Modellanrufe verhindern möchten, lehnen Sie beide ab.

Verbindung zur Claude Console herstellen

`aws-external-anthropic:AssumeConsole`ermöglicht es einem IAM-Prinzipal, sich mit der Claude-Konsole zu verbinden Anthropic-operated . Der Zugriff innerhalb der Konsole wird bei den meisten Vorgängen immer noch von IAM gesteuert, aber ein Teil der Administratorvorgänge — hauptsächlich Nutzungsansichten, für die es keine entsprechende IAM-API gibt — ist auf die Funktion beschränkt, mit der der Principal verbunden ist.

Es gibt zwei Funktionen:

- **`developer`**— ermöglicht die Operationen, die ein Entwickler der Claude-Plattform auf AWS für die tägliche Arbeit benötigt: Inferenz von der Konsole aus ausführen, Workspace-Daten lesen, persönliche Nutzung einsehen.
- **`admin`**— ermöglicht zusätzlich Konsolenoperationen nur für Administratoren, einschließlich kontoübergreifender Nutzungsansichten und administrativer Einstellungen, die nicht durch IAM-Aktionen angezeigt werden.

Steuern Sie mit dem Bedingungsschlüssel für die Aktion, welche Funktion ein Principal anfordern kann: `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`Zuschüsse `AssumeConsole` ohne Kapazitätsbeschränkung. Für jede engere Gewährung — Konsolenzugriff nur für Entwickler oder Zugriff nur für Administratoren —

fügen Sie eine benutzerdefinierte Richtlinie hinzu, die die oben dargestellte Bedingung abdeckt. `AssumeConsole` `aws-external-anthropic:Capability`

Anrufe mit Inhaber-Tokens

`aws-external-anthropic:CallWithBearerToken`autorisiert den SigV4-free Anforderungspfad, der verwendet wird, wenn ein API-Schlüssel als Bearer-Token präsentiert wird. Jeder Principal, der sich mit einem API-Schlüssel authentifiziert, benötigt diese Aktion im Ziel-Workspace. Dies gilt unabhängig davon, ob Sie den `Authorization: Bearer` Header direkt verwenden `ANTHROPIC_AWS_API_KEY` oder festlegen.

Dies ist für API-Schlüsselaufrufer zusätzlich zu den Inferenzaktionen (`CreateInferenceCreateBatchInference`, usw.) erforderlich. `CallWithBearerToken`Andernfalls werden API-Schlüsselanfragen abgelehnt, bevor die Prüfung der Inferenzautorisierung abgeschlossen ist. SigV4-Anrufer benötigen diese Aktion nicht.

`AnthropicReadOnlyAccess`, `AnthropicInferenceAccessAnthropicLimitedAccess`, und `AnthropicFullAccess` alle beinhalten. `CallWithBearerToken` Wenn Sie eine benutzerdefinierte Richtlinie für den API-Schlüsselzugriff schreiben, fügen Sie sie explizit hinzu.

Verbund ausgehender Web-Identitätsverbund (für Konsolenzugriff erforderlich)

Die Claude Console läuft in der anthropischen Infrastruktur, nicht in AWS. Wenn ein IAM-Principal aufruft`AssumeConsole`, gibt AWS STS ein Web-Identity-Token aus, das für die anthropische Zielgruppe bestimmt ist. Die Anthropic-Konsole akzeptiert dann dieses Token und richtet die Verbundsitzung ein. Damit dies funktioniert, muss das AWS-Konto den ausgehenden Web-Identitätsverbund für die `aws-external-anthropic` Zielgruppe zulassen.

Principals, die anrufen, `AssumeConsole` benötigen zusätzlich zur Aktion die folgenden STS-Berechtigungen: `aws-external-anthropic:AssumeConsole`

- **`sts:GetWebIdentityToken`**— ermöglicht die Ausgabe des Web-Identitätstokens, das von der Anthropic-Konsole verwendet wird.
- **`sts:TagGetWebIdentityToken`**— ermöglicht das Anhängen von Sitzungs-Tags an das Web-Identity-Token. Claude Platform auf AWS verwendet diese Tags, um der Konsole die Fähigkeiten und den Workspace-Kontext des Principals zu vermitteln.

Nehmen Sie beides in die Vertrauensrichtlinie jeder Rolle auf, die Konsolenbenutzer übernehmen, oder in die inline/managed Richtlinie für Benutzeridentitäten, die `AssumeConsole` direkt aufrufen. `AnthropicFullAccess`schließt beide STS-Aktionen ein. Umgebungen, die `sts:*` auf SCP- oder Berechtigungsgrenzebene verweigern, müssen diese beiden Aktionen ausdrücklich zulassen, damit der Konsolenverbund erfolgreich ist.

IAM-Aktionen für Claude Platform auf AWS

IAM-Aktionsreferenz für die Steuerung des Zugriffs auf die Claude-Plattform auf AWS über AWS-Richtlinien.

Claude Platform auf AWS verwendet AWS IAM für die Zugriffskontrolle. Jede API-Route ist einer IAM-Aktion im Namespace zugeordnet. `aws-external-anthropic` Auf dieser Seite werden alle Aktionen, die Routen, die jede Aktion autorisiert, und die verwalteten Richtlinien aufgeführt, die für allgemeine Zugriffsmuster verfügbar sind. Informationen zur Plattformeinrichtung und Authentifizierung finden Sie unter Was ist Claude Platform on AWS? .

Service-Details

IAM-Dienstpräfix	`aws-external-anthropic`
Ressourcentypen	`workspace`
Workspace-ARN	`arn:aws:aws-external-anthropic:{region}:{account-id}:workspace/{workspace-id}`

Die ARN-Region ist immer gefüllt und entspricht der Region, an die der Workspace gebunden ist. Das Ressourcensegment ist die markierte Workspace-ID (`wrkspc_…`), derselbe Wert, den Sie in der `anthropic-workspace-id` Kopfzeile übergeben.

Aktionen

Der Service definiert 65 Aktionen in 15 Gruppen. Aktionen folgen der `VerbNoun` AWS-Konvention und verwenden Verbdisziplin, `Get*` sodass `List*` Platzhalter eine saubere, schreibgeschützte Grenze bilden.

Inferenz

Action	Autorisierte Routen
`CreateInference`	`POST /v1/messages`

Action	Autorisierte Routen
CountTokens	POST /v1/messages/count_tokens

Batch-Verarbeitung

Action	Autorisierte Strecken
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}

Modelle

Action	Autorisierte Strecken
GetModel	GET /v1/models/{id}
ListModels	GET /v1/models

Dateien

Action	Autorisierte Strecken
CreateFile	POST /v1/files
GetFile	GET /v1/files/{id} GET /v1/files/{id}/content
ListFiles	GET /v1/files

Action	Autorisierte Strecken
`DeleteFile`	`DELETE /v1/files/{id}`

> **Note**
>
> `GetFile`autorisiert sowohl das Herunterladen von Metadaten als auch von Inhalten. Ein Principal mit schreibgeschütztem Zugriff kann Dateibytes herunterladen und nicht nur Dateien auflisten.

Skills

Action	Autorisierte Routen
`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**
>
> Das Löschen einer einzelnen Skill-Version wird zugeordnet`UpdateSkill`, nicht`DeleteSkill`. Eine Richtlinie, die dies ablehnt, erlaubt `aws-external-anthropic:Delete*` immer noch das Löschen von Versionen. Verweigern Sie dies

`UpdateSkill` und `CreateSkill` auch, wenn Sie jegliche Änderung Ihrer Fertigkeiten verhindern möchten.

Benutzerprofile

Action	Autorisierte Strecken
`CreateUserProfile`	`POST /v1/user_profiles`
`GetUserProfile`	`GET /v1/user_profiles/{id}`
`ListUserProfiles`	`GET /v1/user_profiles`
`UpdateUserProfile`	`POST /v1/user_profiles/{id}`

Warning

Beim IAM-Aktionsabgleich wird nicht zwischen Groß- und Kleinschreibung unterschieden. Der Platzhalter `aws-external-anthropic:*File` entsprichtCreateFile, und `GetFileDeleteFile`, entspricht aber nicht `ListFiles` (was auf „Dateien“ und nicht auf „Datei“ endet). Es entspricht auch zu vielCreateUserProfile, und `GetUserProfile` zwar deshalb, `UpdateUserProfile` weil „Profil“ auf „Datei“ endet. Wenn Sie nur Datei-API-Aktionen gewähren oder verweigern möchten, zählen Sie sie explizit auf (,,,DeleteFile) `CreateFile GetFileListFiles`, anstatt ein `*File` Suffixmuster zu verwenden.

Kundendienstmitarbeiters (Kundendienstmitarbeiter)

Agentendefinitionen für Claude Managed Agents.

Action	Autorisierte Routen
`CreateAgent`	Der Agent erstellt Routen
`GetAgent`	Der Agent liest Routen

Action	Autorisierte Routen
`ListAgents`	Die Routen der Agenten auflisten
`UpdateAgent`	Routen für Agenten-Updates
`ArchiveAgent`	Routen für Agenten archivieren

Sitzungen

Agentensitzungen und Ereignisverlauf.

Action	Autorisierte Routen
`CreateSession`	Sitzung: Routen erstellen
`GetSession`	Leserouten der Sitzung
`ListSessions`	Routen in der Sitzungsliste
`UpdateSession`	Routen zur Sitzungsaktualisierung
`ArchiveSession`	Routen zum Archivieren von Sitzungen
`DeleteSession`	Routen aus der Sitzung löschen

Umgebungen

Cloud-Container-Konfigurationen für Sitzungen.

Action	Autorisierte Routen
`CreateEnvironment`	Umgebung, Routen erstellen
`GetEnvironment`	Umgebung liest Routen
`ListEnvironments`	Umgebung listet Routen auf

Action	Autorisierte Routen
`UpdateEnvironment`	Routen zur Aktualisierung der Umgebung
`ArchiveEnvironment`	Routen für die Umgebung archivieren
`DeleteEnvironment`	Umgebung: Routen löschen
`ProcessEnvironmentWork`	Self-hosted Routen für Umweltarbeiter

Tresore

Tresore für Anmeldeinformationen für die Sitzungsauthentifizierung.

Action	Autorisierte Routen
`CreateVault`	Tresor: Routen erstellen
`GetVault`	Routen im Tresor lesen
`ListVaults`	Tresor listet Routen auf
`UpdateVault`	Routen für Vault-Updates
`ArchiveVault`	Routen im Tresor archivieren
`DeleteVault`	Routen im Tresor löschen

> **Note**
>
> Tresorvorgänge werden als CloudTrail Verwaltungsereignisse (und nicht als Datenereignisse) klassifiziert, da Tresore Anmeldeinformationen enthalten und von der standardmäßigen Auditprotokollierung profitieren. Andere Vorgänge von Claude Managed Agents (Agenten, Sitzungen, Umgebungen, Speicherspeicher) sind Datenereignisse.

Speicherspeicher

Persistenter Agentenspeicher.

Action	Autorisierte Routen
`CreateMemoryStore`	Speicher speichern, Routen erstellen
`GetMemoryStore`	Speicher speichert Leserouten
`ListMemoryStores`	Im Speicher werden Routen aufgelistet
`UpdateMemoryStore`	Routen zur Aktualisierung von Speicherspeichern
`ArchiveMemoryStore`	Routen für das Speichern von Speicherarchiven
`DeleteMemoryStore`	Speicher speichern, Routen löschen

> **Note**
>
> `GetMemoryStore`liest Speicherinhalte. Der `Get*` Platzhalter in einer verwalteten oder benutzerdefinierten Richtlinie gewährt daher Lesezugriff auf den Speicher. Geltungsbereich von Richtlinien explizit, wenn der Speicherinhalt eingeschränkt werden muss.

Webhooks

Benachrichtigungen über Lifecycle-Ereignisse für Sitzungen.

Action	Autorisierte Routen
`ListWebhooks`	`GET /v1/webhooks`
`GetWebhook`	`GET /v1/webhooks/{webhook_endpoint_id}`
`CreateWebhook`	`POST /v1/webhooks`
`UpdateWebhook`	`POST /v1/webhooks/{webhook_endpoint_id}`

Action	Autorisierte Routen
`DeleteWebhook`	`DELETE /v1/webhooks/{webhook_endpoint_id}`
`RotateWebhookSecret`	`POST /v1/webhooks/{webhook_endpoint_id}/regenerate_signing_secret`

Arbeitsbereiche

Action	Autorisierte Strecken
`CreateWorkspace`	`POST /v1/organizations/workspaces`
`GetWorkspace`	`GET /v1/organizations/workspaces/{id}`
`ListWorkspaces`	`GET /v1/organizations/workspaces`
`UpdateWorkspace`	`POST /v1/organizations/workspaces/{id}`
`ArchiveWorkspace`	`POST /v1/organizations/workspaces/{id}/archive`

Bei der Registrierung wird ein Standard-Workspace bereitgestellt; siehe Workspaces.

Authentifizierung

Action	Autorisierte Routen
`CallWithBearerToken`	(Keine)

`CallWithBearerToken`ist eine Auth-Layer-Berechtigung, die einen Principal autorisiert, sich über einen API-Schlüssel (Bearer-Token) statt über AWS SigV4 zu authentifizieren. Sie ist keiner Route zugeordnet. Erteilen Sie es zusammen mit den Routenzuordnungsaktionen, die der API-Schlüsselinhaber ausführen soll.

Konsolenzugriff

Action	Autorisierte Routen
`AssumeConsole`	(Keine)

`AssumeConsole`autorisiert einen Principal, die Claude-Konsole für eine Claude-Plattform im AWS-Workspace über den AWS Console Federation Flow zu öffnen. Es ist keiner Route zugeordnet. Gewähren Sie es Principals, die auf der Serviceseite Claude Platform on AWS in der AWS-Konsole auf Claude Console öffnen können sollten. Der `aws-external-anthropic:Capability` Bedingungsschlüssel in der `AssumeConsole` Aktion steuert, welche Konsolenfunktion (Entwickler oder Administrator) ein Principal anfordern kann. Einzelheiten finden Sie in den IAM-Richtlinien und unter Verwenden der Claude-Konsole für den Anmeldevorgang.

Warning

`aws-external-anthropic:AssumeConsole`leitet eine Claude Console-Sitzung ein, die unabhängig von der Sitzungsdauer des Anrufers bis zu 12 Stunden dauert. Ein Anrufer, dessen IAM-Sitzung kürzer als 12 Stunden dauert, kann trotzdem eine Konsolensitzung erhalten, deren Quellanmeldedaten länger gültig sind. Beschränken Sie diese Berechtigung auf Prinzipale, die Zugriff auf die Claude-Konsole benötigen, und widerrufen Sie sie umgehend, wenn sie nicht mehr benötigt wird.

Note

Aktionen, die nach dem Verbund in der Claude-Konsole ausgeführt werden, werden nicht in AWS aufgezeichnet CloudTrail oder sind nicht über IAM zuzuordnen. Dazu gehören globale Aktivitäten, die sich auf den Arbeitsbereich beziehen, wie z. B. das Anzeigen von Nutzungsberichten. Wenn Sie einen Prüfpfad für Konsolenaktivitäten benötigen, verwenden Sie stattdessen die in der Claude Console verfügbaren Auditprotokolle. CloudTrail

Route-to-action Zuordnung

In der folgenden Tabelle sind alle Routen auf der Claude-Plattform auf AWS sowie die IAM-Aktion aufgeführt, die erforderlich ist, um sie aufzurufen. Die IAM-Aktion der stabilen Route autorisiert auch Anfragen, die den Header verwenden. `anthropic-beta` CloudTrail klassifiziert jede Aktion entweder als Datenereignis (umfangreiche Operationen auf Datenebene) oder als Managementereignis (Operationen auf Kontrollebene).

Methode	Route	IAM-Aktion	CloudTrail Ereignistyp
`POST`	`/v1/messages`	`CreateInference`	Daten
`POST`	`/v1/messages/count_tokens`	`CountTokens`	Daten
`POST`	`/v1/messages/batches`	`CreateBatchInference`	Daten
`GET`	`/v1/messages/batches`	`ListBatchInferences`	Daten
`GET`	`/v1/messages/batches/{id}`	`GetBatchInference`	Daten
`GET`	`/v1/messages/batches/{id}/results`	`GetBatchInference`	Daten
`POST`	`/v1/messages/batches/{id}/cancel`	`CancelBatchInference`	Daten
`DELETE`	`/v1/messages/batches/{id}`	`DeleteBatchInference`	Daten
`GET`	`/v1/models`	`ListModels`	Daten
`GET`	`/v1/models/{id}`	`GetModel`	Daten
`POST`	`/v1/files`	`CreateFile`	Daten
`GET`	`/v1/files`	`ListFiles`	Daten

Methode	Route	IAM-Aktion	CloudTrail Ereignistyp
GET	/v1/files/{id}	GetFile	Daten
GET	/v1/files/{id}/content	GetFile	Daten
DELETE	/v1/files/{id}	DeleteFile	Daten
POST	/v1/skills	CreateSkill	Daten
GET	/v1/skills	ListSkills	Daten
GET	/v1/skills/{id}	GetSkill	Daten
DELETE	/v1/skills/{id}	DeleteSkill	Daten
POST	/v1/skills/{id}/versions	UpdateSkill	Daten
GET	/v1/skills/{id}/versions	GetSkill	Daten
GET	/v1/skills/{id}/versions/{version}	GetSkill	Daten
GET	/v1/skills/{id}/versions/{version}/content	GetSkill	Daten
DELETE	/v1/skills/{id}/versions/{version}	UpdateSkill	Daten
POST	/v1/user_profiles	CreateUserProfile	Daten
GET	/v1/user_profiles	ListUserProfiles	Daten
GET	/v1/user_profiles/{id}	GetUserProfile	Daten
POST	/v1/user_profiles/{id}	UpdateUserProfile	Daten
POST	/v1/organizations/workspaces	CreateWorkspace	Verwaltung

Methode	Route	IAM-Aktion	CloudTrail Ereignistyp
`GET`	`/v1/organizations/workspaces`	`ListWorkspaces`	Verwaltung
`GET`	`/v1/organizations/workspaces/{id}`	`GetWorkspace`	Verwaltung
`POST`	`/v1/organizations/workspaces/{id}`	`UpdateWorkspace`	Verwaltung
`POST`	`/v1/organizations/workspaces/{id}/archive`	`ArchiveWorkspace`	Verwaltung

Die Routen von Claude Managed Agents (Agenten, Sitzungen, Umgebungen, Tresore, Speicherspeicher) folgen demselben Route-zu-Action-Muster. Die vollständige Zuordnung pro Route finden Sie in der Referenz zu Anthropic IAM-Aktionen. Tresor-Routen sind Management-Ereignisse; Agenten-, Sitzungs-, Umgebungs- und Speicherspeicherrouten sind Datenereignisse.

Routen, die sich nicht auf der Claude-Plattform auf AWS befinden, werden am Gateway standardmäßig verweigert.

Weitere Informationen finden Sie auch unter

- IAM-Richtlinien, zum Beispiel Richtlinien und verwaltete Richtlinien
- AWS IAM-Benutzerhandbuch für IAM-Richtliniensyntax und Bewertungslogik
- CloudTrail AWS-Benutzerhandbuch für die Konfiguration der Audit-Protokollierung

Dokumentverlauf

In der folgenden Tabelle werden wichtige Änderungen am Claude Platform on AWS-Benutzerhandbuch beschrieben.

Änderungen	Beschreibung	Date
Erste Veröffentlichung	Das Benutzerhandbuch für die Claude-Pl attform auf AWS wurde veröffentlicht.	7. Mai 2026

Die vorliegende Übersetzung wurde maschinell erstellt. Im Falle eines Konflikts oder eines Widerspruchs zwischen dieser übersetzten Fassung und der englischen Fassung (einschließlich infolge von Verzögerungen bei der Übersetzung) ist die englische Fassung maßgeblich.