



AWS Well-Architected Framework

# Säule der Leistungseffizienz



# Säule der Leistungseffizienz: AWS Well-Architected Framework

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

Die Handelsmarken und Handelsaufmachung von Amazon dürfen nicht in einer Weise in Verbindung mit nicht von Amazon stammenden Produkten oder Services verwendet werden, durch die Kunden irregeführt werden könnten oder Amazon in schlechtem Licht dargestellt oder diskreditiert werden könnte. 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

Zusammenfassung und Einführung .....	1
Einführung .....	1
Leistungseffizienz .....	3
Designprinzipien .....	3
Definition .....	4
Auswahl der Architektur .....	5
PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features .....	5
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF01-BP02 Einholen von Rat beim Cloud-Anbieter oder einem geeigneten Partner, um mehr über Architekturmuster und bewährte Methoden zu erfahren .....	8
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF01-BP03 Berücksichtigen Sie die Kosten bei architektonischen Entscheidungen .....	10
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF01-BP04 Evaluieren, wie sich Kompromisse auf Kunden und Architektureffizienz auswirken .....	12
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF01-BP05 Verwenden Sie Richtlinien und Referenzarchitekturen .....	14
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF01-BP06 Verwenden von Benchmarking, um architektonische Entscheidungen zu treffen .....	16
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF01-BP07 Verwenden eines datengesteuerten Ansatzes für architektonische Entscheidungen .....	19
Implementierungsleitfaden .....	6
Ressourcen .....	7
Computer und Hardware .....	22
PERF02-BP01 Auswählen der besten Datenverarbeitungsoptionen für die Workload .....	22
Implementierungsleitfaden .....	6

Implementierungsschritte .....	6
Ressourcen .....	7
PERF02-BP02 Verstehen verfügbarer Konfigurationen und Features für die Datenverarbeitung .....	26
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7
PERF02-BP03 Sammeln Sie rechnerbezogene Metriken .....	30
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7
PERF02-BP04 Konfigurieren und richtiges Dimensionieren von Datenverarbeitungsressourcen .....	33
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF02-BP05 Dynamisches Skalieren von Datenverarbeitungsressourcen .....	35
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF02-BP06 Verwenden von optimierten hardwarebasierten Datenverarbeitungsbeschleunigern .....	39
Implementierungsleitfaden .....	6
Ressourcen .....	7
Datenverwaltung .....	42
PERF03-BP01 Verwenden eines speziell entwickelten Datenspeichers, der die Datenzugriffs- und Speicheranforderungen am besten unterstützt .....	42
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF03-BP02 Bewerten verfügbarer Konfigurationsoptionen für den Datenspeicher .....	55
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF03-BP03 Erfassen und Aufzeichnen von Metriken zur Datenspeicherleistung .....	60
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7
PERF03-BP04 Implementieren von Strategien zur Verbesserung der Abfrageleistung im Datenspeicher .....	63

Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF03-BP05 Implementieren von Datenzugriffsmustern, die Caching nutzen .....	65
Implementierungsleitfaden .....	6
Ressourcen .....	7
Netzwerk und Bereitstellung von Inhalten .....	70
PERF04-BP01 Verstehen der Auswirkungen des Netzwerks auf die Leistung .....	71
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP02 Evaluieren verfügbarer Netzwerk-Features .....	74
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP03 Auswählen von entsprechend dedizierter Konnektivität oder VPN für Ihre Workload .....	81
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP04 Lastausgleich verwenden, um den Datenverkehr auf mehrere Ressourcen zu verteilen .....	84
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP05 Auswählen leistungsfördernder Netzwerkprotokolle .....	89
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP06 Wählen Sie den Standort Ihres Workloads basierend auf den Netzwerkanforderungen .....	92
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP07 Optimieren der Netzwerkkonfiguration basierend auf Metriken .....	98
Implementierungsleitfaden .....	6
Ressourcen .....	7
Prozess und Kultur .....	104
PERF05-BP01 Festlegen wichtiger Leistungskennzahlen (KPIs) zum Messen des Zustands und der Leistung der Workload .....	106
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7

PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren .....	109
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP03 Definieren eines Prozesses zum Verbessern der Workload-Leistung .....	112
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP04 Testen Sie Ihre Arbeitslast .....	114
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP05 Verwenden von Automatisierung zur proaktiven Behebung leistungsbezogener Probleme .....	116
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP06 Behalten Sie Ihren Workload und Ihre Services up-to-date .....	119
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7
PERF05-BP07 Regelmäßiges Überprüfen von Metriken .....	121
Implementierungsleitfaden .....	6
Ressourcen .....	7
Schlussfolgerung .....	124
Mitwirkende .....	125
Weitere Informationen .....	126
Dokumentversionen .....	127
Hinweise .....	129
AWS Glossar .....	130

# Säule „Leistungseffizienz“ – AWS-Well-Architected-Framework

Veröffentlichungsdatum: 6. November 2024 ([Dokumentversionen](#))

In diesem Whitepaper geht es speziell um die Säule „Leistungseffizienz“ des AWS-Well-Architected-Framework. Kunden finden hier Informationen zur Anwendung von bewährten Methoden für die Konzeption, Bereitstellung und Wartung von AWS-Umgebungen.

## Einführung

Das [AWS-Well-Architected-Framework](#) unterstützt Sie dabei, die Vor- und Nachteile der Entscheidungen nachzuvollziehen, die Sie beim Erstellen von Workloads in AWS treffen. Das Framework hilft Ihnen, bewährte Architekturmethoden für den Entwurf und Betrieb zuverlässiger, sicherer, effizienter, kostengünstiger und nachhaltiger Workloads in der Cloud zu ermitteln. Es bietet Ihnen die Möglichkeit, Ihre Architekturen konsistent auf die Einhaltung bewährter Methoden zu prüfen und Verbesserungspotenzial zu identifizieren. Wir sind der Meinung, dass eine gute Workload-Architektur die Wahrscheinlichkeit für den geschäftlichen Erfolg deutlich erhöht.

Das Framework basiert auf den folgenden sechs Säulen:

- Operational Excellence
- Sicherheit
- Zuverlässigkeit
- Leistungseffizienz
- Kostenoptimierung
- Nachhaltigkeit

In diesem Dokument geht es um die Anwendung der Prinzipien der Säule „Leistungseffizienz“ auf Ihre Workloads. In herkömmlichen On-Premises-Umgebungen stellt das Erzielen einer konsistent hohen Leistung eine Herausforderung dar. Die Umsetzung der Prinzipien in diesem Whitepaper hilft Ihnen beim Aufbau von Architekturen in AWS, die im Laufe der Zeit effizient für eine nachhaltige Leistung sorgen. Die Leitfäden und bewährten Verfahren in diesem Dokument verteilen sich auf fünf Schwerpunktbereiche, die als Leitprinzipien für die Entwicklung leistungsfähiger Cloud-Lösungen in AWS dienen. Schwerpunktbereiche:

- [Auswahl der Architektur](#)
- [Computer und Hardware](#)
- [Datenverwaltung](#)
- [Netzwerk und Bereitstellung von Inhalten](#)
- [Prozess und Kultur](#)

Dieses Dokument richtet sich an Personen in technologischen Funktionen wie Chief Technology Officers (CTOs), Architekten, Entwickler und Mitglieder des Betriebsteams. Sie erfahren darin mehr über die bewährten Methoden und Strategien von AWS zum Entwickeln einer leistungsfähigen Cloud-Architektur.

# Leistungseffizienz

Die Säule der Leistungseffizienz betrifft die Fähigkeit zur effizienten Nutzung von Cloud-Ressourcen, um die Leistungsanforderungen zu erfüllen, sowie die Möglichkeit zur Aufrechterhaltung dieser Effizienz bei Nachfrageänderungen und einer Weiterentwicklung der Technologien.

Themen

- [Designprinzipien](#)
- [Definition](#)

## Designprinzipien

Die folgenden Designprinzipien können Ihnen dabei helfen, effiziente Workloads in der Cloud zu erzielen und zu verwalten.

- **Demokratisieren fortschrittlicher Technologien:** Vereinfachen Sie die Implementierung fortschrittlicher Technologien für Ihr Team, indem Sie komplexe Aufgaben an Ihren Cloud-Anbieter delegieren. Statt Ihr IT-Team aufzufordern, sich näher über das Hosten und Ausführen einer neuen Technologie zu informieren, sollten Sie die Technologie als Service nutzen. Zum Beispiel sind keine SQL Datenbanken, Medientranscodierung und maschinelles Lernen alles Technologien, für die spezielles Fachwissen erforderlich ist. In der Cloud kann Ihr Team diese Technologien als Service nutzen und sich auf die Produktentwicklung konzentrieren, ohne sich um die Bereitstellung und Verwaltung von Ressourcen kümmern zu müssen.
- **Werden Sie innerhalb weniger Minuten global:** Durch die Bereitstellung Ihrer Workloads in mehreren AWS Regionen auf der ganzen Welt können Sie Ihren Kunden bei minimalen Kosten eine geringere Latenz und ein besseres Erlebnis bieten.
- **Nutzung von Serverless-Architekturen:** Aufgrund der in der Cloud verwendeten Serverless-Architekturen brauchen Sie für herkömmliche Datenverarbeitungsaktivitäten keine physischen Server mehr auszuführen und zu verwalten. Serverless-Speicherservices können beispielsweise als statische Websites genutzt werden, wodurch sich Webserver erübrigen. Ihren Code können Sie von Ereignisservices hosten lassen. Auf diese Weise entfällt nicht nur die Verwaltung physischer Server, sondern auch die Transaktionskosten sinken, da verwaltete Services in der Cloud-Umgebung ausgeführt werden.
- **Vermehrtes Experimentieren:** Mit virtuellen und automatisierbaren Ressourcen können Sie schnell unterschiedliche Konfigurationen, Instance- oder Speichertypen miteinander vergleichen.

- Berücksichtigen des technischen Verständnisses: Nutzen Sie den für Ihre Anforderungen optimalen Technologieansatz. Berücksichtigen Sie bei der Auswahl der Datenbank oder von Speicher für Ihre Workload beispielsweise die Datenzugriffsmuster.

## Definition

Konzentrieren Sie sich auf die folgenden Bereiche, um Leistungseffizienz in der Cloud zu erzielen:

- [Auswahl der Architektur](#)
- [Computer und Hardware](#)
- [Datenverwaltung](#)
- [Netzwerk und Bereitstellung von Inhalten](#)
- [Prozess und Kultur](#)

Um eine leistungsstarke Architektur sicherzustellen, empfiehlt sich für deren Entwicklung ein datenbasierter Ansatz. Sammeln Sie zu allen Aspekten der Architektur Daten, angefangen vom allgemeinen Design bis hin zur Auswahl und Konfiguration der Ressourcentypen.

Wenn Sie Ihre Optionen regelmäßig überprüfen, stellen Sie sicher, dass Sie die Vorteile der sich ständig weiterentwickelnden AWS Cloud nutzen. Durch Überwachung erkennen Sie Abweichungen von der erwarteten Leistung. Zur Leistungssteigerung der Architektur können Sie auch Kompromisse eingehen, beispielsweise durch Komprimierung oder Caching, oder indem Sie hinsichtlich der Konsistenz mehr Toleranz einräumen.

# Auswahl der Architektur

Die optimale Lösung für eine bestimmte Workload variiert und Lösungen sind häufig eine Kombination mehrerer Ansätze. Well-Architected-Workloads nutzen mehrere Lösungen und ermöglichen verschiedene Features zur Verbesserung der Leistung.

AWS-Ressourcen sind in vielen Typen und Konfigurationen verfügbar, wodurch es einfacher ist, einen Ansatz zu finden, der Ihren Anforderungen weitgehend entspricht. Sie können zudem Optionen nutzen, die sich in Ihrer On-Premises-Infrastruktur nicht ohne Weiteres umsetzen ließen. Nehmen wir beispielsweise den verwalteten Service Amazon DynamoDB. Dieser bietet eine vollständig verwaltete NoSQL-Datenbank mit einer Latenz im einstelligen Millisekundenbereich ungeachtet des Volumens.

In diesem Schwerpunktbereich finden Sie Anleitungen und bewährte Methoden zur Auswahl effizienter, leistungsstarker Cloud-Ressourcen und Architekturmuster.

## Bewährte Methoden

- [PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features](#)
- [PERF01-BP02 Einholen von Rat beim Cloud-Anbieter oder einem geeigneten Partner, um mehr über Architekturmuster und bewährte Methoden zu erfahren](#)
- [PERF01-BP03 Kosten bei Architekturentscheidungen berücksichtigen](#)
- [PERF01-BP04 Evaluieren, wie sich Kompromisse auf Kunden und Architektureffizienz auswirken](#)
- [PERF01-BP05 Benutzungsrichtlinien und Referenzarchitekturen](#)
- [PERF01-BP06 Verwenden von Benchmarking, um architektonische Entscheidungen zu treffen](#)
- [PERF01-BP07 Verwenden eines datengesteuerten Ansatzes für architektonische Entscheidungen](#)

## PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features

Informieren Sie sich kontinuierlich über verfügbare Services und Konfigurationen, die Ihnen helfen, bessere architektonische Entscheidungen zu treffen und die Leistungseffizienz Ihrer Workload-Architektur zu verbessern.

### Typische Anti-Muster:

- Sie verwenden die Cloud als gemeinsam genutztes Rechenzentrum.

- Sie modernisieren die Anwendung nach der Migration in die Cloud nicht.
- Sie verwenden nur einen Speichertyp für alle Objekte, die gespeichert werden müssen.
- Sie verwenden Instance-Typen, die am besten zu Ihren aktuellen Standards passen, bei Bedarf jedoch größer sind.
- Von Ihnen werden Technologien bereitgestellt und verwaltet, die als verwaltete Services verfügbar sind.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie neue Services und Konfigurationen in Betracht ziehen, können Sie möglicherweise die Leistung erheblich verbessern, die Kosten senken und den Aufwand für die Aufrechterhaltung der Workload optimieren. Es kann Ihnen auch dabei helfen, die Wertschöpfung für Cloud-fähige Produkte zu beschleunigen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

AWS veröffentlicht kontinuierlich neue Services und Features, mit denen die Leistung verbessert und die Kosten von Cloud-Workloads gesenkt werden können. Es ist entscheidend, mit diesen neuen Services und Features auf dem Laufenden zu bleiben, um die Leistungseffizienz in der Cloud aufrechtzuerhalten. Die Modernisierung der Workload-Architektur hilft Ihnen auch dabei, die Produktivität zu beschleunigen, Innovationen voranzutreiben und mehr Wachstumschancen zu erschließen.

## Implementierungsschritte

- Inventarisieren Sie die Workload-Software und -Architektur für verwandte Services. Entscheiden Sie, über welche Produktkategorie Sie mehr erfahren möchten.
- Erkunden Sie die AWS-Angebote, um die relevanten Services und Konfigurationsoptionen zu identifizieren und kennenzulernen, mit denen Sie die Leistung verbessern und die Kosten und die betriebliche Komplexität reduzieren können.
  - [Amazon Web Services Cloud](#)
  - [AWS Academy](#)
  - [Neuerungen bei AWS](#)
  - [AWS-Blog](#)
  - [AWS Skill Builder](#)
  - [AWS-Veranstaltungen und -Webinare](#)

- [AWS Training und Zertifizierungen](#)
- [AWS-YouTube-Kanal](#)
- [AWS-Workshops](#)
- [AWS-Communitys](#)
- Verwenden Sie [Amazon Q](#), um relevante Informationen und Tipps zu Services zu erhalten.
- Verwenden Sie Sandbox- bzw. Nicht-Produktionsumgebungen, um neue Services zu erlernen und mit ihnen zu experimentieren, ohne dass zusätzliche Kosten anfallen.
- Informieren Sie sich kontinuierlich über neue Cloud-Services und -Features.

## Ressourcen

### Zugehörige Dokumente:

- [Übersicht über Amazon Web Services](#)
- [Amazon EC2 features](#)
- [Lernen Sie Schritt für Schritt mit einem Lehrplan für AWS-Partner](#)
- [AWS Training and Certification](#)
- [My learning path to become an AWS solutions architect](#)
- [AWS-Architekturzentrum](#)
- [AWS Partner Network](#)
- [AWS Lösungsbibliothek](#)
- [AWS Knowledge Center](#)
- [Moderne Anwendungen in AWS entwickeln](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – What's new with Amazon EC2](#)
- [AWS re:Invent 2022 – Reduce your operational and infrastructure costs with Amazon ECS](#)
- [AWS re:Invent 2023 – Mit AWS die Effizienz, Agilität und Innovation der Cloud nutzen](#)
- [AWS re:Invent 2022 – ML-Modelle für Inferenz mit hoher Leistung und niedrigen Kosten bereitstellen](#)
- [This is my Architecture](#)

Zugehörige Beispiele:

- [AWS Beispiele](#)
- [AWS-SDK-Beispiele](#)

## PERF01-BP02 Einholen von Rat beim Cloud-Anbieter oder einem geeigneten Partner, um mehr über Architekturmuster und bewährte Methoden zu erfahren

Greifen Sie bei Ihren architektonischen Entscheidungen auf die Ressourcen von Cloud-Unternehmen, wie etwa Dokumentation, Lösungsarchitekten, professionelle Services oder einen geeigneten Partner zurück. Diese Ressourcen helfen Ihnen dabei, Ihre Architektur zu überprüfen und zu verbessern, um so die Leistung zu optimieren.

Typische Anti-Muster:

- Sie verwenden AWS als gängigen Cloud-Anbieter.
- Sie verwenden AWS-Services auf eine Weise, für die sie nicht konzipiert wurden.
- Sie befolgen alle Anweisungen, ohne Ihren Geschäftskontext zu berücksichtigen.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie sich Rat bei einem Cloud-Anbieter oder einem geeigneten Partner einholen, können Sie die richtige Architektur für die Workload wählen und Entscheidungen mit größerer Zuversicht treffen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: Mittel

### Implementierungsleitfaden

AWS bietet eine breite Palette an Anleitungen, Dokumentation und Ressourcen, die Sie bei der Entwicklung und Verwaltung effizienter Cloud-Workloads unterstützen können. Die AWS-Dokumentation enthält Codebeispiele, Tutorials und detaillierte Serviceerklärungen. Zusätzlich zur Dokumentation bietet AWS Trainings- und Zertifizierungsprogramme, Lösungsarchitekten und professionelle Services, die Kunden dabei helfen können, verschiedene Aspekte von Cloud-Services zu entdecken und eine effiziente Cloud-Architektur in AWS zu implementieren.

Nutzen Sie diese Ressourcen, um Einblicke in wertvolles Wissen und bewährte Methoden zu gewinnen, Zeit zu sparen und bessere Ergebnisse in der AWS Cloud zu erzielen.

## Implementierungsschritte

- Lesen Sie die AWS-Dokumentation und -Anleitungen und befolgen Sie die bewährten Methoden. Diese Ressourcen können Ihnen helfen, Services effektiv auszuwählen und zu konfigurieren und eine bessere Leistung zu erzielen.
  - [AWS-Dokumentation](#) (wie Benutzerhandbücher und Whitepapers)
  - [AWSBlog](#)
  - [AWS Training und Zertifizierungen](#)
  - [AWS-YouTube-Kanal](#)
- Nehmen Sie an AWS-Partnerveranstaltungen (wie AWS Global Summits, AWS re:Invent, Benutzergruppen und Workshops) teil, um von AWS-Experten mehr über bewährte Methoden für die Nutzung von AWS-Services zu lernen.
  - [Lernen Sie Schritt für Schritt mit einem Lehrplan für AWS-Partner](#)
  - [AWS-Veranstaltungen und -Webinare](#)
  - [AWS-Workshops](#)
  - [AWS-Communitys](#)
- Wenden Sie sich an AWS, wenn Sie zusätzliche Anleitungen oder Produktinformationen benötigen. AWS Solutions Architects und [AWS Professional Services](#) liefern Ratschläge für die Implementierung von Lösungen. [AWS Partner](#) bieten AWS-Fachwissen, damit Sie in Ihrem Unternehmen flexibel agieren und Innovationen nutzen können.
- Verwenden Sie [Support](#), wenn Sie technischen Support benötigen, um einen Service effektiv nutzen zu können. [Unsere Support-Pläne](#) bieten Ihnen die richtige Kombination aus Tools und Zugang zu Fachwissen, um die Grundlagen für Ihren Erfolg mit AWS zu legen, ohne dabei Themen wie Leistungsoptimierung, Risikomanagement und Kostenkontrolle zu vernachlässigen.

## Ressourcen

Zugehörige Dokumente:

- [AWS-Architekturzentrum](#)
- [AWS Partner Network](#)
- [AWS-Lösungsbibliothek](#)
- [AWS Knowledge Center](#)
- [AWS Enterprise Support](#)

## Zugehörige Videos:

- [This is my Architecture](#)
- [AWS re:Invent 2.023 - Advanced event-driven patterns with Amazon EventBridge](#)
- [AWS re:Invent 2023 – Implementierung verteilter Designmuster in AWS](#)
- [AWS re:Invent 2023 – Anwendungsarchitektur als Code](#)

## Zugehörige Beispiele:

- [AWS Beispiele](#)
- [AWS-SDK-Beispiele](#)
- [AWS-Analytics-Referenzarchitektur](#)

# PERF01-BP03 Kosten bei Architekturentscheidungen berücksichtigen

Berücksichtigen Sie die Kosten bei Ihren architektonischen Entscheidungen, um die Ressourcennutzung und Leistungseffizienz der Cloud-Workloads zu verbessern. Wenn Sie sich der Kostenauswirkungen der Cloud-Workload bewusst sind, ist es wahrscheinlicher, dass Sie effiziente Ressourcen nutzen und verschwenderische Methoden reduzieren.

## Typische Anti-Muster:

- Sie verwenden nur eine Instance-Familie.
- Sie bewerten keine lizenzierten Lösungen verglichen mit Open-Source-Lösungen.
- Sie definieren keine Speicher-Lebenszyklusrichtlinien.
- Sie überprüfen keine neuen Dienste und Funktionen von AWS Cloud
- Sie nutzen nur Blockspeicher.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie die Kosten bei Ihrer Entscheidungsfindung berücksichtigen, können Sie effizientere Ressourcen einsetzen und andere Investitionen in Betracht ziehen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Die Kostenoptimierung von Workloads kann die Ressourcennutzung verbessern und Verschwendung bei einer Cloud-Workload vermeiden. Die Berücksichtigung der Kosten bei architektonischen Entscheidungen beinhaltet in der Regel die richtige Dimensionierung der Workload-Komponenten und die Schaffung von Elastizität. Dies führt zu einer verbesserten Leistungseffizienz von Cloud-Workloads.

### Implementierungsschritte

- Legen Sie Kostenziele wie Budgetlimits für die Cloud-Workload fest.
- Identifizieren Sie die wesentlichen Komponenten (wie Instances und Speicher), die die Kosten der Workload erhöhen. Sie können [AWS -Preisrechner](#) und [AWS Cost Explorer](#) verwenden, um die wichtigsten Kostentreiber in Ihrer Workload zu identifizieren.
- Informieren Sie sich über die [Preismodelle](#) in der Cloud, z. B. On-Demand, Reserved Instances, Savings Plans und Spot Instances.
- Verwenden Sie die [bewährten Methoden zur Kostenoptimierung von Well-Architected](#), um diese Schlüsselkomponenten im Hinblick auf die Kosten zu optimieren.
- Überwachen und analysieren Sie kontinuierlich die Kosten, um Möglichkeiten zur Kostenoptimierung in der Workload zu identifizieren.
  - Nutzen Sie [AWS -Budgets](#), um bei nicht akzeptablen Kosten Warnungsmeldungen zu erhalten.
  - Verwenden Sie [AWS Compute Optimizer](#) oder [AWS Trusted Advisor](#), um Empfehlungen zur Kostenoptimierung zu erhalten.
  - Nutzen Sie [AWS Cost Anomaly Detection](#), um das automatisierte Erkennen von Kostenanomalien mit Ursachenanalyse zu erhalten.

## Ressourcen

Zugehörige Dokumente:

- [Was ist AWS Billing and Cost Management?](#)
- [Kostenoptimierung mit AWS](#)
- [Wahl einer AWS Kostenmanagementstrategie](#)
- [Ein Leitfaden für Anfänger zum AWS Kostenmanagement](#)
- [Eine detaillierte Übersicht über das Cost Intelligence Dashboard](#)

- [AWS -Architekturzentrum](#)
- [AWS -Lösungsbibliothek](#)
- [AWS Knowledge Center](#)

Zugehörige Videos:

- [This is my Architecture](#)
- [AWS re:Invent 2023 — Was ist neu bei AWS der Kostenoptimierung](#)
- [AWS re:Invent 2023 — Optimieren Sie Kosten und Leistung und verfolgen Sie die Fortschritte bei der Minderung](#)
- [AWS re:Invent 2023 — bewährte Methoden zur Optimierung der Speicherkosten AWS](#)
- [AWS re:Invent 2023 — Optimieren Sie die Kosten in Ihren Umgebungen mit mehreren Konten](#)

Zugehörige Beispiele:

- [AWS Compute Optimizer Demo-Code](#)
- [Workshop zur Kostenoptimierung](#)
- [Technische Playbooks zur Implementierung von Cloud Financial Management](#)
- [Startoptimierung: Optimierung der Anwendungsleistung für maximale Effizienz](#)
- [Workshop zur Serverless-Optimierung \(Leistung und Kosten\)](#)
- [Skalierung kostengünstiger Architekturen](#)

## PERF01-BP04 Evaluieren, wie sich Kompromisse auf Kunden und Architektureffizienz auswirken

Ermitteln Sie beim Evaluieren von leistungsbezogenen Verbesserungen, welche gewählten Optionen sich auf Ihre Kunden und die Effizienz der Workloads auswirken. Wenn sich die Systemleistung beispielsweise bei Verwendung eines Schlüssel-Wert-Datenspeichers erhöht, sollten Sie unbedingt ermitteln, welche Auswirkungen sich bei einem dauerhaften Einsatz für die Kunden ergeben würden.

Typische Anti-Muster:

- Sie gehen davon aus, dass alle Leistungsgewinne implementiert werden sollten, auch wenn es Kompromisse für die Implementierung gibt.

- Änderungen an Workloads werden nur dann ausgewertet, wenn ein Leistungsproblem einen kritischen Punkt erreicht hat.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie potenzielle leistungsbezogene Verbesserungen bewerten, müssen Sie entscheiden, ob die Kompromisse für die Änderungen angesichts der Workload-Anforderungen akzeptabel sind. In einigen Fällen müssen Sie möglicherweise zusätzliche Kontrollen implementieren, um Kompromisse zu kompensieren.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

Identifizieren Sie kritische Bereiche in der Architektur in Bezug auf Leistung und Kundenauswirkung. Stellen Sie fest, welche Verbesserungen möglich und welche Kompromisse damit verbunden sind und wie sich diese auf das System und das Benutzererlebnis auswirken. So lässt sich beispielsweise durch Caching von Daten die Leistung deutlich steigern. Es ist aber eine eindeutige Strategie erforderlich, mit der festgelegt wird, wie und wann Cache-Daten aktualisiert oder ungültig werden, um unerwünschtes Systemverhalten zu verhindern.

## Implementierungsschritte

- Verstehen Sie Ihre Workload-Anforderungen und SLAs.
- Definieren Sie klare Bewertungsfaktoren. Faktoren können sich auf Kosten, Zuverlässigkeit, Sicherheit und Leistung der Workload beziehen.
- Wählen Sie die Architektur und Services, die Ihren Anforderungen entsprechen.
- Führen Sie Experimente und Machbarkeitsstudien (POCs) durch, um Kompromissfaktoren und Auswirkungen auf Kunden und Architektureffizienz zu bewerten. In der Regel verbrauchen hochverfügbare, leistungsstarke und sichere Workloads mehr Cloud-Ressourcen und bieten gleichzeitig ein besseres Kundenerlebnis. Machen Sie sich ein Bild von den Kompromissen in Bezug auf Komplexität, Leistung und Kosten Ihrer Workloads. In der Regel geht die Priorisierung von zwei der Faktoren auf Kosten des dritten.

## Ressourcen

Zugehörige Dokumente:

- [Amazon Builders' Library](#)

- [Quick-KPIs](#)
- [Amazon CloudWatch RUM](#)
- [X-Ray-Dokumentation](#)
- [Resilienzmuster und Kompromisse verstehen, um eine effiziente Architektur in der Cloud zu entwickeln](#)

Zugehörige Videos:

- [Optimize applications through Amazon CloudWatch RUM](#)
- [AWS re:Invent 2023 – Kapazität, Verfügbarkeit, Kosteneffizienz: Wählen Sie drei Optionen aus](#)
- [AWS re:Invent 2023 – Erweiterte Integrationsmuster und Kompromisse für lose gekoppelte Systeme](#)

Zugehörige Beispiele:

- [Messen der Seitenladezeit mit Amazon CloudWatch Synthetics](#)
- [Webclient von Amazon CloudWatch RUM](#)

## PERF01-BP05 Benutzungsrichtlinien und Referenzarchitekturen

Verwenden Sie interne Richtlinien und vorhandene Referenzarchitekturen bei der Auswahl von Services und Konfigurationen, um die Workload effizienter zu gestalten und zu implementieren.

Typische Anti-Muster:

- Sie erlauben eine Vielzahl von Technologien, was sich auf den Verwaltungsaufwand Ihres Unternehmens auswirken kann.

Vorteile der Nutzung dieser bewährten Methode: Durch Festlegung einer Richtlinie für die Architektur-, Technologie und Anbietersauswahl können Entscheidungen schnell getroffen werden.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Interne Richtlinien bei der Auswahl von Ressourcen und Architektur bieten Standards und Leitlinien, die bei Architekturentscheidungen zu beachten sind. Diese Richtlinien vereinfachen den Entscheidungsprozess bei der Auswahl des richtigen Cloud-Service und können zur Verbesserung der Leistungseffizienz beitragen. Stellen Sie die Workload mithilfe von Richtlinien oder Referenzarchitekturen bereit. Integrieren Sie die Services in Ihre Cloud-Bereitstellung. Überprüfen Sie anschließend anhand von Leistungstests, dass Sie die eigenen Leistungsanforderungen weiterhin erfüllen können.

### Implementierungsschritte

- Verstehen Sie die Anforderungen der Cloud-Workload genau.
- Überprüfen Sie die internen und externen Richtlinien, um die relevantesten zu ermitteln.
- Verwenden Sie die entsprechenden Referenzarchitekturen, die von AWS bereitgestellt werden, oder die branchenweit anerkannten bewährten Methoden.
- Schaffen Sie ein Kontinuum, das aus Richtlinien, Standards, Referenzarchitekturen und präskriptiven Richtlinien für häufig auftretende Situationen besteht. Auf diese Weise können Ihre Teams schneller vorankommen. Passen Sie die Komponenten gegebenenfalls an die Branche an.
- Prüfen Sie diese Richtlinien und Referenzarchitekturen für die Workload in Sandbox-Umgebungen.
- Halten Sie up-to-date sich an Industriestandards und AWS Updates, um sicherzustellen, dass Ihre Richtlinien und Referenzarchitekturen zur Optimierung Ihrer Cloud-Workloads beitragen.

## Ressourcen

Zugehörige Dokumente:

- [AWS -Architekturzentrum](#)
- [AWS Partner Network](#)
- [AWS -Lösungsbibliothek](#)
- [AWS Knowledge Center](#)
- [AWS Architektur-Blog](#)

Zugehörige Videos:

- [This is my Architecture](#)
- [AWS re:Invent 2022 — Steigern Sie den Wert Ihres Unternehmens mit SAP Referenzarchitektur AWS](#)

Zugehörige Beispiele:

- [AWS Beispiele](#)
- [AWS SDKBeispiele](#)

## PERF01-BP06 Verwenden von Benchmarking, um architektonische Entscheidungen zu treffen

Führen Sie einen Benchmark-Vergleich für eine vorhandene Workload durch, um sich ein Bild über deren Leistung in der Cloud zu verschaffen, und treffen Sie architektonische Entscheidungen auf der Grundlage dieser Daten.

Typische Anti-Muster:

- Sie verlassen sich auf gängige Benchmarks, die für die Workload-Merkmale nicht aufschlussreich sind.
- Sie verlassen sich auf Kundenfeedback und Kundenwahrnehmung als einzige Benchmark.

Vorteile der Einführung dieser bewährten Methode: Durch Benchmarking Ihrer aktuellen Implementierung können Sie Leistungsverbesserungen messen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

### Implementierungsleitfaden

Kombinieren Sie Benchmarking mit synthetischen Tests, um die Leistung Ihrer Workload-Komponenten zu bewerten. Benchmarking lässt sich in der Regel schneller als Lasttests einrichten und dient zur Bewertung der Technologie einer bestimmten Komponente. Ein Benchmarking wird oft zu Beginn eines neuen Projekts durchgeführt, wenn Sie noch keine vollständige Lösung für einen Lasttest haben.

Sie können wahlweise eigene Benchmark-Tests erstellen oder einen branchenüblichen Standardtest wie etwa [TPC-DS](#) für das Benchmarking Ihrer Workloads verwenden. Branchen-Benchmarks sind

zum Vergleich von Umgebungen nützlich. Benutzerdefinierte Benchmarks eignen sich zum Prüfen spezieller Arten von Vorgängen, die Sie in der Architektur ausführen möchten.

Beim Benchmarking ist es wichtig, die Testumgebung entsprechend vorzubereiten, um aussagekräftige Ergebnisse zu erzielen. Führen Sie denselben Benchmark-Test mehrmals aus, um sicherzustellen, dass alle Varianzen im Laufe der Zeit ermittelt wurden.

Da sich Benchmarks in der Regel schneller als Lasttests ausführen lassen, können Sie früher in der Bereitstellungs-pipeline eingesetzt werden und schneller Feedback zu Leistungsabweichungen liefern. Wenn Sie eine wesentliche Veränderung einer Komponente oder eines Services bewerten, können Sie schnell ermitteln, ob der Aufwand für die Korrektur gerechtfertigt ist. Die Verwendung von Benchmarking in Verbindung mit Lasttests ist wichtig, da letztere Auskunft über die Leistung der Workload in der Produktion geben.

## Implementierungsschritte

- Planen und Definieren:
  - Definieren Sie die Ziele, Baselines, Testszenarien, Metriken (wie CPU-Auslastung, Latenz oder Durchsatz) und KPIs für Ihren Benchmark.
  - Konzentrieren Sie sich auf die Benutzeranforderungen in Bezug auf das Benutzererlebnis und Faktoren wie Reaktionszeit und Barrierefreiheit.
  - Identifizieren Sie ein Benchmarking-Tool, das für Ihre Workload geeignet ist. Sie können AWS-Services (wie [Amazon CloudWatch](#)) oder ein Drittanbieter-Tool verwenden, das mit der Workload kompatibel ist.
- Konfiguration und Verwendung:
  - Richten Sie Ihre Umgebung ein und konfigurieren Sie Ihre Ressourcen.
  - Implementieren Sie Überwachungs- und Protokollierungsfunktionen, um Testergebnisse zu erfassen.
- Benchmarking und Überwachung:
  - Führen Sie die Benchmark-Tests durch und überwachen Sie die Metriken während des Tests.
- Analyse und Dokumentation:
  - Dokumentieren Sie Ihren Benchmarking-Prozess und die entsprechenden Erkenntnisse.
  - Analysieren Sie die Ergebnisse, um Engpässe, Trends und Verbesserungsmöglichkeiten zu identifizieren.

- Verwenden Sie die Testergebnisse, um die Architektur betreffende Entscheidungen zu fällen und die Workload anzupassen. Dies kann die Änderung von Services oder die Einführung neuer Features beinhalten.
- Optimierung und Wiederholung:
  - Passen Sie die Ressourcenkonfigurationen und -zuweisungen auf der Grundlage Ihrer Benchmarks an.
  - Testen Sie Ihre Workload nach der Anpassung erneut, um Ihre Verbesserungen zu überprüfen.
  - Dokumentieren Sie Ihre Erkenntnisse und wiederholen Sie den Prozess, um weitere Verbesserungsmöglichkeiten zu identifizieren.

## Ressourcen

### Zugehörige Dokumente:

- [AWS Architekturzentrum](#)
- [AWS Partner Network](#)
- [AWS Lösungsbibliothek](#)
- [AWS Knowledge Center](#)
- [Amazon CloudWatch RUM](#)
- [Amazon CloudWatch Synthetics](#)
- [Genomics workflows, Part 5: automated benchmarking](#)
- [Benchmark and optimize endpoint deployment in Amazon SageMaker AI JumpStart](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Benchmarking AWS Lambda cold starts](#)
- [Benchmarking stateful services in the cloud](#)
- [This is my Architecture](#)
- [Optimize applications through Amazon CloudWatch RUM](#)
- [Demo von Amazon CloudWatch Synthetics](#)

### Zugehörige Beispiele:

- [AWS Beispiele](#)
- [AWS-SDK-Beispiele](#)
- [Verteilte Lasttests](#)
- [Messen der Seitenladezeit mit Amazon CloudWatch Synthetics](#)
- [Webclient von Amazon CloudWatch RUM](#)

## PERF01-BP07 Verwenden eines datengesteuerten Ansatzes für architektonische Entscheidungen

Definieren Sie einen klaren, datengesteuerten Ansatz für architektonische Entscheidungen, um sicherzustellen, dass die richtigen Cloud-Services und -Konfigurationen verwendet werden, um Ihre spezifischen Geschäftsanforderungen zu erfüllen.

Typische Anti-Muster:

- Sie gehen davon aus, dass die aktuelle Architektur statisch ist und im Laufe der Zeit nicht aktualisiert werden sollte.
- Ihre architektonischen Entscheidungen basieren auf Vermutungen und Annahmen.
- Sie führen im Laufe der Zeit Änderungen an der Architektur ein, ohne sie zu begründen.

Vorteile der Nutzung dieser bewährten Methode: Durch einen klar definierten Ansatz für architektonische Entscheidungen verwenden Sie Daten, um das Workload-Design zu beeinflussen und im Laufe der Zeit fundierte Entscheidungen zu treffen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

### Implementierungsleitfaden

Nutzen Sie interne Erfahrungen und Kenntnisse im Zusammenhang mit der Cloud oder ziehen Sie externe Ressourcen heran, wie etwa veröffentlichte Anwendungsbeispiele oder Whitepapers, um Ressourcen und Services in der Architektur auszuwählen. Sie sollten über einen klar definierten Prozess verfügen, der das Experimentieren und Benchmarking mit den Services fördert, die in der Workload verwendet werden könnten.

Backlogs für kritische Workloads sollten nicht nur aus Benutzerszenarien bestehen, die für das Unternehmen und die Benutzer relevante Funktionen bereitstellen, sondern auch aus technischen

Szenarien, die ein architektonisches System für die Workload bilden. Dieses System stützt sich auf neue technologische Fortschritte sowie neue Services und nimmt diese auf der Grundlage von Daten und entsprechender Begründung an. Dies stellt sicher, dass die Architektur zukunftssicher bleibt und nicht stagniert.

## Implementierungsschritte

- Arbeiten Sie mit wichtigen Stakeholdern zusammen, um die Workload-Anforderungen zu definieren, einschließlich Überlegungen zu Leistung, Verfügbarkeit und Kosten. Berücksichtigen Sie Faktoren wie die Anzahl der Benutzer und das Nutzungsmuster für die Workload.
- Erstellen Sie ein Architektursystem oder einen Technologie-Backlog, der zusammen mit dem funktionalen Backlog priorisiert wird.
- Bewerten und beurteilen Sie verschiedene Cloud-Services (weitere Informationen finden Sie unter [PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features](#)).
- Erkunden Sie verschiedene Architekturmuster wie Microservices oder Serverless, die Ihren Leistungsanforderungen entsprechen (weitere Informationen finden Sie unter [PERF01-BP02 Einholen von Rat beim Cloud-Anbieter oder einem geeigneten Partner, um mehr über Architekturmuster und bewährte Methoden zu erfahren](#)).
- Ziehen Sie andere Teams, Architekturdiagramme und Ressourcen zu Rate, z. B. AWS Solution Architects, [AWS-Architekturzentrum](#) und [AWS Partner Network](#), um die richtige Architektur für Ihre Workload auszuwählen.
- Definieren Sie Leistungsmetriken wie Durchsatz und Reaktionszeit, anhand derer Sie die Leistung der Workload bewerten können.
- Experimentieren Sie und verwenden Sie definierte Metriken, um die Leistung der ausgewählten Architektur zu validieren.
- Überwachen Sie kontinuierlich und nehmen Sie bei Bedarf Anpassungen vor, um die optimale Leistung der Architektur aufrechtzuerhalten.
- Dokumentieren Sie Ihre gewählte Architektur und Entscheidungen als Referenz für zukünftige Updates und Erkenntnisse.
- Überprüfen und aktualisieren Sie den Ansatz zur Architekturauswahl kontinuierlich auf der Grundlage von Erkenntnissen, neuen Technologien und Metriken, die auf eine notwendige Änderung oder ein Problem im aktuellen Ansatz hinweisen.

# Ressourcen

## Zugehörige Dokumente:

- [AWS-Lösungsbibliothek](#)
- [AWS Knowledge Center](#)
- [Architekturmodelle für die Erstellung von datengesteuerten End-to-End-Anwendungen in AWS](#)

## Zugehörige Videos:

- [This is my Architecture](#)
- [AWS re:Invent 2021 – Das datengesteuerte Unternehmen: Von der Vision zum Mehrwert](#)
- [AWS re:Invent 2022 – Bereitstellung nachhaltiger, leistungsstarker Architekturen](#)
- [AWS re:Invent 2023 – Optimieren der Kosten und Leistung sowie Verfolgen der Fortschritte bei der Schadensbegrenzung](#)
- [AWS re:Invent 2022 – AWS-Optimierung: Umsetzbare Schritte für sofortige Ergebnisse](#)

## Zugehörige Beispiele:

- [AWS Beispiele](#)
- [AWS SDK-Beispiele](#)

# Computer und Hardware

Die optimale Datenverarbeitungsoption für eine bestimmte Workload kann sich je nach Anwendungsdesign, Nutzungsmustern und Konfigurationseinstellungen unterscheiden. Architekturen können verschiedene Datenverarbeitungsoptionen für verschiedene Komponenten verwenden und verschiedene Features zur Verbesserung der Leistung bieten. Die Wahl der falschen Datenverarbeitungslösung für eine Architektur kann die Leistungseffizienz schmälern.

In diesem Schwerpunktbereich finden Sie Anleitungen und bewährte Methoden zur Identifizierung und Optimierung von Datenverarbeitungsoptionen für Leistungseffizienz in der Cloud.

## Bewährte Methoden

- [PERF02-BP01 Auswählen der besten Datenverarbeitungsoptionen für die Workload](#)
- [PERF02-BP02 Verstehen verfügbarer Konfigurationen und Features für die Datenverarbeitung](#)
- [PERF02-BP03 Erfassung rechnerbezogener Metriken](#)
- [PERF02-BP04 Konfigurieren und richtiges Dimensionieren von Datenverarbeitungsressourcen](#)
- [PERF02-BP05 Dynamisches Skalieren von Datenverarbeitungsressourcen](#)
- [PERF02-BP06 Verwenden von optimierten hardwarebasierten Datenverarbeitungsbeschleunigern](#)

## PERF02-BP01 Auswählen der besten Datenverarbeitungsoptionen für die Workload

Wenn Sie die für die Workload am besten geeignete Datenverarbeitungsoption auswählen, können Sie die Leistung verbessern, unnötige Infrastrukturkosten reduzieren und den Betriebsaufwand für die Aufrechterhaltung der Workload senken.

### Typische Anti-Muster:

- Sie verwenden dieselbe Datenverarbeitungsoption, die On-Premises verwendet wurde.
- Ihnen fehlt es an Bewusstsein für Cloud-Datenverarbeitungsoptionen, -Features und -lösungen und wie diese Lösungen die Datenverarbeitungsleistung verbessern können.
- Sie stellen eine bestehende Datenverarbeitungsoption zu viel bereit, um Skalierungs- oder Leistungsanforderungen zu erfüllen, wenn eine alternative Datenverarbeitungsoption den Workload-Merkmalen besser entsprechen würde.

Vorteile der Nutzung dieser bewährten Methode: Durch die Ermittlung der Anforderungen an die Datenverarbeitung und deren Bewertung anhand der verfügbaren Optionen können Sie die Workload ressourceneffizienter gestalten.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

Zur Optimierung der Cloud-Workloads im Hinblick auf Leistungseffizienz ist es wichtig, die am besten geeigneten Datenverarbeitungsoptionen für Ihren Anwendungsfall und Ihre Leistungsanforderungen auszuwählen. AWS bietet eine Vielzahl von Datenverarbeitungsoptionen, die auf unterschiedliche Workloads in der Cloud zugeschnitten sind. Sie können beispielsweise [Amazon EC2](#) verwenden, um virtuelle Server zu starten und zu verwalten, [AWS Lambda](#), um Code auszuführen, ohne Server bereitstellen oder verwalten zu müssen, [Amazon ECS](#) oder [Amazon EKS](#), um Container auszuführen und zu verwalten, oder [AWS Batch](#), um große Datenmengen parallel zu verarbeiten. Basierend auf Ihren Skalierungs- und Datenverarbeitungsanforderungen sollten Sie die optimale Datenverarbeitungslösung für Ihre Situation auswählen und konfigurieren. Sie können auch erwägen, mehrere Arten von Datenverarbeitungslösungen in einer einzigen Workload zu verwenden, da jede ihre eigenen Vor- und Nachteile hat.

Die folgenden Schritte führen Sie durch die Auswahl der richtigen Datenverarbeitungsoptionen, die Ihren Workload-Eigenschaften und Leistungsanforderungen entsprechen.

## Implementierungsschritte

- Verstehen Sie Ihre Workload-Datenverarbeitungsanforderungen. Die zu berücksichtigenden wesentlichen Anforderungen umfassen Anforderungen an Datenverarbeitung, Datenverkehrsmuster, Datenzugriffsmuster, Skalierung und Latenz.
- Erfahren Sie mehr über verschiedene [AWS-Datenverarbeitungsservices](#) für Ihre Workload. Weitere Informationen finden Sie unter [PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features](#). Hier finden Sie einige wichtige AWS-Datenverarbeitungsoptionen, ihre Eigenschaften und gängige Anwendungsfälle:

AWS-Service	Schlüsselmerkmale	Häufige Anwendungsfälle
<a href="#">Amazon Elastic Compute Cloud (Amazon EC2)</a>	Verfügt über eine spezielle Option für Hardware, Lizenzanforderungen,	Lift-and-Shift-Migrationen, monolithische Anwendung

AWS-Service	Schlüsselmerkmale	Häufige Anwendungsfälle
	eine große Auswahl an verschiedenen Instance-Familien, Prozessortypen und Beschleuniger der Datenverarbeitung	, hybride Umgebungen, Enterprise-Anwendungen
<a href="#">Amazon Elastic Container Service (Amazon ECS)</a> , <a href="#">Amazon Elastic Kubernetes Service (Amazon EKS)</a>	Einfache Bereitstellung, konsistente Umgebungen, skalierbar	Microservices, Hybrid-Umgebungen
<a href="#">AWS Lambda</a>	<a href="#">Serverless-Datenverarbeitungsservice</a> , der Code als Reaktion auf Ereignisse ausführt und die zugrunde liegenden Ressourcen für die Datenverarbeitung automatisch verwaltet.	Microservices, ereignisgesteuerte Anwendungen
<a href="#">AWS Batch</a>	Effiziente und dynamische Bereitstellung und Skalierung von <a href="#">Amazon Elastic Container Service (Amazon ECS)</a> , <a href="#">Amazon Elastic Kubernetes Service (Amazon EKS)</a> und <a href="#">AWS Fargate</a> -Ressourcen der Datenverarbeitung, mit der Option, On-Demand- oder Spot Instances je nach Ihren Jobanforderungen zu verwenden	HPC, ML-Modelle trainieren
<a href="#">Amazon Lightsail</a>	Vorkonfigurierte Linux- und Windows-Anwendung für die Ausführung kleiner Workloads	Einfache Webanwendungen, benutzerdefinierte Website

- Bewerten Sie die Kosten (wie stündliche Gebühr oder Datenübertragung) und den Verwaltungsaufwand (wie Patching und Skalierung), die mit jeder Datenverarbeitungsoption verbunden sind.
- Führen Sie Experimente und Benchmarking in einer Nicht-Produktionsumgebung durch, um herauszufinden, welche Datenverarbeitungsoption Ihre Workload-Anforderungen am besten erfüllt.
- Nachdem Sie experimentiert und die neue Datenverarbeitungslösung ermittelt haben, planen Sie die Migration und überprüfen Sie die Leistungsmetriken.
- Verwenden Sie AWS-Überwachungstools wie [Amazon CloudWatch](#) und Optimierungsservices wie [AWS Compute Optimizer](#), um die Datenverarbeitungsressourcen kontinuierlich auf der Grundlage realer Nutzungsmuster zu optimieren.

## Ressourcen

### Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon-EC2-Instance-Typen](#)
- [Amazon EKS Containers: Amazon EKS Worker Nodes](#)
- [Amazon ECS Containers: Amazon ECS Container Instances](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)
- [Prescriptive Guidance for Containers](#)
- [Prescriptive Guidance for Serverless](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – AWS Graviton: Das beste Preis-Leistungs-Verhältnis für Ihre AWS-Workloads](#)
- [AWS re:Invent 2023 – New Amazon Elastic Compute Cloud generative AI capabilities in AMS](#)
- [AWS re:Invent 2023 – What's new with Amazon Elastic Compute Cloud](#)
- [AWS re:Invent 2023 – Smart savings: Amazon Elastic Compute Cloud cost-optimization strategies](#)
- [AWS re:Invent 2021 – Powering next-gen Amazon Elastic Compute Cloud: Deep dive on the Nitro System](#)
- [AWS re:Invent 2019 – Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS](#)

- [AWS re:Invent 2.019 – Amazon Elastic Compute Cloud foundations](#)
- [AWS re:Invent 2022 – ML-Modelle für Inferenz mit hoher Leistung und niedrigen Kosten bereitstellen](#)
- [AWS re:Invent 2019 – Optimieren von Leistung und Kosten für die Datenverarbeitung bei AWS](#)
- [Grundlagen von Amazon EC2](#)
- [Stellen Sie ML-Modelle für Inference mit hoher Leistung und niedrigen Kosten bereit](#)

Zugehörige Beispiele:

- [Migration der Webanwendung zu Containern](#)
- [Ausführen eines Serverless-„Hello World“](#)
- [Workshop zu Amazon EKS](#)
- [Workshop zu Amazon EC2](#)
- [Effiziente und belastbare Workloads mit Amazon Elastic Compute Cloud Auto Scaling](#)
- [Migration zu AWS Graviton mit Container Services](#)

## PERF02-BP02 Verstehen verfügbarer Konfigurationen und Features für die Datenverarbeitung

Informieren Sie sich über die verfügbaren Konfigurationsoptionen und Features für den Datenverarbeitungsservice, damit Sie die richtige Menge an Ressourcen bereitstellen und die Leistungseffizienz verbessern können.

Typische Anti-Muster:

- Sie bewerten keine Datenverarbeitungsoptionen oder verfügbaren Instance-Familien anhand der Workload-Merkmale.
- Sie stellen zu viele Datenverarbeitungsressourcen bereit, um Anforderungen von Nachfragespitzen zu erfüllen.

Vorteile der Nutzung dieser bewährten Methode: Machen Sie sich mit den AWS-Features und -Konfigurationen für die Datenverarbeitung vertraut, sodass Sie eine Datenverarbeitungslösung verwenden können, die für die Workload-Merkmale und -Anforderungen optimiert ist.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Jede Datenverarbeitungslösung verfügt über einzigartige Konfigurationen und Features, um unterschiedliche Workload-Merkmale und -Anforderungen zu unterstützen. Erfahren Sie, wie diese Optionen die Workload ergänzen, und finden Sie heraus, welche Konfigurationsoptionen am besten für Ihre Anwendung geeignet sind. Beispiele für diese Optionen sind Instance-Familien, Größen, Features (GPU, E/A), Bursting, Zeitüberschreitungen, Funktionsgrößen, Container-Instances und Gleichzeitigkeit. Wenn Ihre Workload die gleiche Rechenoption für mehr als vier Wochen verwendet hat und sie davon ausgehen, dass die Eigenschaften in Zukunft gleich bleiben, können Sie mithilfe von [AWS Compute Optimizer](#) herausfinden, ob Ihre aktuelle Datenverarbeitungsoption aus CPU- und Speicherebene für die Workloads geeignet ist.

## Implementierungsschritte

- Verstehen Sie die Workload-Anforderungen (wie CPU-Bedarf, Arbeitsspeicher und Latenz).
- Lesen Sie die AWS-Dokumentation und die bewährten Methoden, um mehr über empfohlene Konfigurationsoptionen zu erfahren, mit denen Sie die Rechenleistung verbessern können. Hier finden Sie einige wichtige Konfigurationsoptionen, die Sie in Betracht ziehen sollten:

Konfigurationsoption	Beispiele
Instance-Typ	<ul style="list-style-type: none"> <li>• <a href="#">Für Datenverarbeitung optimierte</a> Instances eignen sich ideal für Workloads, die ein hohes vCPU-/Arbeitsspeicherverhältnis erfordern.</li> <li>• <a href="#">Arbeitsspeicheroptimierte</a> Instances bieten große Mengen an Arbeitsspeicher, um arbeitsspeicherintensive Workloads zu unterstützen.</li> <li>• <a href="#">Speicheroptimierte</a> Instances wurden für Workloads entworfen, die hohen, sequenziellen Lese- und Schreibzugriff (IOPS) auf lokalen Speicher erfordern.</li> </ul>
Preismodell	<ul style="list-style-type: none"> <li>• Mit <a href="#">On-Demand-Instances</a> können Sie die Datenverarbeitungskapazität nach Sekunde oder Stunde ohne langfristige Verpflichtung</li> </ul>

Konfigurationsoption	Beispiele
	<p>tungen verwenden. Diese Instances eignen sich für Bursting über die Leistungsbasis hinaus.</p> <ul style="list-style-type: none"><li>• <a href="#">Savings Plans</a> bieten erhebliche Einsparungen gegenüber On-Demand-Instances im Austausch gegen die Verpflichtung, eine bestimmte Menge an Rechenleistung für einen Zeitraum von ein oder drei Jahren zu nutzen.</li><li>• <a href="#">Spot Instances</a> ermöglichen es Ihnen, ungenutzte Instance-Kapazitäten mit einem Rabatt für Ihre zustandslosen, fehlertoleranten Workloads zu nutzen.</li></ul>
Auto Scaling	Nutzen Sie die <a href="#">Auto-Scaling</a> -Konfiguration zur Anpassung der Datenverarbeitungsressourcen an die Datenverkehrsmuster.
Dimensionierung	<ul style="list-style-type: none"><li>• Nutzen Sie <a href="#">Compute Optimizer</a> zum Erhalt von Machine-Learning-gestützten Empfehlungen dazu, welche Datenverarbeitungskonfiguration am besten Ihren Datenverarbeitungsmerkmalen entspricht.</li><li>• Mit <a href="#">AWS Lambda Power Tuning</a> können Sie die beste Konfiguration für Ihre Lambda-Funktion auswählen.</li></ul>
Hardwarebasierte Computing-Beschleuniger	<ul style="list-style-type: none"><li>• <a href="#">Beschleunigte Computing-Instances</a> führen Funktionen wie die Grafikverarbeitung oder Datenmusterzuordnung effizienter aus als CPU-basierte Alternativen.</li><li>• Nutzen Sie für Machine-Learning-Workloads spezielle Hardware, die auf Ihre Workload abgestimmt ist, z. B. <a href="#">AWS Trainium</a>, <a href="#">AWS Inferentia</a> und <a href="#">Amazon EC2 DL1</a>.</li></ul>

# Ressourcen

## Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon-EC2-Instance-Typen](#)
- [Processor State Control for Your Amazon EC2 Instance](#)
- [Amazon EKS Containers: Amazon EKS Worker Nodes](#)
- [Amazon ECS Containers: Amazon ECS Container Instances](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)

## Zugehörige Videos:

- [AWS re:Invent 2023 – AWS Graviton: Das beste Preis-Leistungs-Verhältnis für Ihre AWS-Workloads](#)
- [AWS re:Invent 2023 – New Amazon EC2 generative AI capabilities in AWS-Managementkonsole](#)
- [AWS re:Invent 2023 – What's new with Amazon EC2](#)
- [AWS re:Invent 2023 – Smart savings: Amazon EC2 cost-optimization strategies](#)
- [AWS re:Invent 2021 – Powering next-gen Amazon EC2: Deep dive on the Nitro System](#)
- [AWS re:Invent 2019 – Amazon EC2 foundations](#)
- [AWS re:Invent 2022 – Optimizing Amazon EKS for performance and cost on AWS](#)

## Zugehörige Beispiele:

- [Demo-Code für Compute Optimizer](#)
- [Workshop zu Spot Instances von Amazon EC2](#)
- [Effiziente und belastbare Workloads mit Amazon EC2 AWS Auto Scaling](#)
- [Workshop für Graviton-Entwickler](#)
- [AWS für Microsoft-Workloads Immersion Day](#)
- [AWS für Linux-Workloads Immersion Day](#)
- [AWS Compute Optimizer-Demo-Code](#)
- [Workshop zu Amazon EKS](#)

## PERF02-BP03 Erfassung rechnerbezogener Metriken

Erfassen und verfolgen Sie Datenverarbeitungsmetriken, um die Leistung der Rechenressourcen besser zu verstehen und deren Leistung und Auslastung zu verbessern.

Typische Anti-Muster:

- Sie suchen ausschließlich manuell mithilfe von Protokolldateien nach Metriken.
- Sie verwenden nur die Standardmetriken, die von der Überwachungssoftware aufgezeichnet wurden.
- Sie überprüfen Metriken nur dann, wenn ein Problem vorliegt.

Vorteile der Nutzung dieser bewährten Methode: Die Erfassung von Leistungsmetriken hilft Ihnen dabei, die Anwendungsleistung an den Geschäftsanforderungen auszurichten, um sicherzustellen, dass Sie Ihre Workload-Anforderungen erfüllen. Es kann Ihnen auch dabei helfen, die Ressourcenleistung und -nutzung in der Workload kontinuierlich zu verbessern.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

### Implementierungsleitfaden

Cloud-Workloads können große Mengen an Daten generieren, wie Metriken, Protokolle und Ereignisse. In der AWS Cloud ist das Sammeln von Metriken ein entscheidender Schritt zur Verbesserung von Sicherheit, Kosteneffizienz, Leistung und Nachhaltigkeit. AWS bietet eine breite Palette leistungsbezogener Kennzahlen mithilfe von Überwachungsdiensten wie [Amazon](#), CloudWatch um Ihnen wertvolle Einblicke zu bieten. Metriken wie CPU Auslastung, Speicherauslastung, Festplatten-I/O sowie eingehende und ausgehende Netzwerkdaten können Aufschluss über Auslastungsgrade oder Leistungsengpässe geben. Nutzen Sie diese Metriken im Rahmen eines datengestützten Ansatzes, der Ihnen die aktive Feinabstimmung und Optimierung der von der Workload genutzten Ressourcen ermöglicht. Im Idealfall sollten Sie alle Metriken zu Ihren Datenverarbeitungsressourcen auf einer einzigen Plattform erfassen und Aufbewahrungsrichtlinien implementieren, um Kosten- und Betriebsziele zu unterstützen.

### Implementierungsschritte

- Identifizieren Sie, welche Leistungsmetriken für Ihre Workload relevant sind. Sie sollten Metriken zur Ressourcennutzung und zum Betrieb der Cloud-Workload (wie Reaktionszeit und Durchsatz) erfassen.

- [EC2Amazon-Standardmetriken](#)
- [ECSAmazon-Standardmetriken](#)
- [EKSAmazon-Standardmetriken](#)
- [Lambda-Standardmetriken](#)
- [EC2Speicher- und Festplattenmetriken von Amazon](#)
- Wählen Sie die richtige Protokollierungs- und Überwachungslösung für Ihre Workload aus und richten Sie sie ein.
  - [AWS -native Beobachtbarkeit](#)
  - [AWS Distribution für OpenTelemetry](#)
  - [Amazon Managed Service for Prometheus](#)
- Definieren Sie den erforderlichen Filter und die erforderliche Aggregation für die Metriken auf der Grundlage Ihrer Workload-Anforderungen.
  - [Quantifizieren Sie benutzerdefinierte Anwendungsmetriken mit Amazon CloudWatch Logs und Metrikfiltern](#)
  - [Sammeln Sie benutzerdefinierte Metriken mit CloudWatch strategischem Tagging von Amazon](#)
- Konfigurieren Sie Richtlinien zur Datenaufbewahrung für Ihre Metriken so, dass sie Ihren Sicherheits- und Betriebszielen entsprechen.
  - [Standardmäßige Datenspeicherung für Metriken CloudWatch](#)
  - [Standardmäßige Datenspeicherung für CloudWatch Protokolle](#)
- Erstellen Sie bei Bedarf Alarme und Benachrichtigungen für Ihre Metriken, damit Sie proaktiv auf leistungsbezogene Probleme reagieren können.
  - [Erstellen Sie mithilfe der CloudWatch Amazon-Anomalieerkennung Alarme für benutzerdefinierte Metriken](#)
  - [Erstellen Sie mit Amazon Metriken und Alarme für bestimmte Webseiten CloudWatch RUM](#)
- Verwenden Sie die Automatisierung, um die Kundendienstmitarbeiter für die Metrik- und Protokollaggregation einzusetzen.
  - [AWS Systems Manager Automatisierung](#)
  - [OpenTelemetrySammler](#)

## Ressourcen

Zugehörige Dokumente:

- [Überwachung und Beobachtbarkeit](#)
- [Bewährte Methoden: Implementierung von Observability mit AWS](#)
- [CloudWatch Amazon-Dokumentation](#)
- [Erfassen Sie mit dem Agenten Metriken und Protokolle von EC2 Amazon-Instances und lokalen Servern CloudWatch](#)
- [Zugreifen auf Amazon CloudWatch Logs für AWS Lambda](#)
- [CloudWatch Logs mit Container-Instances verwenden](#)
- [Veröffentlichen von benutzerdefinierten Metriken](#)
- [AWS Answers: Zentralisierte Protokollierung](#)
- [AWS Dienste, die CloudWatch Metriken veröffentlichen](#)
- [Überwachung von Amazon EKS am AWS Fargate](#)

#### Zugehörige Videos:

- [AWS re:Invent 2023 — \[LAUNCH\] Anwendungsüberwachung für moderne Workloads](#)
- [AWS re:Invent 2023 — Implementierung der Anwendungsbeobachtbarkeit](#)
- [AWS re:Invent 2023 — Aufbau einer effektiven Strategie für Beobachtbarkeit](#)
- [AWS re:Invent 2023 — Nahtlose Beobachtbarkeit mit Distro für AWS OpenTelemetry](#)
- [Leistungsmanagement für Anwendungen aktiviert AWS](#)

#### Zugehörige Beispiele:

- [AWS Immersion Day für Linux-Workloads — Amazon CloudWatch](#)
- [Überwachung von ECS Amazon-Clustern und Containern](#)
- [Überwachung mit CloudWatch Amazon-Dashboards](#)
- [EKSAmazon-Werkstatt](#)

## PERF02-BP04 Konfigurieren und richtiges Dimensionieren von Datenverarbeitungsressourcen

Konfigurieren und passen Sie die Größe der Datenverarbeitungsressourcen so an, dass sie den Leistungsanforderungen der Workloads entsprechen, und vermeiden Sie zu wenig oder zu stark ausgelastete Ressourcen.

Typische Anti-Muster:

- Sie ignorieren Ihre Workload-Leistungsanforderungen, was zu über- oder unterdimensionierten Datenverarbeitungsressourcen führt.
- Sie wählen nur die größte oder kleinste verfügbare Instance für alle Workloads aus.
- Sie verwenden nur eine Instance-Familie, um die Verwaltung zu vereinfachen.
- Sie ignorieren Empfehlungen von AWS Cost Explorer oder Compute Optimizer zur richtigen Dimensionierung.
- Sie bewerten die Workload nicht erneut auf die Eignung neuer Instance-Typen.
- Sie zertifizieren nur eine kleine Anzahl von Instance-Konfigurationen für Ihre Organisation.

Vorteile der Nutzung dieser bewährten Methode: Die richtige Dimensionierung der Datenverarbeitungsressourcen gewährleistet einen optimalen Betrieb in der Cloud, indem eine Über- und Unterdimensionierung von Ressourcen vermieden wird. Die richtige Dimensionierung der Datenverarbeitungsressourcen führt in der Regel zu einer besseren Leistung und einem besseren Kundenerlebnis bei gleichzeitiger Senkung der Kosten.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

### Implementierungsleitfaden

Die richtige Dimensionierung ermöglicht es Organisationen, ihre Cloud-Infrastruktur effizient und kostengünstig zu betreiben und gleichzeitig ihre Geschäftsanforderungen zu erfüllen. Eine übermäßige Bereitstellung von Cloud-Ressourcen kann zu zusätzlichen Kosten führen, während eine unzureichende Bereitstellung zu schlechter Leistung und einem negativen Kundenerlebnis führen kann. AWS bietet Tools wie [AWS Compute Optimizer](#) und [AWS Trusted Advisor](#), die historische Daten nutzen, um Empfehlungen zur richtigen Größe Ihrer Rechenressourcen zu geben.

## Implementierungsschritte

- Wählen Sie einen Instance-Typ, der am besten zu Ihren Anforderungen passt:
  - [How do I choose the appropriate Amazon EC2 instance type for my workload?](#)
  - [Attributbasierte Auswahl von Instance-Typen für Amazon-EC2-Flotte](#)
  - [Erstellen einer Auto-Scaling-Gruppe mit attributbasierter Auswahl des Instance-Typs](#)
  - [Optimieren Ihrer Kubernetes-Datenverarbeitungskosten mit der Karpenter-Konsolidierung](#)
- Analysieren Sie die verschiedenen Leistungsmerkmale Ihrer Workload und bewerten Sie, wie sich diese auf Arbeitsspeicher, Netzwerk und CPU-Auslastung auswirken. Wählen Sie anhand dieser Daten die für das Profil und die Leistungsziele der Workloads am besten geeigneten Ressourcen aus.
- Überwachen Sie Ihren Ressourcenverbrauch mithilfe von AWS-Überwachungstools wie Amazon CloudWatch.
- Wählen Sie die richtige Konfiguration für die Datenverarbeitungsressource aus.
  - Prüfen Sie für kurzzeitige Workloads die [Amazon-CloudWatch-Metriken der Instance](#) wie CPUUtilization, um festzustellen, ob die Instance zu wenig oder zu viel genutzt wird.
  - Prüfen Sie für stabile Workloads in regelmäßigen Intervallen AWS-Dimensionierungstools wie etwa AWS Compute Optimizer und AWS Trusted Advisor, um Möglichkeiten zur Optimierung und zur korrekten Dimensionierung der Datenverarbeitungsressource zu erkennen.
- Testen Sie Konfigurationsänderungen in einer Nicht-Produktionsumgebung, bevor Sie sie in einer Live-Umgebung implementieren.
- Bewerten Sie neue Datenverarbeitungsangebote und vergleichen Sie sie mit den Anforderungen Ihrer Workload.

## Ressourcen

Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon-EC2-Instance-Typen](#)
- [Amazon ECS Containers: Amazon ECS Container Instances](#)
- [Amazon EKS Containers: Amazon EKS Worker Nodes](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)
- [Processor State Control for Your Amazon EC2 Instance](#)

## Zugehörige Videos:

- [Grundlagen von Amazon EC2](#)
- [AWS re:Invent 2023 – AWS Graviton: Das beste Preis-Leistungs-Verhältnis für Ihre AWS-Workloads](#)
- [AWS re:Invent 2023 – New Amazon EC2 generative AI capabilities in AWS-Managementkonsole](#)
- [AWS re:Invent 2023 – What's new with Amazon EC2](#)
- [AWS re:Invent 2023 – Smart savings: Amazon EC2 cost-optimization strategies](#)
- [AWS re:Invent 2021 – Powering next-gen Amazon EC2: Deep dive on the Nitro System](#)
- [AWS re:Invent 2019 – Amazon EC2 foundations](#)

## Zugehörige Beispiele:

- [AWS Compute Optimizer-Demo-Code](#)
- [Workshop zu Amazon EKS](#)
- [Empfehlungen zur Dimensionierung](#)

# PERF02-BP05 Dynamisches Skalieren von Datenverarbeitungsressourcen

Nutzen Sie die Elastizität der Cloud, um die Datenverarbeitungsressourcen dynamisch nach oben oder unten zu skalieren, um Ihren Bedürfnissen zu entsprechen und eine Über- oder Unterdimensionierung von Kapazitäten für die Workload zu vermeiden.

## Typische Anti-Muster:

- Sie reagieren auf Alarme, indem Sie die Kapazität manuell erhöhen.
- Sie verwenden dieselben Dimensionierungsrichtlinien (in der Regel statische Infrastruktur) wie bei On-Premises.
- Sie belassen die erhöhte Kapazität nach dem Hochskalieren, anstatt wieder herunterzuskalieren.

Vorteile der Nutzung dieser bewährten Methode: Durch das Konfigurieren und Testen der Elastizität von Datenverarbeitungsressourcen können Sie Geld sparen, Leistungsbenchmarks einhalten und die Zuverlässigkeit verbessern, wenn sich der Datenverkehr ändert.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

AWS bietet Ihnen die Flexibilität, Ressourcen dynamisch durch verschiedene Skalierungsmechanismen nach oben oder unten zu skalieren, um Bedarfsänderungen gerecht zu werden. In Kombination mit Datenverarbeitungsmetriken ermöglicht eine dynamische Skalierung Workloads, automatisch auf Änderungen zu reagieren und die optimalen Datenverarbeitungsressourcen zu nutzen, um die Zielvorgabe zu erreichen.

Sie können verschiedene Ansätze nutzen, um das Angebot an Ressourcen auf die Nachfrage abzustimmen.

- Ansatz zur Zielverfolgung: Überwachen Sie Ihre Skalierungsmetriken und erhöhen oder verringern Sie die Kapazität automatisch Ihrem Bedarf entsprechend.
- Prädiktive Skalierung: Skalieren Sie in Erwartung täglicher und wöchentlicher Trends.
- Zeitplanbasierter Ansatz: Legen Sie Ihren eigenen Skalierungszeitplan entsprechend vorhersehbaren Laständerungen fest.
- Skalierung von Services: Wählen Sie Services (wie Serverless), die auf automatische Skalierung ausgelegt sind.

Sie müssen sicherstellen, dass Workload-Bereitstellungen sowohl Hoch- als auch Herunterskalierungsereignisse verarbeiten können.

## Implementierungsschritte

- Datenverarbeitungs-Instances, Container und Funktionen bieten Mechanismen für Elastizität, sei es in Kombination mit AutoScaling oder als Feature des Service. Hier finden Sie einige Beispiele für automatische Skalierungsmechanismen:

Autoscaling-Mechanismus	Aktion
<a href="#">Amazon EC2 Auto Scaling</a>	Zur Sicherstellung, dass die richtige Anzahl verfügbarer <a href="#">Amazon-EC2</a> -Instances vorhanden ist, um die Benutzerlast für Ihre Anwendung zu bewältigen.

Autoscaling-Mechanismus	Aktion
<a href="#">Application Auto Scaling</a>	Zur automatischen Skalierung der Ressourcen für einzelne AWS-Services über Amazon EC2 hinaus, wie <a href="#">AWS Lambda</a> -Funktionen oder <a href="#">Amazon Elastic Container Service (Amazon ECS)</a> -Services.
<a href="#">Kubernetes Cluster Autoscaler/Karpenter</a>	Zur automatischen Skalierung von Kubernetes-Clustern.

- Skalieren wird häufig im Zusammenhang mit Datenverarbeitungs-Services wie Amazon-EC2-Instances oder AWS Lambda-Funktionen behandelt. Denken Sie auch daran, die Konfiguration von nicht Daten verarbeitenden Services in Betracht zu ziehen, z. B. [AWS Glue](#), um die Nachfrage zu decken.
- Stellen Sie sicher, dass die Metriken für die Skalierung den Merkmalen der bereitgestellten Workload entsprechen. Wenn Sie eine Anwendung zur Video-Transkodierung bereitstellen, wird eine CPU-Auslastung von 100 % erwartet, weshalb dies nicht die Hauptmetrik sein sollte. Verwenden Sie stattdessen die Tiefe der Aufgabenwarteschlange für die Transkodierung. Sie können bei Bedarf eine [benutzerdefinierte Metrik](#) für Ihre Skalierungsrichtlinie verwenden. Beachten Sie bei der Wahl geeigneter Metriken die folgenden Hinweise für Amazon EC2:
  - Es muss sich um eine gültige Nutzungsmetrik handeln, die beschreibt, wie stark eine Instance genutzt wird.
  - Der Wert der Metrik muss sich proportional zur Anzahl der Instances in der Auto-Scaling-Gruppe erhöhen oder verringern.
- Achten Sie darauf, für Ihre Auto-Scaling-Gruppe eine [dynamische Skalierung](#) anstelle einer [manuellen Skalierung](#) zu verwenden. Außerdem empfiehlt es sich, bei der dynamischen Skalierung [Skalierungsrichtlinien zur Zielverfolgung](#) zu verwenden.
- Prüfen Sie, ob Workload-Bereitstellungen mit beiden Skalierungen (nach oben und unten) umgehen können. Sie können beispielsweise den [Aktivitätsverlauf](#) verwenden, um eine Skalierungsaktivität für eine Auto-Scaling-Gruppe zu überprüfen.
- Evaluieren Sie Ihre Workload auf vorhersagbare Muster und skalieren Sie proaktiv, wenn Sie vorhergesagte und geplante Änderungen der Nachfrage erwarten. Mit der prädiktiven Skalierung können Sie die Notwendigkeit einer Überbereitstellung von Kapazitäten vermeiden. Weitere Einzelheiten finden Sie unter [Prädiktive Skalierung mit Amazon EC2 Auto Scaling](#).

# Ressourcen

## Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon-EC2-Instance-Typen](#)
- [Amazon ECS Containers: Amazon ECS Container Instances](#)
- [Amazon EKS Containers: Amazon EKS Worker Nodes](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)
- [Processor State Control for Your Amazon EC2 Instance](#)
- [Deep Dive on Amazon ECS Cluster Auto Scaling](#)
- [Vorstellung von Karpenter – Open-Source-Kubernetes-Cluster-Autoscaler mit hoher Leistung](#)

## Zugehörige Videos:

- [AWS re:Invent 2023 – AWS Graviton: Das beste Preis-Leistungs-Verhältnis für Ihre AWS-Workloads](#)
- [AWS re:Invent 2023 – New Amazon EC2 generative AI capabilities in AWS Management Console](#)
- [AWS re:Invent 2023 – What’s new with Amazon EC2](#)
- [AWS re:Invent 2023 – Smart savings: Amazon EC2 cost-optimization strategies](#)
- [AWS re:Invent 2021 – Powering next-gen Amazon EC2: Deep dive on the Nitro System](#)
- [AWS re:Invent 2019 – Amazon EC2 foundations](#)

## Zugehörige Beispiele:

- [Beispiele für Amazon EC2 Auto Scaling-Gruppe](#)
- [Workshop zu Amazon EKS](#)
- [Skalieren von Amazon-EKS-Workloads durch die Ausführung auf IPv6](#)

# PERF02-BP06 Verwenden von optimierten hardwarebasierten Datenverarbeitungsbeschleunigern

Verwenden Sie Hardwarebeschleuniger, um bestimmte Funktionen effizienter auszuführen als CPU-basierte Alternativen.

Typische Anti-Muster:

- Sie haben in der Workload keine Benchmark einer Allzweck-Instance verglichen mit einer speziell entwickelten Instance durchgeführt, die eine höhere Leistung und niedrigere Kosten bieten kann.
- Sie verwenden hardwarebasierte Datenverarbeitungsbeschleuniger für Aufgaben, die mithilfe von CPU-basierten Alternativen effizienter sein können.
- Sie überwachen die GPU-Nutzung nicht.

Vorteile der Nutzung dieser bewährten Methode: Durch die Verwendung hardwarebasierter Beschleuniger wie Grafikprozessoren (GPUs) und Field-Programmable Gate Arrays (FPGAs) können Sie bestimmte Verarbeitungsfunktionen effizienter ausführen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Beschleunigte Computing-Instances bieten Zugriff auf hardwarebasierte Datenverarbeitungsbeschleuniger wie GPUs und FPGAs. Diese Hardwarebeschleuniger führen bestimmte Funktionen wie die Grafikverarbeitung oder Datenmusterzuordnung effizienter aus als CPU-basierte Alternativen. Viele beschleunigte Workloads, wie Rendering, Transkodierung und Machine Learning, sind sehr variabel im Bezug auf die Ressourcennutzung. Betreiben Sie diese Hardware nur so lange wie nötig und nehmen Sie sie automatisch außer Betrieb, wenn sie nicht mehr benötigt wird, um die allgemeine Leistungseffizienz zu verbessern.

## Implementierungsschritte

- Ermitteln Sie, welche [beschleunigten Computing-Instances](#) für Ihre Anforderungen geeignet sind.
- Nutzen Sie für Machine-Learning-Workloads spezielle Hardware, die auf Ihre Workload abgestimmt ist, z. B. [AWS Trainium](#), [AWS Inferentia](#) und [Amazon EC2 DL1](#). AWS Inferentia-Instances wie Inf2-Instances [bieten eine um bis zu 50 % bessere Leistung pro Watt als vergleichbare Amazon-EC2-Instances](#).

- Erfassen Sie Nutzungsmetriken für Ihre beschleunigten Computing-Instances. Sie können den CloudWatch-Agent beispielsweise verwenden, um Metriken wie `utilization_gpu` und `utilization_memory` für Ihre GPUs zu sammeln, wie im [Artikel zum Erfassen von NVIDIA-GPU-Metriken mit Amazon CloudWatch](#) beschrieben.
- Optimieren Sie Code, Netzwerkbetrieb und die Einstellungen von Hardwarebeschleunigern, um sicherzustellen, dass die zugrunde liegende Hardware optimal genutzt wird.
  - [Optimieren der GPU-Einstellungen](#)
  - [GPU-Überwachung und -Optimierung im Deep-Learning-AMI](#)
  - [Optimieren von E/A für die GPU-Leistungsoptimierung von Deep-Learning-Training in Amazon SageMaker AI](#)
- Verwenden Sie die aktuellen leistungsstarken Bibliotheken und GPU-Treiber.
- Automatisieren Sie die Freigabe nicht genutzter GPU-Instances.

## Ressourcen

Zugehörige Dokumente:

- [Working with GPUs on Amazon Elastic Container Service](#)
- [GPU-Instances](#)
- [Instances mit AWS Trainium](#)
- [Instances mit AWS Inferentia](#)
- [Let's Architect! Erstellen von Architekturen mit benutzerdefinierten Chips und Beschleunigern](#)
  
- [Beschleunigte Datenverarbeitung](#)
- [Amazon EC2 VT1 Instances](#)
- [How do I choose the appropriate Amazon EC2 instance type for my workload?](#)
- [Choose the best AI accelerator and model compilation for computer vision inference with Amazon SageMaker AI](#)

Zugehörige Videos:

- AWS re:Invent 2021 – [How to select Amazon Elastic Compute Cloud GPU instances for deep learning](#)

- [AWS re:Invent 2022 – \[NEW LAUNCH!\] Introducing AWS Inferentia2-based Amazon EC2 Inf2 instances](#)
- [AWS re:Invent 2022 – Accelerate deep learning and innovate faster with AWS Trainium](#)
- [AWS re:Invent 2022 – Deep learning on AWS with NVIDIA: From training to deployment](#)

Zugehörige Beispiele:

- [Amazon SageMaker AI und NVIDIA GPU Cloud \(NGC\)](#)
- [Verwendung von SageMaker AI mit Trainium und Inferentia für optimierte Workloads für Deep-Learning-Training und Inferenzierung](#)
- [Optimierung von NLP-Modellen mit Instances für Amazon Elastic Compute Cloud Inf1 in Amazon SageMaker AI](#)

# Datenverwaltung

Die optimale Datenverwaltungslösung für ein bestimmtes System hängt vom Datentyp (Block, Datei oder Objekt), den Zugriffsmustern (zufällig oder sequenziell), dem erforderlichen Durchsatz, der Zugriffshäufigkeit (online, offline, Archiv), der Aktualisierungshäufigkeit (WORM, dynamisch) sowie den Verfügbarkeits- und Lebensdauernanforderungen ab. Well-Architected-Workloads verwenden zweckgebundene Datenspeicher, die verschiedene Features zur Verbesserung der Leistung ermöglichen.

In diesem Schwerpunktbereich werden Anleitungen und bewährte Methoden zur Optimierung der Datenspeicherung, der Bewegungs- und Zugriffsmuster sowie der Leistungseffizienz von Datenspeichern bereitgestellt.

## Bewährte Methoden

- [PERF03-BP01 Verwenden eines speziell entwickelten Datenspeichers, der die Datenzugriffs- und Speicheranforderungen am besten unterstützt](#)
- [PERF03-BP02 Bewerten verfügbarer Konfigurationsoptionen für den Datenspeicher](#)
- [PERF03-BP03 Erfassen und Aufzeichnen von Metriken zur Datenspeicherleistung](#)
- [PERF03-BP04 Implementieren von Strategien zur Verbesserung der Abfrageleistung im Datenspeicher](#)
- [PERF03-BP05 Implementieren von Datenzugriffsmustern, die Caching nutzen](#)

## PERF03-BP01 Verwenden eines speziell entwickelten Datenspeichers, der die Datenzugriffs- und Speicheranforderungen am besten unterstützt

Machen Sie sich mit Datenmerkmalen (wie Freigabe, Größe, Cache-Größe, Zugriffsmuster, Latenz, Durchsatz und Persistenz von Daten) vertraut, um die richtigen, speziell entwickelten Datenspeicher (Speicher oder Datenbank) für die Workload auszuwählen.

### Typische Anti-Muster:

- Sie halten an einem Datenspeicher fest, da es interne Erfahrungen und Wissen über eine bestimmte Datenbanklösung gibt.

- Sie gehen davon aus, dass für alle Workloads ähnliche Datenspeicher- und Zugriffsanforderungen gelten.
- Sie haben keinen Datenkatalog zur Inventarisierung Ihrer Datenbestände eingeführt.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie die Datenmerkmale und -anforderungen verstehen, können Sie die effizienteste und leistungsfähigste Speichertechnologie ermitteln, die für Ihre Workload-Anforderungen geeignet ist.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

Stellen Sie bei der Auswahl und Implementierung von Datenspeicher sicher, dass die Abfrage-, Skalierungs- und Speichermerkmale die Workload-Datenanforderungen unterstützen. AWS bietet zahlreiche Datenspeicher- und Datenbanktechnologien, darunter Blockspeicher, Objektspeicher, Streaming-Speicher, Dateisystem-, relationale, Schlüsselwert-, Dokument-, In-Memory-, Graph-, Zeitreihen- und Ledger-Datenbanken. Jede Datenverwaltungslösung hat verfügbare Optionen und Konfigurationen, um Ihre Anwendungsfälle und Datenmodelle zu unterstützen. Wenn Sie die Merkmale und Anforderungen der Daten verstehen, können Sie sich von monolithischer Speichertechnologie und restriktiven Einheitsansätzen lösen und sich auf eine angemessene Datenverwaltung konzentrieren.

## Implementierungsschritte

- Führen Sie eine Bestandsaufnahme der verschiedenen Datentypen durch, die in Ihrer Workload vorhanden sind.
- Verstehen und dokumentieren Sie Datenmerkmale und -anforderungen, einschließlich:
  - Datentyp (strukturiert, semistrukturiert, relational)
  - Datenvolumen und -wachstum
  - Lebensdauer von Daten: anhaltend, flüchtig, vorübergehend
  - Anforderungen an AKID (Atomarität, Konsistenz, Isolation, Dauerhaftigkeit)
  - Datenzugriffsmuster (leseintensiv oder schreibintensiv)
  - Latenz
  - Durchsatz
  - IOPS (Eingabe-/Ausgabevorgänge pro Sekunde)

- Aufbewahrungszeitraum
- Erfahren Sie mehr über die verschiedenen Datenspeicher ([Speicher-](#) und [Datenbank-](#)Services), die für Ihre Workload in AWS verfügbar sind und Ihre Datenmerkmale erfüllen können (wie beschrieben unter [PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features](#)). Einige Beispiele für AWS-Speichertechnologien und ihre Schlüsselmerkmale sind:

Typ	AWS-Services	Schlüsselmerkmale
Objektspeicher	<a href="#">Amazon S3</a>	Unbegrenzte Skalierbarkeit, hohe Verfügbarkeit und mehrere Optionen für Barrierefreiheit. Für die Übertragung von Objekten in und aus Amazon S3 und den Zugriff auf diese Objekte können Sie einen Service wie z. B. <a href="#">Transfer Acceleration</a> oder <a href="#">Zugangspunkte</a> verwenden, um Ihren Standort, Ihre Sicherheitsanforderungen und Zugriffsmuster zu unterstützen.
Archivieren von Speichern	<a href="#">Amazon Glacier</a>	Für die Datenarchivierung entwickelt.
Streaming-Speicher	<a href="#">Amazon Kinesis</a> <a href="#">Amazon Managed Streaming für Apache Kafka (Amazon MSK)</a>	Effiziente Erfassung und Speicherung von Streaming-Daten.
Gemeinsames Dateisystem	<a href="#">Amazon Elastic File System (Amazon EFS)</a>	Bereitstellbares Dateisystem, auf das mehrere Arten von Datenverarbeitungslösungen zugreifen können.

Typ	AWS-Services	Schlüsselmerkmale
Gemeinsames Dateisystem	<a href="#">Amazon FSx</a>	Basiert auf den aktuellen AWS-Datenverarbeitungslösungen und unterstützt vier gängige Dateisysteme: NetApp ONTAP, OpenZFS, Windows File Server und Lustre. <a href="#">Die Latenz, der Durchsatz und die IOPS</a> von Amazon FSx variieren je nach Dateisystem und sollten bei der Auswahl des richtigen Dateisystems für Ihre Workload-Anforderungen berücksichtigt werden.
Blockspeicher	<a href="#">Amazon Elastic Block Store (Amazon EBS)</a>	Skalierbarer, hochleistungsfähiger Blockspeicherservice für Amazon Elastic Compute Cloud (Amazon EC2). Amazon EBS umfasst SSD-gestützten Speicher für transaktions- und IOPS-intensive Workloads und HDD-gestützten Speicher für durchsatzintensive Workloads.

Typ	AWS-Services	Schlüsselmerkmale
Relationale Datenbank	<a href="#">Amazon Aurora</a> , <a href="#">Amazon RDS</a> , <a href="#">Amazon Redshift</a> .	Sie unterstützt AKID-Transaktionen (Atomarität, Konsistenz, Isolation und Dauerhaftigkeit) und gewährleistet die referentielle Integrität sowie eine starke Datenkonsistenz. Bei zahlreichen herkömmlichen Anwendungen, Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) und E-Commerce werden relationale Datenbanken zum Speichern der Daten verwendet.
Schlüssel-Werte-Datenbank	<a href="#">Amazon DynamoDB</a>	Für gängige Zugriffsmuster optimiert, üblicherweise zum Speichern und Abrufen großer Datenmengen. Web-Apps mit hohem Datenverkehr, E-Commerce-Systeme und Gaming-Anwendungen sind typische Anwendungsfälle für Schlüssel-Werte-Datenbanken.

Typ	AWS-Services	Schlüsselmerkmale
Dokumentdatenbank	<a href="#">Amazon DocumentDB</a>	Dies wurde entwickelt, um halbstrukturierte Daten als JSON-ähnliche Dokumente zu speichern. Mit diesen Datenbanken können Entwickler Anwendungen wie Content Management, Kataloge und Benutzerprofile schnell erstellen und aktualisieren.
In-Memory-Datenbanken	<a href="#">Amazon ElastiCache</a> , <a href="#">Amazon MemoryDB für Redis</a>	Dies wird für Anwendungen eingesetzt, die einen Echtzeitzugriff auf Daten, die niedrigste Latenz und den höchsten Durchsatz erfordern. Sie können In-Memory-Datenbanken für Anwendungs-Caching , Sitzungsmanagement, Gaming-Bestenlisten, ML-Feature-Store mit niedriger Latenz, Microservices-Messaging-System und einen Streaming-Mechanismus mit hohem Durchsatz verwenden

Typ	AWS-Services	Schlüsselmerkmale
Graphdatenbank	<a href="#">Amazon Neptune</a>	Dies ist für Anwendungen gedacht, die in Millionen von Beziehungen zwischen hochgradig vernetzten Diagrammdatensätzen mit Millisekunden-Latenz navigieren und diese abfragen müssen. Viele Unternehmen verwenden Graphdatenbanken für Betrugserkennung, soziale Netzwerke und Empfehlungs-Engines.
Zeitreihendatenbank	<a href="#">Amazon Timestream</a>	Diese erfasst, generiert und gewinnt auf effiziente Weise Einblicke aus Daten, die sich im Laufe der Zeit ändern. IoT-Anwendungen, DevOps und industrielle Telemetrie können Zeitreihendatenbanken nutzen.

Typ	AWS-Services	Schlüsselmerkmale
Wide-Column-Datenbanken	<a href="#">Amazon Keyspaces (für Apache Cassandra)</a>	Es werden Tabellen, Zeilen und Spalten verwendet, aber im Gegensatz zu einer relationalen Datenbank können sich die Namen und das Format der Spalten von Zeile zu Zeile in derselben Tabelle unterscheiden. In der Regel werden Wide Column-Speicher in umfangreichen Branchen-Apps für Gerätewartung, Flottenverwaltung und Routenoptimierung eingesetzt.
Ledger	<a href="#">Amazon Quantum Ledger Database (Amazon QLDB)</a>	Dies bietet eine zentrale und vertrauenswürdige Instanz für die Verwaltung einer skalierbaren, unveränderlichen und kryptografisch überprüfbaren Aufzeichnung von Transaktionen für jede Anwendung. Ledger-Datenbanken werden für Datensatzsysteme, Lieferketten, Registrierungen und sogar Banktransaktionen verwendet.

- Wenn Sie eine Datenplattform aufbauen, nutzen Sie eine [moderne Datenarchitektur](#) in AWS, um Ihren Data Lake, Ihr Data Warehouse und Ihre speziell entwickelten Datenspeicher zu integrieren.
- Die wichtigsten Fragen, die Sie bei der Auswahl eines Datenspeichers für Ihre Workload berücksichtigen müssen, lauten wie folgt:

Frage	Worauf Sie achten sollten
Wie sind die Daten strukturiert?	<ul style="list-style-type: none"><li>• Wenn die Daten nicht strukturiert sind, erwägen Sie einen Objektspeicher wie <a href="#">Amazon S3</a> oder eine NoSQL-Datenbank wie <a href="#">Amazon DocumentDB</a>.</li><li>• Erwägen Sie für Schlüssel-Werte-Daten <a href="#">DynamoDB</a>, <a href="#">Amazon ElastiCache (Redis OSS)</a> oder <a href="#">Amazon MemoryDB</a>.</li></ul>
Welches Maß an referentieller Integrität ist erforderlich?	<ul style="list-style-type: none"><li>• Bei Fremdschlüsseinschränkungen können relationale Datenbanken wie <a href="#">Amazon RDS</a> und <a href="#">Aurora</a> diese Integritätsebene bieten.</li><li>• Üblicherweise würden Sie innerhalb eines NoSQL-Datenmodells Ihre Daten in ein einzelnes Dokument oder eine Sammlung von Dokumenten denormalisieren, die in einer einzelnen Anfrage abgerufen werden können, anstatt Daten in Dokumenten oder Tabellen zusammenzufügen.</li></ul>
Ist AKID-Compliance (Atomarität, Konsistenz, Isolation, Dauerhaftigkeit) erforderlich?	<ul style="list-style-type: none"><li>• Wenn mit relationalen Datenbanken zusammenhängende AKID-Eigenschaften erforderlich sind, erwägen Sie eine relationale Datenbank wie <a href="#">Amazon RDS</a> und <a href="#">Aurora</a>.</li><li>• Wenn eine starke Konsistenz für eine <a href="#">NoSQL-Datenbank</a> erforderlich ist, können Sie strikt konsistente Lesevorgänge mit <a href="#">DynamoDB</a> verwenden.</li></ul>

Frage	Worauf Sie achten sollten
<p>Wie ändern sich die Speicheranforderungen im Laufe der Zeit? Wie beeinflusst dies die Skalierbarkeit?</p>	<ul style="list-style-type: none"> <li>• Serverless-Datenbanken wie <a href="#">DynamoDB</a> und <a href="#">Amazon Quantum Ledger Database (Amazon QLDB)</a> skalieren dynamisch.</li> <li>• Relationale Datenbanken haben oftmals Obergrenzen bei bereitgestelltem Speicher und müssen mithilfe von Mechanismen wie Sharding horizontal partitioniert werden, sobald sie diese Grenzen erreicht haben.</li> </ul>
<p>Wie hoch ist der Anteil der Leseabfragen im Verhältnis zu den Schreibabfragen? Könnte Caching die Leistung verbessern?</p>	<ul style="list-style-type: none"> <li>• Leseintensive Workloads könnten von einer Caching-Ebene, wie <a href="#">ElastiCache</a> oder <a href="#">DAX</a> profitieren, wenn es sich bei der Datenbank um DynamoDB handelt.</li> <li>• Lesevorgänge können auch zu Read Replicas mit relationalen Datenbanken ausgelagert werden, wie <a href="#">Amazon RDS</a>.</li> </ul>
<p>Wird Speicher und Modifizierung (OLTP – Online Transaction Processing) oder Abruf und Berichterstattung (OLAP – Online Analytical Processing) eine höhere Priorität eingeräumt?</p>	<ul style="list-style-type: none"> <li>• Erwägen Sie für Read-as-is-Transaktionsverarbeitung mit hohem Durchsatz eine NoSQL-Datenbank wie DynamoDB.</li> <li>• Verwenden Sie Amazon RDS für hohen Durchsatz und komplexe Lesemuster (wie Join) mit Konsistenz.</li> <li>• Erwägen Sie für analytische Abfragen eine spaltenbasierte Datenbank wie <a href="#">Amazon Redshift</a> oder das Exportieren von Daten zu Amazon S3 und das Durchführen von Analytik mithilfe von <a href="#">Athena</a> oder <a href="#">Amazon Quick</a>.</li> </ul>

Frage	Worauf Sie achten sollten
Welches Ausmaß an Stabilität erfordern die Daten?	<ul style="list-style-type: none"><li>• Aurora repliziert Ihre Daten automatisch in drei Availability Zones innerhalb einer Region, was bedeutet, dass Ihre Daten hochbeständig sind und eine geringere Wahrscheinlichkeit von Datenverlust besteht.</li><li>• DynamoDB wird automatisch in mehreren Availability Zones repliziert und bietet hohe Verfügbarkeit und Datenstabilität.</li><li>• Amazon S3 bietet eine Langlebigkeit mit 11 Neunen. Viele Datenbankservices wie Amazon RDS und DynamoDB unterstützen das Exportieren von Daten zu Amazon S3 für Langzeitaufbewahrung und Archivierung.</li></ul>
Besteht der Wunsch, sich von kommerziellen Datenbank-Engines oder Lizenzkosten zu entfernen?	<ul style="list-style-type: none"><li>• Ziehen Sie Open-Source-Engines wie PostgreSQL und MySQL auf Amazon RDS oder Aurora in Erwägung.</li><li>• Nutzen Sie <a href="#">AWS Database Migration Service</a> und <a href="#">AWS Schema Conversion Tool</a> zum Migrieren von kommerziellen Datenbank-Engines zu Open Source-Lösungen.</li></ul>
Was ist die Betriebserwartung an die Datenbank? Ist der Umstieg zu verwalteten Services eine Priorität?	<ul style="list-style-type: none"><li>• Das Verwenden von Amazon RDS anstatt von Amazon EC2 und DynamoDB oder Amazon DocumentDB anstatt eine NoSQL-Datenbank selbst zu hosten, kann den Betriebsaufwand verringern.</li></ul>

Frage	Worauf Sie achten sollten
Wie erfolgt derzeit der Zugriff auf die Datenbank? Handelt es sich nur um einen Anwendungszugriff oder gibt es Business-Intelligence (BI)-Benutzer und andere Standardanwendungen?	<ul style="list-style-type: none"><li>• Wenn Sie von externen Tools abhängig sind, müssen Sie möglicherweise mit der Datenbank, die unterstützt wird, die Kompatibilität aufrecht erhalten. Amazon RDS ist vollständig kompatibel mit den unterschiedlichen Engine-Versionen, die unterstützt werden, einschließlich Microsoft SQL Server, Oracle, MySQL und PostgreSQL.</li></ul>

- Führen Sie Experimente und Benchmarking in einer Nicht-Produktionsumgebung durch, um herauszufinden, welcher Datenspeicher Ihre Workload-Anforderungen erfüllen kann.

## Ressourcen

Zugehörige Dokumente:

- [Typen von Amazon-EBS-Volumes](#)
- [Amazon-EC2-Speicher](#)
- [Amazon EFS: Amazon EFS-Performance](#)
- [Leistung von Amazon FSx für Lustre](#)
- [Leistung von Amazon FSx für Windows File Server](#)
- [Amazon Glacier: Glacier-Dokumentation](#)
- [Amazon S3: Überlegungen zu Anfragerate und Leistung](#)
- [Cloud-Speicher mit AWS](#)
- [Amazon EBS I/O Characteristics](#)
- [Cloud-Datenbanken mit AWS](#)
- [AWS Datenbank-Caching](#)
- [DynamoDB Accelerator](#)
- [Bewährte Methoden mit Amazon Aurora](#)
- [Leistung von Amazon Redshift](#)

- [Die zehn besten Leistungstipps für Amazon Athena](#)
- [Bewährte Methoden für Amazon Redshift Spectrum](#)
- [Bewährte Methoden für Amazon DynamoDB](#)
- [Choose between Amazon EC2 and Amazon RDS](#)
- [Bewährte Methoden für die Implementierung von Amazon ElastiCache](#)

#### Zugehörige Videos:

- [AWS re:Invent 2023: Improve Amazon Elastic Block Store efficiency and be more cost-efficient](#)
- [AWS re:Invent 2023: Optimizing storage price and performance with Amazon Simple Storage Service](#)
- [AWS re:Invent 2023: Building and optimizing a data lake on Amazon Simple Storage Service](#)
- [AWS re:Invent 2022: Building modern data architectures on AWS](#)
- [AWS re:Invent 2022: Building data mesh architectures on AWS](#)
- [AWS re:Invent 2023: Deep dive into Amazon Aurora and its innovations](#)
- [AWS re:Invent 2023: Advanced data modeling with Amazon DynamoDB](#)
- [AWS re:Invent 2022: Modernize apps with purpose-built databases](#)
- [Ausführliche Beschreibung von Amazon DynamoDB: Erweiterte Entwurfsmuster](#)

#### Zugehörige Beispiele:

- [AWS Workshop „Speziell entwickelte Datenbanken](#)
- [Databases for Developers](#)
- [AWS Modern Data Architecture Immersion Day](#)
- [Build a Data Mesh on AWS](#)
- [Amazon-S3-Beispiele](#)
- [Optimierung von Datenmustern mithilfe von Amazon Redshift Data Sharing](#)
- [Datenbankmigrationen](#)
- [MS SQL Server – AWS Database Migration Service \(AWS DMS\)-Replikationsdemo](#)
- [Praktischer Workshop zur Datenbankmodernisierung](#)
- [Beispiele zu Amazon Neptune](#)

## PERF03-BP02 Bewerten verfügbarer Konfigurationsoptionen für den Datenspeicher

Machen Sie sich mit den verschiedenen Features und Konfigurationsoptionen vertraut, die für Ihre Datenspeicher verfügbar sind, und bewerten Sie sie, um Speicherplatz und Leistung für Ihre Workload zu optimieren.

Typische Anti-Muster:

- Sie verwenden nur einen Speichertyp, z. B. Amazon EBS, für alle Workloads.
- Sie verwenden bereitgestellte IOPS für alle Workloads, ohne reale Tests auf allen Speicherebenen durchzuführen.
- Ihnen fehlt das Bewusstsein für die Wahl der Konfigurationsoptionen der Datenverwaltungslösung.
- Sie verlassen sich ausschließlich auf das Vergrößern der Instance-Größe, ohne andere verfügbare Konfigurationsoptionen in Betracht zu ziehen.
- Sie testen die Skalierungsoptionen Ihres Datenspeichers nicht.

Vorteile der Nutzung dieser bewährten Methode: Indem Sie Datenspeicherkonfigurationen erkunden und mit ihnen experimentieren, können Sie möglicherweise Infrastrukturkosten senken, die Leistung verbessern und den Aufwand zur Verwaltung Ihrer Workloads verringern.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

### Implementierungsleitfaden

Für eine Workload können je nach Datenspeicher- und Zugriffsanforderungen ein oder mehrere Datenspeicher verwendet werden. Zur Optimierung der Leistungseffizienz und Kosten müssen Sie Datenzugriffsmuster auswerten, um die entsprechenden Datenspeicherkonfigurationen zu bestimmen. Während Sie die Datenspeicheroptionen erkunden, sollten Sie unterschiedliche Aspekte in Betracht ziehen. Dazu zählen Speicheroptionen, Arbeitsspeicher, Rechenvorgänge, Read Replica, Konsistenzanforderungen, Verbindungs-Pooling und Caching-Optionen. Experimentieren Sie mit diesen unterschiedlichen Konfigurationsoptionen, um Metriken zur Leistungseffizienz zu verbessern.

### Implementierungsschritte

- Verstehen Sie die aktuellen Konfigurationen (wie Instance-Typ, Speichergröße oder Version der Datenbank-Engine) des Datenspeichers.

- Lesen Sie die AWS-Dokumentation und die bewährten Methoden, um mehr über empfohlene Konfigurationsoptionen zu erfahren, mit denen Sie die Leistung für den Datenspeicher verbessern können. Die wichtigsten Datenspeicheroptionen, die Sie in Betracht ziehen sollten, sind die folgenden:

Konfigurationsoption	Beispiele
Auslagern von Lesevorgängen (wie Read Replicas und Caching)	<ul style="list-style-type: none"><li>• Bei DynamoDB-Tabellen können Sie Lesevorgänge mithilfe von DAX für Caching auslagern.</li><li>• Sie können einen Cluster für Amazon ElastiCache (Redis OSS) erstellen und Ihre Anwendung so konfigurieren, dass sie zuerst aus dem Cache liest und dann auf die Datenbank zurückfällt, wenn das angeforderte Element nicht vorhanden ist.</li><li>• Relationale Datenbanken wie Amazon RDS und Aurora sowie bereitgestellte NoSQL-Datenbanken wie Neptune und Amazon DocumentDB unterstützen alle das Hinzufügen von Read Replicas, um die Lesevorgänge der Workload auszulagern.</li><li>• Serverless-Datenbanken wie DynamoDB skalieren automatisch. Stellen Sie sicher, dass Sie ausreichend Read Capacity Units (RCU) bereitstellen, um die Workload zu verarbeiten.</li></ul>

Konfigurationsoption	Beispiele
Skalieren von Schreibvorgängen (wie Partitionsschlüssel-Sharding oder Einführung einer Warteschlange)	<ul style="list-style-type: none"><li>• Bei relationalen Datenbanken können Sie die Größe der Instance erhöhen, um eine erhöhte Workload zu bewältigen, oder die bereitgestellten IOPs erhöhen, um einen erhöhten Durchsatz in den zugrunde liegenden Speicher zu ermöglichen.</li><li>• Sie können vor Ihrer Datenbank auch eine Warteschlange einrichten, anstatt direkt in die Datenbank zu schreiben. Mithilfe dieses Musters können Sie die Datenerfassung von der Datenbank entkoppeln und die Flow-Rate steuern, sodass die Datenbank nicht überwältigt wird.</li><li>• Das Batching Ihrer Schreibanforderungen, anstatt mehrere kurzlebige Transaktionen zu erstellen, kann Ihnen dabei helfen, den Durchsatz bei relationalen Datenbanken mit hohem Schreibvolumen zu verbessern.</li><li>• Serverless-Datenbanken wie DynamoDB können den Schreibdurchsatz automatisch skalieren oder indem die bereitgestellten Kapazitätseinheiten für Schreibvorgänge (Write Capacity Units, WCU) abhängig vom Kapazitätsmodus angepasst werden.</li><li>• Es können immer noch Probleme mit heißen Partitionen auftreten, wenn Sie die Durchsatzgrenzen für einen bestimmten Partitionsschlüssel erreichen. Dies kann verhindert werden, indem Sie einen Partitionsschlüssel auswählen, der gleichmäßiger verteilt ist, oder indem Sie die Schreibvorgänge des Partitionsschlüssels in Shards aufteilen.</li></ul>

Konfigurationsoption	Beispiele
Richtlinien zum Verwalten des Lebenszyklus von Datensätzen	<ul style="list-style-type: none"> <li>• Mit <a href="#">Amazon-S3-Lebenszyklen</a> können Sie Ihre Objekte während ihres gesamten Lebenszyklus verwalten. Wenn die Zugriffsmuster unbekannt oder nicht prognostizierbar sind oder sich ändern, können Sie <a href="#">Amazon S3 Intelligent-Tiering</a> verwenden. Hiermit werden Zugriffsmuster überwacht und Objekte, auf die nicht zugegriffen wurde, automatisch in kostengünstigere Zugriffsebenen verschoben. Anhand der Metriken von <a href="#">Amazon S3 Storage Lens</a> können Sie Optimierungsmöglichkeiten und Lücken im Lebenszyklusmanagement ermitteln.</li> <li>• Das <a href="#">Amazon-EFS-Lebenszyklusmanagement</a> verwaltet den Dateispeicher für Ihre Dateisysteme automatisch.</li> </ul>
Verbindungsmanagement und Pooling	<ul style="list-style-type: none"> <li>• Amazon-RDS-Proxy kann mit Amazon RDS und Aurora verwendet werden, um Verbindungen mit der Datenbank zu verwalten.</li> <li>• Serverless-Datenbanken wie DynamoDB haben keine ihnen zugewiesenen Verbindungen, aber ziehen Sie die bereitgestellte Kapazität sowie automatische Skalierungsrichtlinien in Betracht, um Datenverkehrsspitzen zu bewältigen.</li> </ul>

- Führen Sie Experimente und Benchmarking in einer Nicht-Produktionsumgebung durch, um herauszufinden, welche Konfigurationsoption Ihre Workload-Anforderungen erfüllen kann.
- Nachdem Sie experimentiert haben, planen Sie die Migration und überprüfen Sie die Leistungsmetriken.

- Verwenden Sie AWS-Tools zur Überwachung (wie [Amazon CloudWatch](#)) und Optimierung (wie [Amazon S3 Storage Lens](#)), um den Datenspeicher kontinuierlich anhand realer Nutzungsmuster zu optimieren.

## Ressourcen

### Zugehörige Dokumente:

- [Cloud-Speicher mit AWS](#)
- [Typen von Amazon-EBS-Volumes](#)
- [Amazon-EC2-Speicher](#)
- [Amazon EFS: Amazon EFS-Performance](#)
- [Leistung von Amazon FSx für Lustre](#)
- [Leistung von Amazon FSx für Windows File Server](#)
- [Amazon Glacier: Glacier-Dokumentation](#)
- [Amazon S3: Überlegungen zu Anfragerate und Leistung](#)
- [Amazon EBS I/O Characteristics](#)
- [Cloud-Datenbanken mit AWS](#)
- [AWS-Datenbank-Caching](#)
- [DynamoDB Accelerator](#)
- [Bewährte Methoden mit Amazon Aurora](#)
- [Leistung von Amazon Redshift](#)
- [Die zehn besten Leistungstipps für Amazon Athena](#)
- [Bewährte Methoden für Amazon Redshift Spectrum](#)
- [Bewährte Methoden für Amazon DynamoDB](#)

### Zugehörige Videos:

- [AWS re:Invent 2023: Improve Amazon Elastic Block Store efficiency and be more cost-efficient](#)
- [AWS re:Invent 2023: Optimize storage price and performance with Amazon Simple Storage Service](#)
- [AWS re:Invent 2023: Building and optimizing a data lake on Amazon Simple Storage Service](#)

- [AWS re:Invent 2023: What's new with AWS file storage](#)
- [AWS re:Invent 2023: Dive deep into Amazon DynamoDB](#)

Zugehörige Beispiele:

- [AWS-Workshop „Speziell entwickelte Datenbanken“](#)
- [Databases for Developers](#)
- [AWS Modern Data Architecture Immersion Day](#)
- [Amazon EBS Autoscale](#)
- [Amazon-S3-Beispiele](#)
- [Amazon DynamoDB-Beispiele](#)
- [Beispiele von AWS-Datenbankmigration](#)
- [Workshop zur Modernisierung von Datenbanken](#)
- [Arbeiten mit Parametern auf Ihrem Amazon RDS für Postgress DB](#)

## PERF03-BP03 Erfassen und Aufzeichnen von Metriken zur Datenspeicherleistung

Verfolgen und zeichnen Sie relevante Leistungsmetriken für Ihren Datenspeicher auf, um zu verstehen, wie Ihre Datenverwaltungslösungen funktionieren. Mithilfe dieser Metriken können Sie Ihren Datenspeicher optimieren, überprüfen, ob Ihre Workload-Anforderungen erfüllt werden, und sich einen klaren Überblick über die Workload-Leistung verschaffen.

Typische Anti-Muster:

- Sie suchen ausschließlich manuell mithilfe von Protokolldateien nach Metriken.
- Sie veröffentlichen Metriken nur in internen Tools, die von Ihrem Team verwendet werden, und Sie haben kein umfassendes Bild Ihrer Workload.
- Sie verwenden nur die Standardmetriken, die von der Überwachungssoftware Ihrer Wahl aufgezeichnet wurden.
- Sie überprüfen Metriken nur dann, wenn ein Problem vorliegt.
- Sie überwachen Metriken nur auf Systemebene und erfassen keine Datenzugriffs- und Nutzungsmetriken.

Vorteile der Nutzung dieser bewährten Methode: Das Einrichten einer Leistungsbasislinie hilft Ihnen dabei, das normale Verhalten und die Anforderungen von Workloads zu verstehen. Abnorme Muster können schneller identifiziert und behoben werden, was die Leistung und Zuverlässigkeit des Datenspeichers erhöht.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

Um die Leistung der Datenspeicher zu überwachen, müssen Sie mehrere Leistungsmetriken über einen bestimmten Zeitraum aufzeichnen. Auf diese Weise können Sie Anomalien erkennen und die Leistung anhand von Geschäftsmetriken messen, um sicherzustellen, dass Sie die Anforderungen Ihrer Workload erfüllen.

Metriken sollten das zugrunde liegende System, das den Datenspeicher unterstützt, sowie die Datenbankmetriken enthalten. Die Metriken des zugrunde liegenden Systems können die CPU-Auslastung, den Arbeitsspeicher, den verfügbaren Festplattenspeicher, Festplatten-E/A, das Cache-Trefferverhältnis und Metriken zum eingehenden und ausgehenden Netzwerkdatenverkehr umfassen, während die Datenspeichermetriken die Transaktionen pro Sekunde, die häufigsten Abfragen, die durchschnittlichen Abfrageraten, Antwortzeiten, die Indexauslastung, Tabellenschlösser, Abfragezeitüberschreitungen und die Anzahl offener Verbindungen enthält. Diese Daten sind von entscheidender Bedeutung, um festzustellen, wie leistungsfähig die Workload ist und wie die Datenverwaltungslösung genutzt wird. Nutzen Sie diese Metriken im Rahmen eines datengestützten Ansatzes, der Ihnen die Feinabstimmung und Optimierung der von der Workload genutzten Ressourcen ermöglicht.

Nutzen Sie Tools, Bibliotheken und Systeme zum Aufzeichnen von Messungen zur Datenbankanleistung.

## Implementierungsschritte

- Identifizieren Sie die wichtigsten Leistungsmetriken, die der Datenspeicher verfolgen soll.
  - [Metriken und Dimensionen von Amazon S3](#)
  - [Überwachen von Metriken in einer Amazon-RDS-Instance](#)
  - [Überwachen der DB-Last mit Performance Insights in Amazon RDS](#)
  - [Überblick über „Erweiterte Überwachung“](#)
  - [DynamoDB-Metriken und -Dimensionen](#)
  - [Überwachen von DynamoDB Accelerator](#)

- [Überwachen von Amazon MemoryDB mit Amazon CloudWatch](#)
- [Welche Metriken sollte ich überwachen?](#)
- [Überwachen der Amazon-Redshift-Cluster-Leistung](#)
- [Timestream-Metriken und Dimensionen](#)
- [Amazon-CloudWatch-Metriken für Amazon Aurora](#)
- [Protokollierung und Überwachung in Amazon Keyspaces \(für Apache Cassandra\)](#)
- [Überwachen von Amazon-Neptune-Ressourcen](#)
- Verwenden Sie eine zugelassene Protokollierungs- und Überwachungslösung, um diese Metriken zu erfassen. Mit [Amazon CloudWatch](#) lassen sich Metriken aus sämtlichen Ressourcen Ihrer Architektur erfassen. Sie können auch benutzerdefinierte Metriken erfassen und in Oberflächen-, Geschäfts- oder abgeleiteten Metriken veröffentlichen. Richten Sie mit CloudWatch oder mit Lösungen von Drittanbietern Alarme ein, die auf das Überschreiten von Schwellenwerten hinweisen.
- Prüfen Sie, ob die Datenspeicherüberwachung von einer Machine-Learning-Lösung profitieren kann, die Leistungsanomalien erkennt.
  - [Amazon DevOps Guru für Amazon RDS](#) ermöglicht einen Einblick in Leistungsprobleme und bietet Empfehlungen für Korrekturmaßnahmen.
- Konfigurieren Sie die Datenaufbewahrung in Ihrer Überwachungs- und Protokollierungslösung so, dass sie Ihren Sicherheits- und Betriebszielen entspricht.
  - [Standard-Datenaufbewahrung für CloudWatch-Metriken](#)
  - [Standard-Datenaufbewahrung für CloudWatch-Protokolle](#)

## Ressourcen

### Zugehörige Dokumente:

- [AWS-Datenbank-Caching](#)
- [Die zehn besten Leistungstipps für Amazon Athena](#)
- [Bewährte Methoden mit Amazon Aurora](#)
- [DynamoDB Accelerator](#)
- [Bewährte Methoden für Amazon DynamoDB](#)
- [Bewährte Methoden für Amazon Redshift Spectrum](#)
- [Leistung von Amazon Redshift](#)

- [Cloud-Datenbanken mit AWS](#)
- [Erkenntnisse zur Amazon-RDS-Leistung](#)

Zugehörige Videos:

- [AWS re:Invent 2022 – Performance monitoring with Amazon RDS and Aurora, featuring Autodesk](#)
- [Database Performance Monitoring and Tuning with Amazon DevOps Guru for Amazon RDS](#)
- [AWS re:Invent 2023 – What's new with AWS file storage](#)
- [AWS re:Invent 2023 – Dive deep into Amazon DynamoDB](#)
- [AWS re:Invent 2023 – Building and optimizing a data lake on Amazon S3](#)
- [AWS re:Invent 2023 – What's new with AWS file storage](#)
- [AWS re:Invent 2023 – Dive deep into Amazon DynamoDB](#)
- [Best Practices for Monitoring Redis Workloads on Amazon ElastiCache](#)

Zugehörige Beispiele:

- [Framework zur AWS-Datensatzerfassung und Sammlung von Metriken](#)
- [Workshop zur Amazon-RDS-Überwachung](#)
- [AWS-Workshop „Speziell entwickelte Datenbanken“](#)

## PERF03-BP04 Implementieren von Strategien zur Verbesserung der Abfrageleistung im Datenspeicher

Implementieren Sie Strategien zur Datenoptimierung und Verbesserung der Datenabfrage, um mehr Skalierbarkeit und eine effizientere Leistung für Ihre Workloads zu erzielen.

Typische Anti-Muster:

- Sie partitionieren keine Daten in Ihrem Datenspeicher.
- Sie speichern Daten in nur einem Dateiformat in Ihrem Datenspeicher.
- Sie verwenden keine Indizes in Ihrem Datenspeicher.

Vorteile der Nutzung dieser bewährten Methode: Die Optimierung der Daten- und Abfrageleistung führt zu mehr Effizienz, niedrigeren Kosten und einer verbesserten Benutzererfahrung.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Daten- und Abfrageoptimierung sind wichtige Aspekte der Leistungseffizienz in einem Datenspeicher, da sie sich auf die Leistung und Reaktionsfähigkeit der gesamten Cloud-Workload auswirken. Nicht optimierte Abfragen können zu einem höheren Ressourcenverbrauch und Engpässen führen, wodurch die Gesamteffizienz eines Datenspeichers beeinträchtigt wird.

Die Datenoptimierung umfasst mehrere Techniken, um eine effiziente Datenspeicherung und einen effizienten Datenzugriff zu gewährleisten. Dies trägt auch dazu bei, die Abfrageleistung in einem Datenspeicher zu verbessern. Zu den wichtigsten Strategien gehören Datenpartitionierung, Datenkomprimierung und Datendenormalisierung, mit denen Daten sowohl für die Speicherung als auch für den Zugriff optimiert werden können.

### Implementierungsschritte

- Verstehen und analysieren Sie die kritischen Datenabfragen, die in Ihrem Datenspeicher durchgeführt werden.
- Identifizieren Sie die langsamen Abfragen in Ihrem Datenspeicher und verwenden Sie Abfragepläne, um den aktuellen Status zu verstehen.
  - [Analysieren des Abfrageplans in Amazon Redshift](#)
  - [Verwenden von EXPLAIN und EXPLAIN ANALYZE in Athena](#)
- Implementieren Sie Strategien zur Verbesserung der Abfrageleistung. Einige der wichtigsten Strategien sind:
  - Nutzung eines [spaltenbasierten Dateiformats](#) (wie Parquet oder ORC).
  - Komprimieren von Daten im Datenspeicher, um Speicherplatz und E/A-Betrieb zu reduzieren.
  - Datenpartitionierung zur Aufteilung von Daten in kleinere Teile und zur Reduzierung der Zeit für das Scannen von Daten.
    - [Daten in Athena partitionieren](#)
    - [Partitionen und Datenverteilung](#)
  - Datenindizierung für die gemeinsamen Spalten in der Abfrage.
  - Verwenden Sie materialisierte Ansichten für häufige Abfragen.
    - [Materialisierte Ansichten verstehen](#)
    - [Erstellen von materialisierten Ansichten in Amazon Redshift](#)

- Wählen Sie den richtigen Verknüpfungsvorgang für die Abfrage aus. Wenn Sie zwei Tabellen verknüpfen, geben Sie die größere Tabelle auf der linken Seite der Verknüpfung und die kleinere Tabelle auf der rechten Seite der Verknüpfung an.
- Verteilte Caching-Lösung zur Verbesserung der Latenz und zur Reduzierung der Anzahl von Datenbank-E/A-Vorgängen.
- Regelmäßige Wartung wie [Bereinigung](#), Neuindizierung und [Ausführen von Statistiken](#).
- Experimentieren und testen Sie Strategien in einer Nicht-Produktionsumgebung.

## Ressourcen

### Zugehörige Dokumente:

- [Bewährte Methoden mit Amazon Aurora](#)
- [Leistung von Amazon Redshift](#)
- [Die zehn besten Leistungstipps für Amazon Athena](#)
- [AWS-Datenbank-Caching](#)
- [Bewährte Methoden für die Implementierung von Amazon ElastiCache](#)
- [Daten in Athena partitionieren](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Bewährte Methoden zur Kostenoptimierung für AWS-Speicher](#)
- [AWS re:Invent 2022 – Performance monitoring with Amazon RDS and Aurora, featuring Autodesk](#)
- [Optimize Amazon Athena Queries with New Query Analysis Tools](#)

### Zugehörige Beispiele:

- [AWS Purpose Built Databases Workshop](#)

## PERF03-BP05 Implementieren von Datenzugriffsmustern, die Caching nutzen

Implementieren Sie Zugriffsmuster, die vom Daten-Caching profitieren, damit häufig aufgerufene Daten schnell abgerufen werden können.

## Typische Anti-Muster:

- Sie speichern Daten, die sich häufig ändern.
- Sie verlassen sich auf zwischengespeicherte Daten, als ob sie dauerhaft gespeichert und immer verfügbar wären.
- Sie berücksichtigen nicht die Konsistenz Ihrer zwischengespeicherten Daten.
- Sie überwachen die Effizienz Ihrer Caching-Implementierung nicht.

Vorteile der Nutzung dieser bewährten Methode: Das Speichern von Daten in einem Cache kann die Leselatenz, den Lesedurchsatz, die Benutzererfahrung und die Gesamteffizienz verbessern sowie die Kosten senken.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Ein Cache ist eine Software- oder Hardwarekomponente zum Speichern von Daten, damit zukünftige Abfragen derselben Daten schneller oder effizienter verarbeitet werden können. Die in einem Cache gespeicherten Daten können bei Verlust rekonstruiert werden, indem eine frühere Berechnung wiederholt wird oder die Daten aus einem anderen Datenspeicher abgerufen werden.

Das Caching von Daten kann eine der effektivsten Strategien sein, um die allgemeine Anwendungsleistung zu verbessern und die Belastung Ihrer zugrunde liegenden primären Datenquellen zu verringern. Daten können auf mehreren Ebenen in der Anwendung zwischengespeichert werden, z. B. innerhalb der Anwendung durch Remote-Aufrufe (so genanntes clientseitiges Caching) oder durch Verwendung eines schnellen sekundären Services zur Speicherung der Daten (so genanntes Remote-Caching).

### Clientseitiges Caching

Beim clientseitigen Caching kann jeder Client (eine Anwendung oder ein Service, die bzw. der den Backend-Datenspeicher abfragt) die Ergebnisse seiner eindeutigen Abfragen lokal für einen bestimmten Zeitraum speichern. So kann die Anzahl der Anfragen an einen Datenspeicher im Netzwerk reduziert werden, da zuerst der lokale Client-Cache überprüft wird. Wenn die Ergebnisse nicht vorhanden sind, kann die Anwendung den Datenspeicher abfragen und diese Ergebnisse lokal speichern. Dieses Muster ermöglicht es jedem Client, Daten am nächstgelegenen Ort (dem Client selbst) zu speichern, was zur geringstmöglichen Latenz führt. Clients können auch weiterhin einige

Abfragen bearbeiten, wenn der Backend-Datenspeicher nicht verfügbar ist, wodurch die Verfügbarkeit des Gesamtsystems erhöht wird.

Ein Nachteil dieses Ansatzes besteht darin, dass bei Beteiligung mehrerer Clients diese möglicherweise dieselben zwischengespeicherten Daten lokal speichern. Dies führt sowohl zu doppelten Speichervorgängen als auch zu Dateninkonsistenzen zwischen diesen Clients. So kann z. B. ein Client die Ergebnisse einer Abfrage zwischenspeichern und eine Minute später führt ein anderer Client dieselbe Abfrage aus und erhält ein anderes Ergebnis.

## Remote-Caching

Um das Problem der doppelten Daten zwischen den Clients zu lösen, kann ein schneller externer Service oder ein Remote-Cache verwendet werden, um die abgefragten Daten zu speichern. Anstatt einen lokalen Datenspeicher zu überprüfen, prüft jeder Client den Remote-Cache, bevor er den Backend-Datenspeicher abfragt. Diese Strategie ermöglicht konsistentere Antworten zwischen den Clients, eine bessere Effizienz der gespeicherten Daten und ein höheres Volumen an zwischengespeicherten Daten, da der Speicherplatz unabhängig von den Clients skaliert wird.

Der Nachteil eines Remote-Caches besteht darin, dass das Gesamtsystem möglicherweise eine höhere Latenz aufweist, da ein zusätzlicher Netzwerk-Hop erforderlich ist, um den Remote-Cache zu überprüfen. Das clientseitige Caching kann in Kombination mit dem Remote-Caching verwendet werden, um ein mehrstufiges Caching zu implementieren und die Latenz zu verbessern.

## Implementierungsschritte

- Identifizieren Sie Datenbanken, APIs und Netzwerkservices, die vom Caching profitieren könnten. Services, die hohe Lese-Workloads oder ein hohes Lese-Schreib-Verhältnis aufweisen oder deren Skalierung teuer ist, kommen für das Caching in Frage.
  - [Datenbank-Caching](#)
  - [Aktivieren von API-Caching für verbesserte Reaktionsfähigkeit](#)
- Identifizieren Sie die geeignete Caching-Strategie, die am besten zu Ihrem Zugriffsmuster passt.
  - [Caching-Strategien](#)
  - [AWS-Caching-Lösungen](#)
- Befolgen Sie die [bewährten Methoden für das Caching](#) für Ihren Datenspeicher.
- Konfigurieren Sie eine Cache-Invalidierungsstrategie, z. B. eine Time-to-Live (TTL), für alle Daten, die ein Gleichgewicht zwischen der Aktualität der Daten und der Verringerung der Auslastung des Backend-Datenspeichers herstellt.

- Aktivieren Sie Features wie automatische Verbindungswiederholungen, exponentielles Backoff, clientseitige Timeouts und Verbindungspooling beim Client, sofern verfügbar, um die Leistung und Zuverlässigkeit zu verbessern.
  - [Bewährte Methoden: Redis-Clients und Amazon ElastiCache \(Redis OSS\)](#)
- Überwachen Sie die Cache-Trefferrate mit einem Ziel von mindestens 80 %. Niedrigere Werte können auf eine unzureichende Cache-Größe oder ein Zugriffsmuster hinweisen, das nicht vom Caching profitiert.
  - [Welche Metriken sollte ich überwachen?](#)
  - [Best Practices for Monitoring Redis Workloads on Amazon ElastiCache](#)
  - [Monitoring best practices with Amazon ElastiCache \(Redis OSS\) using Amazon CloudWatch](#)
- Implementieren Sie die [Datenreplikation](#), um Lesevorgänge auf mehrere Instances auszulagern und die Leseleistung und Verfügbarkeit von Daten zu verbessern.

## Ressourcen

### Zugehörige Dokumente:

- [Verwendung von Amazon ElastiCache Well-Architected Lens](#)
- [Monitoring best practices with Amazon ElastiCache \(Redis OSS\) using Amazon CloudWatch](#)
- [Welche Metriken sollte ich überwachen?](#)
- [Whitepaper „Skalierbare Leistung mit Amazon ElastiCache“](#)
- [Caching-Herausforderungen und -Strategien](#)

### Zugehörige Videos:

- [Amazon ElastiCache Learning Path](#)
- [Bewährte Methoden für erfolgreiches Design in Amazon ElastiCache](#)
- [AWS re:Invent 2020 – Bewährte Methoden für erfolgreiches Design in Amazon ElastiCache](#)
- [AWS re:Invent 2023 – \[LAUNCH\] Introducing Amazon ElastiCache Serverless](#)
- [AWS re:Invent 2022 – 5 hervorragende Methoden, um die Datenebene mit Redis neu zu gestalten](#)
- [AWS re:Invent 2021 – Deep dive on Amazon ElastiCache \(Redis OSS\)](#)

### Zugehörige Beispiele:

- [Steigerung der MySQL-Datenbankleistung mit Amazon ElastiCache \(Redis OSS\)](#)

# Netzwerk und Bereitstellung von Inhalten

Welche Netzwerklösung für eine Workload optimal ist, richtet sich nach der Latenz, dem erforderlichen Durchsatz, dem Jitter und der Bandbreite. Die Standortoptionen sind von den physischen Einschränkungen abhängig, z. B. von Benutzer- oder On-Premises-Ressourcen. Diese Einschränkungen können durch Edge-Standorte oder die Ressourcenplatzierung wettgemacht werden.

In AWS wird das Netzwerk virtualisiert und es sind unterschiedliche Typen und Konfigurationen verfügbar. Das erleichtert Ihnen die Anpassung Ihrer Networking-Anforderungen. AWS bietet zur Optimierung des Netzwerkdatenverkehrs Produktfeatures wie Enhanced Networking, für Amazon-EC2-Netzwerke optimierte Instances, Amazon-S3-Übertragungsbeschleunigung und Amazon CloudFront (dynamisch) zur Optimierung des Netzwerk-Datenverkehrs. Zur Verbesserung der Latenz und der Stabilität des Netzwerks finden Sie in AWS zudem Netzwerkfeatures wie die latenzbasierte Weiterleitung mit Amazon Route 53, Amazon-VPC-Endpunkte, AWS Direct Connect und AWS Global Accelerator.

In diesem Schwerpunktbereich werden Anleitungen und bewährte Methoden für die Entwicklung, Konfiguration und den Betrieb effizienter Netzwerk und Inhaltsbereitstellungslösungen in der Cloud bereitgestellt.

## Bewährte Methoden

- [PERF04-BP01 Verstehen der Auswirkungen des Netzwerks auf die Leistung](#)
- [PERF04-BP02 Evaluieren verfügbarer Netzwerk-Features](#)
- [PERF04-BP03 Auswählen von entsprechend dedizierter Konnektivität oder VPN für Ihre Workload](#)
- [PERF04-BP04 Lastausgleich verwenden, um den Datenverkehr auf mehrere Ressourcen zu verteilen](#)
- [PERF04-BP05 Auswählen leistungsfördernder Netzwerkprotokolle](#)
- [PERF04-BP06 Wählen Sie den Standort Ihres Workloads basierend auf den Netzwerkanforderungen](#)
- [PERF04-BP07 Optimieren der Netzwerkkonfiguration basierend auf Metriken](#)

# PERF04-BP01 Verstehen der Auswirkungen des Netzwerks auf die Leistung

Analysieren und verstehen Sie, wie sich netzwerkbezogene Entscheidungen auf Ihre Workload auswirken, sodass Sie eine effiziente Leistung und ein verbessertes Benutzererlebnis erzielen können.

Typische Anti-Muster:

- Der gesamte Datenverkehr fließt durch Ihre bestehenden Rechenzentren.
- Sie leiten den gesamten Datenverkehr durch zentrale Firewalls, anstatt cloudnative Netzwerksicherheitstools zu verwenden.
- Sie stellen AWS Direct Connect-Verbindungen bereit, ohne die tatsächlichen Nutzungsanforderungen zu verstehen.
- Sie berücksichtigen beim Definieren Ihrer Netzwerklösungen die Workload-Eigenschaften und den Verschlüsselungsaufwand nicht.
- Sie verwenden On-Premises-Konzepte und -Strategien für Netzwerklösungen in der Cloud.

Vorteile der Nutzung dieser bewährten Methode: Indem Sie verstehen, wie das Netzwerk die Workload-Leistung beeinflusst, können Sie potenzielle Engpässe erkennen, die Benutzererfahrung verbessern, die Zuverlässigkeit erhöhen und den Betriebsaufwand verringern, während sich die Workload verändert.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

Das Netzwerk ist für die Verbindung zwischen Anwendungskomponenten, Cloud-Services, Edge-Netzwerken und On-Premises-Daten verantwortlich und kann daher die Workload-Leistung wesentlich beeinflussen. Die Benutzererfahrung kann nicht nur durch die Workload-Leistung, sondern auch durch Netzwerklatenz, Bandbreite, Protokolle, Standort, Netzwerküberlastungen, Jitter, Durchsatz und Routingregeln beeinträchtigt werden.

Sie haben eine dokumentierte Liste an Netzwerkanforderungen der Workload, einschließlich Latenz, Paketgröße, Routingregeln, Protokolle und unterstützender Datenverkehrsmuster. Sie überprüfen alle verfügbaren Netzwerklösungen und identifizieren, welcher Service den Netzwerkmerkmalen Ihrer

Workload entspricht. Da cloudbasierte Netzwerke schnell geändert werden können, müssen Sie Ihre Netzwerkarchitektur im Laufe der Zeit weiterentwickeln, um die effiziente Leistung zu verbessern.

### Implementierungsschritte:

- Definieren und dokumentieren Sie die Anforderungen an die Netzwerkleistung, einschließlich Metriken wie Netzwerklatenz, Bandbreite, Protokolle, Standorte, Datenverkehrsmuster (Spitzen und Frequenz), Durchsatz, Verschlüsselung, Überprüfung und Routingregeln.
- Erfahren Sie mehr über wichtige AWS-Netzwerk-Services wie [VPCs](#), [AWS Direct Connect](#), [Elastic Load Balancing \(ELB\)](#) und [Amazon Route 53](#).
- Erfassen Sie die folgenden wichtigen Netzwerkmerkmale:

Merkmale	Tools und Metriken
Grundlegende Netzwerkmerkmale	<ul style="list-style-type: none"> <li>• <a href="#">VPC Flow-Protokolle</a></li> <li>• <a href="#">AWS Transit Gateway-Flow-Protokolle</a></li> <li>• <a href="#">AWS Transit Gateway Metriken</a></li> <li>• <a href="#">AWS PrivateLink Metriken</a></li> </ul>
Merkmale von Anwendungsnetzwerken	<ul style="list-style-type: none"> <li>• <a href="#">Elastic Fabric Adapter</a></li> <li>• <a href="#">AWS App Mesh Metriken</a></li> <li>• <a href="#">Metriken für Amazon API Gateway</a></li> </ul>
Merkmale von Edge-Netzwerken	<ul style="list-style-type: none"> <li>• <a href="#">Amazon-CloudFront-Metriken</a></li> <li>• <a href="#">Amazon-Route-53-Metriken</a></li> <li>• <a href="#">AWS Global Accelerator Metriken</a></li> </ul>
Merkmale hybrider Netzwerke	<ul style="list-style-type: none"> <li>• <a href="#">Direct Connect Metriken</a></li> <li>• <a href="#">AWS Site-to-Site VPN Metriken</a></li> <li>• <a href="#">AWS Client VPN Metriken</a></li> <li>• <a href="#">AWS Cloud-WAN-Metriken</a></li> </ul>
Merkmale von Sicherheitsnetzwerken	<ul style="list-style-type: none"> <li>• <a href="#">Metriken von AWS Shield, AWS WAF und AWS Network Firewall</a></li> </ul>
Nachverfolgungsmerkmale	<ul style="list-style-type: none"> <li>• <a href="#">AWS X-Ray</a></li> </ul>

Merkmale	Tools und Metriken
	<ul style="list-style-type: none"><li>• <a href="#">VPC Reachability Analyzer</a></li><li>• <a href="#">Network Access Analyzer</a></li><li>• <a href="#">Amazon Inspector</a></li><li>• <a href="#">Amazon CloudWatch RUM</a></li></ul>

- Benchmarks für die Netzwerkleistung festlegen und testen:
  - Führen Sie [Benchmark](#)-Tests für den Netzwerkdurchsatz durch, da einige Faktoren die Amazon-EC2-Netzwerkleistung beeinflussen können, wenn sich Instances in derselben VPC befinden. Messen Sie die Netzwerkbandbreite zwischen Linux-Instances von Amazon EC2 in der gleichen VPC.
  - Führen Sie [Lasttests](#) durch, um mit Netzwerklösungen und -optionen zu experimentieren.

## Ressourcen

### Zugehörige Dokumente:

- [Application Load Balancer](#)
- [EC2: Enhanced Networking unter Linux](#)
- [EC2: Enhanced Networking unter Windows](#)
- [EC2: Platzierungsgruppen](#)
- [Enabling Enhanced Networking with the Elastic Network Adapter \(ENA\) on Linux Instances](#)
- [Network Load Balancer](#)
- [Netzwerkprodukte mit AWS](#)
- [Transit-Gateway](#)
- [Umstellung auf latenzbasiertes Routing in Amazon Route 53](#)
- [VPC-Endpunkte](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – AWS-Netzwerkgrundlagen](#)
- [AWS re:Invent 2023 – Welche Vorteile bietet die Vernetzung für Ihre Anwendung?](#)
- [AWS re:Invent 2023 – Erweiterte VPC-Designs und neue Funktionen](#)

- [AWS re:Invent 2023 – Leitfaden für Entwickler von Cloud-Netzwerken](#)
- [AWS re:Invent 2019 – Konnektivität mit AWS und hybriden AWS-Netzwerkarchitekturen](#)
- [AWS re:Invent 2019 – Optimizing Network Performance for Amazon EC2 Instances](#)
- [AWS Summit Online – Verbessern der Leistung von globalen Netzwerken für Anwendungen](#)
- [AWS re:Invent 2020 – Bewährte Methoden für Netzwerke und Tipps für das Well-Architected Framework](#)
- [AWS re:Invent 2020 – Bewährte Methoden für AWS-Netzwerke in umfangreichen Migrationen](#)

Zugehörige Beispiele:

- [AWS Transit Gateway und skalierbare Sicherheitslösungen](#)
- [Workshops zu AWS-Netzwerken](#)
- [Praktischer Workshop zur Netzwerk-Firewall](#)
- [Beobachten und Diagnostizieren Ihres Netzwerks in AWS](#)
- [Finden und Beheben von Netzwerkfehlerkonfigurationen in AWS](#)

## PERF04-BP02 Evaluieren verfügbarer Netzwerk-Features

Prüfen Sie die Netzwerk-Features in der Cloud, mit denen die Leistung unter Umständen verbessert werden kann. Messen Sie die Auswirkungen der Features anhand von Tests, Metriken und Analysen. Nutzen Sie beispielsweise die verfügbaren Features auf Netzwerkebene, um die Latenz, die Netzwerkentfernung oder den Jitter zu reduzieren.

Typische Anti-Muster:

- Sie bleiben innerhalb einer Region, da sich Ihre Firmenzentrale dort befindet.
- Sie verwenden Firewalls anstelle von Sicherheitsgruppen, um den Datenverkehr zu filtern.
- Sie unterbrechen TLS für die Überprüfung des Datenverkehrs, anstatt sich auf Sicherheitsgruppen, Endpunktrichtlinien und andere cloudnative Funktionen zu verlassen.
- Sie nutzen nur eine subnetzbasierte Segmentierung anstelle von Sicherheitsgruppen.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie alle Service-Features und Optionen evaluieren, kann dies die Workload-Leistung verbessern, die Infrastrukturkosten senken, den

Verwaltungsaufwand für die Workload reduzieren und die allgemeine Sicherheit erhöhen. Dank der weltweiten Abdeckung von AWS können Sie Ihren Kunden stets das bestmögliche Netzwerkerlebnis bieten.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

## Implementierungsleitfaden

AWS bietet Services wie [AWS Global Accelerator](#) und [Amazon CloudFront](#), die zur Verbesserung der Netzwerkleistung beitragen können, während die meisten AWS-Services über Produktfeatures (wie das Feature [Amazon S3 Transfer Acceleration](#)) zur Optimierung des Netzwerkverkehrs verfügen.

Sehen Sie sich die verfügbaren Konfigurationsoptionen für das Netzwerk an und finden Sie heraus, wie sich diese auf Ihre Workload auswirken. Die Leistungsoptimierung hängt davon ab, wie diese Optionen mit Ihrer Architektur interagieren und welche Auswirkungen sie auf die gemessene Leistung und auf die Benutzererfahrung haben.

## Implementierungsschritte

- Erstellen Sie eine Liste der Workload-Komponenten.
- Erwägen Sie die Verwendung von [AWS Cloud WAN](#), um das Netzwerk Ihrer Organisation aufzubauen, zu verwalten und zu überwachen, wenn Sie ein einheitliches globales Netzwerk aufbauen.
- Überwachen Sie Ihre globalen und Kernnetzwerke mit [Metriken für Amazon CloudWatch Logs](#). Nutzen Sie [Amazon CloudWatch RUM](#), das Erkenntnisse bietet, die dazu beitragen, das digitale Erlebnis der Benutzer zu identifizieren, zu verstehen und zu verbessern.
- Zeigen Sie die aggregierte Netzwerklatenz zwischen AWS-Regionen und den Availability Zones sowie innerhalb der einzelnen Availability Zones mit [AWS Network Manager](#) an, um einen Einblick in das Verhältnis zwischen der Leistung Ihrer Anwendung und der Leistung des zugrunde liegenden AWS-Netzwerks zu erhalten.
- Verwenden Sie ein vorhandenes CMDB-Tool (Configuration Management Database) oder einen Service wie [AWS Config](#), um eine Bestandsaufnahme Ihrer Workload und seiner Konfiguration zu erstellen.
- Wenn es sich um eine bestehende Workload handelt, ermitteln und dokumentieren Sie die Benchmark für Ihre Leistungsmetriken. Konzentrieren Sie sich dabei auf Engpässe und Bereiche mit Verbesserungspotenzial. Leistungsbezogene Netzwerkmetriken werden je nach geschäftlichen Anforderungen und Workload-Merkmalen für die einzelnen Workloads unterschiedlich sein. Für den

Anfang könnte die Prüfung folgender Metriken für Ihre Workload wichtig sein: Bandbreite, Latenz, Paketverlust, Jitter und erneute Übertragungen.

- Wenn es sich um eine neue Workload handelt, führen Sie [Lasttests](#) durch, um Leistungsengpässe zu ermitteln.
- Prüfen Sie für die ermittelten Leistungsengpässe die Konfigurationsoptionen Ihrer Lösungen, um Möglichkeiten zur Leistungsverbesserung zu finden. Informieren Sie sich über die folgenden wichtigen Netzwerkoptionen und -Features:

Verbesserungsmöglichkeit	Lösung
Netzwerkpfad oder -routen	Verwenden Sie <a href="#">Network Access Analyzer</a> , um Pfade oder Routen zu ermitteln.
Netzwerkprotokolle	Siehe <a href="#">PERF04-BP05 Auswählen leistungsfördernder Netzwerkprotokolle</a>
Netzwerktopologie	<p>Bewerten Sie Ihre betrieblichen und leistungsbezogenen Kompromisse zwischen <a href="#">VPC Peering</a> und <a href="#">AWS Transit Gateway</a> beim Verbinden mehrerer Konten. AWS Transit Gateway vereinfacht die Art und Weise, wie Sie all Ihre VPCs miteinander verbinden, die sich über Tausende von AWS-Konten-Netzwerken bis hin zu On-Premises-Netzwerken erstrecken können. Teilen Sie Ihr AWS Transit Gateway zwischen mehreren Konten mit <a href="#">AWS Resource Access Manager</a>.</p> <p>Siehe <a href="#">PERF04-BP03 Auswählen von entsprechend dedizierter Konnektivität oder VPN für Ihre Workload</a></p>
Netzwerkservices	<a href="#">AWS Global Accelerator</a> ist ein Netzwerkservice, der die Leistung des Benutzerdatenverkehrs unter Verwendung der globalen Netzwerkinfrastruktur von AWS um bis zu 60 % verbessert.

Verbesserungsmöglichkeit	Lösung
	<p><a href="#">Amazon CloudFront</a> kann die Leistung Ihrer Workload-Inhaltsbereitstellung und Latenz weltweit verbessern.</p> <p>Verwenden Sie <a href="#">Lambda@edge</a>, um Funktionen auszuführen, die den Inhalt, den CloudFront bereitstellt, besser an die Benutzer anpassen, die Latenz reduzieren und die Leistung erhöhen.</p> <p>Amazon Route 53 bietet Optionen für <a href="#">latenzbasiertes Routing</a>, <a href="#">Geolocation-Routing</a>, <a href="#">Geoproximity-Routing</a> und <a href="#">IP-basiertes Routing</a>, mit denen Sie die Leistung Ihrer Workloads für eine globale Zielgruppe verbessern können. Ermitteln Sie, welche Routing-Option Ihre Workload-Leistung optimieren würde. Prüfen Sie dazu Ihren Workload-Datenverkehr und den Standort bei der globalen Verteilung Ihrer Workload.</p>

Verbesserungsmöglichkeit	Lösung
Speicherressourcen-Features	<p>Durch das Feature <a href="#">Amazon S3 Transfer Acceleration</a> profitieren externe Benutzer beim Hochladen von Daten in Amazon S3 von den Netzwerkoptimierungen von CloudFront. Dies erleichtert die Übertragung großer Datenmengen von Remotestandorten ohne spezielle Konnektivität zur AWS Cloud.</p> <p><a href="#">Multi-Region-Zugangspunkte in Amazon S3</a> replizieren Inhalte in mehreren Regionen und vereinfachen die Workload durch die Bereitstellung eines Zugangspunkts. Bei Verwendung eines Multi-Region-Zugangspunkts können Sie Daten anfordern oder in Amazon S3 schreiben, wobei der Service den Bucket mit der geringsten Latenz ermittelt.</p>

Verbesserungsmöglichkeit	Lösung
Compute-Ressourcen-Features	<p>Die von Amazon-EC2-Instances, Container n und Lambda-Funktionen verwendeten <a href="#">Elastic-Network-Schnittstellen (ENI)</a> sind pro Fluss begrenzt. Überprüfen Sie Ihre Platzierungsgruppen, um Ihren <a href="#">EC2-Netzwerkdurchsatz</a> zu optimieren. Um Engpässe auf Pro-Fluss-Basis zu vermeiden, sollten Sie Ihre Anwendung so gestalten, dass mehrere Flüsse verwendet werden. Um Ihre datenverarbeitungsbezogenen Netzwerkmetriken zu überwachen und Einblicke in diese Metriken zu erhalten, verwenden Sie CloudWatch-Metriken und <a href="#">ethtool</a>. Der <code>ethtool</code>-Befehl ist im ENA-Treiber enthalten und stellt zusätzliche netzwerkbezogene Metriken zur Verfügung, die als <a href="#">benutzerdefinierte Metrik</a> in CloudWatch veröffentlicht werden können.</p> <p><a href="#">Amazon Elastic Network Adapters (ENA)</a> ermöglichen eine weitere Optimierung, indem sie für Ihre Instances innerhalb einer <a href="#">Cluster-Placement-Gruppe</a> einen besseren Durchsatz bereitstellen.</p> <p><a href="#">Elastic Fabric Adapter (EFA)</a> ist eine Netzwerkschnittstelle für Amazon-EC2-Instances, mit der Sie Workloads, für die ein hohes Maß an Kommunikation zwischen Knoten anfällt, unter AWS bedarfsgesteuert ausführen können.</p> <p>Für <a href="#">Amazon EBS optimierte Instances</a> verwenden einen optimierten Konfigurations-Stack und bieten zusätzliche, dediziert</p>

Verbesserungsmöglichkeit	Lösung
	e Kapazität, um die Amazon EBS-I/O zu erhöhen.

## Ressourcen

### Zugehörige Dokumente:

- [Application Load Balancer](#)
- [EC2: Enhanced Networking unter Linux](#)
- [EC2: Enhanced Networking unter Windows](#)
- [EC2: Platzierungsgruppen](#)
- [Enabling Enhanced Networking with the Elastic Network Adapter \(ENA\) on Linux Instances](#)
- [Network Load Balancer](#)
- [Netzwerkprodukte mit AWS](#)
- [Umstellung auf latenzbasiertes Routing in Amazon Route 53](#)
- [VPC-Endpunkte](#)
- [VPC Flow Logs](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Sind Sie bereit für Neues? Gestaltung von Netzwerken für Wachstum und Flexibilität](#)
- [AWS re:Invent 2023 – Erweiterte VPC-Designs und neue Funktionen](#)
- [AWS re:Invent 2023 – Leitfaden für Entwickler von Cloud-Netzwerken](#)
- [AWS re:Invent 2022 – Ausführliche Beschreibung der AWS-Netzwerkinfrastruktur](#)
- [AWS re:Invent 2019 – Konnektivität mit AWS und hybriden AWS-Netzwerkarchitekturen](#)
- [AWS re:Invent 2018 – Optimizing Network Performance for Amazon EC2 Instances](#)
- [AWS Global Accelerator](#)

### Zugehörige Beispiele:

- [AWS Transit Gateway und skalierbare Sicherheitslösungen](#)

- [Workshops zu AWS-Netzwerken](#)
- [Überwachung und Diagnose Ihres Netzwerks](#)
- [Finden und Beheben von Netzwerkfehlfunktionen in AWS](#)

## PERF04-BP03 Auswählen von entsprechend dedizierter Konnektivität oder VPN für Ihre Workload

Wenn Hybrid-Konnektivität für die Verbindung von On-Premises- und Cloud-Ressourcen erforderlich ist, stellen Sie ausreichend Bandbreite bereit, um Ihre Leistungsanforderungen zu erfüllen. Schätzen Sie die Anforderungen an Bandbreite und Latenz für Ihre hybride Workload ab. Diese Zahlen dienen als Grundlage für die Größenanpassung.

Typische Anti-Muster:

- Sie evaluieren nur VPN-Lösungen für Ihre Netzwerk-Verschlüsselungsanforderungen.
- Sie bewerten keine Optionen für Sicherung oder redundante Verbindungen.
- Sie identifizieren nicht alle Workload-Anforderungen (Verschlüsselung, Protokoll, Bandbreite und Traffic-Bedarf).

Vorteile der Nutzung dieser bewährten Methode: Durch die Auswahl und Konfiguration geeigneter Konnektivitätslösungen wird die Zuverlässigkeit Ihrer Workloads erhöht und die Leistung maximiert. Indem Sie die Workload-Anforderungen identifizieren, im Voraus planen und hybride Lösungen evaluieren, verringern Sie teure physische Netzwerkänderungen sowie den Betriebsaufwand und verkürzen die Amortisationszeit.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

### Implementierungsleitfaden

Entwickeln Sie eine hybride Netzwerkarchitektur, die auf Ihren Bandbreitenanforderungen basiert. [Direct Connect](#) ermöglicht es Ihnen, Ihr On-Premises-Netzwerk privat mit AWS zu verbinden. Sie ist geeignet, wenn Sie eine hohe Bandbreite und eine geringe Latenz bei gleichbleibender Leistung benötigen. Eine VPN-Verbindung stellt eine sichere Verbindung über das Internet her. Sie wird verwendet, wenn lediglich eine temporäre Verbindung erforderlich ist, wenn die Kosten eine Rolle spielen, oder wenn bei der Verwendung von Direct Connect darauf gewartet wird, dass eine resiliente physische Netzwerkkonnektivität hergestellt wird.

Wenn Ihre Bandbreitenanforderungen hoch sind, könnten Sie mehrere Direct Connect oder VPN-Services in Betracht ziehen. Der Lastausgleich für den Datenverkehr kann über die Services hinweg erfolgen. Allerdings empfehlen wir aufgrund der Latenz- und Bandbreitenunterschiede keinen Lastausgleich zwischen Direct Connect und VPN.

## Implementierungsschritte

- Schätzen Sie die Anforderungen an Bandbreite und Latenz für Ihre bestehenden Anwendungen ab.
  - Für bestehende Workloads, die auf AWS umgestellt werden, nutzen Sie die Daten aus Ihren internen Systemen zur Überwachung des Netzwerks.
  - Bei neuen oder bestehenden Workloads, für die Sie keine Monitoring-Daten haben, beraten Sie sich mit den Besitzern der Produkte, um angemessene Metriken für die Leistung zu bestimmen und ein gutes Benutzererlebnis zu gewährleisten.
- Wählen Sie eine dedizierte Verbindung oder ein VPN als Konnektivitätsoption aus. Je nach den Anforderungen der Workload (Verschlüsselung, Bandbreite und Traffic-Bedarf) können Sie entweder AWS Direct Connect oder [Site-to-Site VPN](#) (oder beides) auswählen. Das folgende Diagramm kann Ihnen bei der Wahl der geeigneten Verbindungsart helfen.
  - [AWS Direct Connect](#) liefert dedizierte Konnektivität für die AWS-Umgebung, von 50 Mbit/s bis zu 100 Gbit/s, entweder über dedizierte Verbindungen oder über gehostete Verbindungen. So erhalten Sie eine verwaltete und kontrollierte Latenz und bereitgestellte Bandbreite, damit sich Ihre Workload effizient mit anderen Umgebungen verbinden kann. Mit einem AWS Direct Connect-Partner können Sie eine End-to-End-Konnektivität aus mehreren Umgebungen nutzen und so ein erweitertes Netzwerk mit konsistenter Leistung bereitstellen. AWS bietet eine Skalierung der Bandbreite für Direct Connect-Verbindungen entweder über native 100 Gbit/s, Link Aggregation Group (LAG) oder BGP Equal-Cost Multipath (ECMP).
  - Die AWS [Site-to-Site VPN](#) bietet einen verwalteten VPN-Service, der das IPsec (Internet Protocol Security) unterstützt. Wenn eine VPN-Verbindung erstellt wird, besteht die VPN-Verbindung aus zwei Tunneln, um eine hohe Verfügbarkeit zu gewährleisten.
- Folgen Sie der AWS-Dokumentation, um eine geeignete Verbindungsoption auszuwählen:
  - Wenn Sie sich für die Verwendung von Direct Connect entscheiden, wählen Sie die entsprechende Bandbreite für Ihre Konnektivität aus.
  - Wenn Sie eine AWS Site-to-Site VPN über mehrere Standorte hinweg nutzen, um eine Verbindung mit einer AWS-Region herzustellen, verwenden Sie eine [beschleunigte Site-to-Site-VPN-Verbindung](#), um die Netzwerkleistung zu verbessern.

- Wenn Ihr Netzwerkdesign aus einer IPsec-VPN-Verbindung über [AWS Direct Connect](#) besteht, sollten Sie die Verwendung von privatem IP-VPN in Betracht ziehen, um die Sicherheit zu verbessern und eine Segmentierung zu erzielen. [AWS Site-to-Site Private IP VPN](#) wird auf einer virtuellen Transit-Schnittstelle (VIF) bereitgestellt.
- [AWS Direct Connect SiteLink](#) ermöglicht die Schaffung redundanter Verbindungen mit niedriger Latenz zwischen Ihren Rechenzentren weltweit, indem Daten über den schnellsten Weg zwischen [AWS Direct Connect-Standorten](#) gesendet werden, unter Umgehung von AWS-Regionen.
- Überprüfen Sie Ihr Konnektivitäts-Setup, bevor Sie es in der Produktion einsetzen. Führen Sie Sicherheits- und Leistungstests durch, um sicherzustellen, dass das Setup Ihre Anforderungen an Bandbreite, Zuverlässigkeit, Latenz und Compliance erfüllt.
- Überwachen Sie regelmäßig die Leistung und Nutzung Ihrer Konnektivität und optimieren Sie sie bei Bedarf.

Flussdiagramm zur deterministischen Leistung

## Ressourcen

Zugehörige Dokumente:

- [Netzwerkprodukte mit AWS](#)
- [AWS Transit Gateway](#)
- [VPC-Endpunkte](#)
- [Building a Scalable and Secure Multi-VPC AWS Network Infrastructure](#)
- [Client-VPN](#)

Zugehörige Videos:

- [AWS re:Invent 2023 – Aufbau einer hybriden Netzwerkkonnektivität mit AWS](#)
- [AWS re:Invent 2023 – Sichere Remote-Verbindung zu AWS](#)
- [AWS re:Invent 2022 – Optimizing performance with Amazon CloudFront](#)
- [AWS re:Invent 2019 – Konnektivität mit AWS und hybriden AWS-Netzwerkarchitekturen](#)

- [AWS re:Invent 2020 – AWS Transit Gateway Connect](#)

Zugehörige Beispiele:

- [AWS Transit Gateway und skalierbare Sicherheitslösungen](#)
- [AWS Workshops zu -Netzwerken](#)

## PERF04-BP04 Lastausgleich verwenden, um den Datenverkehr auf mehrere Ressourcen zu verteilen

Verteilen Sie den Datenverkehr auf mehrere Ressourcen oder Services, damit Ihren Workload von der Elastizität der Cloud zu profitieren kann. Sie können den Lastausgleich auch nutzen, um die Terminierung von Verschlüsselung auszulagern. So lässt sich die Leistung und Zuverlässigkeit optimieren und der Datenverkehr effektiv verwalten und weiterleiten.

Typische Anti-Muster:

- Sie berücksichtigen bei der Wahl des Load-Balancer-Typs nicht die Anforderungen Ihrer Workload.
- Sie nutzen die Features des Load Balancers nicht zur Optimierung der Leistung.
- Die Workload ist direkt mit dem Internet verbunden, ohne dass ein Load Balancer zum Einsatz kommt.
- Sie leiten den gesamten Internetverkehr über vorhandene Load Balancer weiter.
- Sie nutzen einen generischen TCP-Lastausgleich und lassen die SSL-Verschlüsselung von den einzelnen Rechenknoten verarbeiten.

Vorteile der Nutzung dieser bewährten Methode: Ein Load Balancer verarbeitet die variierende Last des Anwendungsdatenverkehrs in einer einzigen oder in mehreren Availability Zones und ermöglicht eine hohe Verfügbarkeit, Auto Scaling sowie eine bessere Nutzung für Ihre Workload.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

### Implementierungsleitfaden

Load Balancer fungieren als Eingangspunkt für Ihre Workload und verteilen den Datenverkehr von dort aus auf Ihre Backend-Ziele – wie Computing-Instances oder Container –, um die Nutzung zu verbessern.

Die Wahl des richtigen Load-Balancer-Typs ist der erste Schritt zur Optimierung Ihrer Architektur. Starten Sie mit einer Auflistung Ihrer Workload-Merkmale wie Protokoll (z. B. TCP, HTTP, TLS oder WebSockets), Zieltyp (z. B. Instances, Container oder Serverless), Anwendungsanforderungen (z. B. langfristige Verbindungen, Benutzerauthentifizierung oder Stickiness) und Platzierung (z. B. Region, lokale Zone, Outposts oder Zonenisolierung).

AWS stellt für Ihre Anwendungen mehrere Modelle zur Verwendung des Lastausgleichs bereit. [Application Load Balancer](#) eignet sich optimal für den Lastausgleich von HTTP- und HTTPS-Datenverkehr. Sie profitieren dadurch von einer erweiterten Anfrageweiterleitung, die es Ihnen ermöglicht, moderne Anwendungsarchitekturen mit Microservices und Containern bereitzustellen.

[Network Load Balancer](#) eignet sich optimal für den Lastausgleich von TCP-Datenverkehr, wenn eine hohe Leistung erforderlich ist. Hiermit lassen sich mit konstant geringer Latenz Millionen Anforderungen pro Sekunde und plötzliche Datenverkehrsspitzen oder schwankende Datenverkehrsmuster verarbeiten.

[Elastic Load Balancing](#) ermöglicht die integrierte Zertifikatverwaltung und SSL/TLS-Entschlüsselung. Auf diese Weise können Sie die SSL-Einstellungen des Load Balancer flexibel zentral verwalten und CPU-intensive Arbeitsschritte für Ihre Workload auslagern.

Nachdem Sie sich für den richtigen Load Balancer entschieden haben, können Sie damit beginnen, seine Features zu nutzen, um die Belastung Ihres Backends durch den Datenverkehr zu verringern.

So können Sie beispielsweise sowohl mit Application Load Balancer (ALB) als auch mit Network Load Balancer (NLB) die SSL/TLS-Verschlüsselung auslagern, was die Möglichkeit bietet, den CPU-intensiven TLS-Handshake bei Ihren Zielen zu vermeiden und die Verwaltung der Zertifikate zu verbessern.

Wenn Sie SSL/TLS-Offloading in Ihrem Load Balancer konfigurieren, übernimmt dieser die Verschlüsselung des Datenverkehrs von und zu den Clients. Er leitet den Datenverkehr dann unverschlüsselt an Ihre Backends weiter, wodurch Ihre Backend-Ressourcen entlastet werden und die Reaktionszeit für die Clients verbessert wird.

Application Load Balancer kann außerdem HTTP/2-Datenverkehr ausliefern, ohne dass Sie ihn auf Ihren Zielen unterstützen müssen. Diese einfache Entscheidung kann die Reaktionszeit Ihrer Anwendung verbessern, da HTTP/2 TCP-Verbindungen effizienter nutzt.

Bei der Definition der Architektur sollten Sie die Anforderungen an die Latenz Ihrer Workload berücksichtigen. Wenn Sie beispielsweise eine latenzempfindliche Anwendung haben, können Sie

sich für Network Load Balancer mit einer extrem niedrigen Latenz entscheiden. Alternativ können Sie Ihre Workload auch näher an Ihre Kunden heranbringen, indem Sie Application Load Balancer in [AWS Local Zones](#) oder sogar in [AWS Outposts](#) nutzen.

Eine weitere Überlegung für latenzempfindliche Workloads ist der zonenübergreifende Lastausgleich. Beim zonenübergreifenden Lastausgleich nimmt jeder Load-Balancer-Knoten eine Verteilung des Datenverkehrs auf die registrierten Ziele in allen zulässigen Availability Zones vor.

Verwenden Sie die Auto-Scaling-Integration für Ihren Load Balancer. Einer der Schlüssel für ein leistungsfähiges System ist die richtige Größenanpassung Ihrer Backend-Ressourcen. Zu diesem Zweck können Sie Load-Balancer-Integrationen für Backend-Zielressourcen nutzen. Mithilfe der Load-Balancer-Integration mit Auto-Scaling-Gruppen werden Ziele je nach Bedarf als Reaktion auf den eingehenden Datenverkehr zum Load Balancer hinzugefügt oder aus ihm entfernt. Load Balancer können für containerisierte Workloads außerdem mit [Amazon ECS](#) und [Amazon EKS](#) integriert werden.

- [Amazon ECS – Service-Lastausgleich](#)
- [Application Load Balancing in Amazon EKS](#)
- [Network Load Balancing in Amazon EKS](#)

## Implementierungsschritte

- Definieren Sie Ihre Anforderungen an den Lastausgleich, einschließlich Datenverkehrsvolumen, Verfügbarkeit und Anwendungsskalierbarkeit.
- Wählen Sie den richtigen Load-Balancer-Typ für Ihre Anwendung.
  - Verwenden Sie den Application Load Balancer für HTTP/HTTPS-Workloads.
  - Verwenden Sie den Network Load Balancer für Nicht-HTTP-Workloads, die TCP oder UDP nutzen.
  - Verwenden Sie eine Kombination aus beidem ([ALB als Ziel des NLB](#)), wenn Sie die Features beider Produkte nutzen möchten. Dies ist zum Beispiel möglich, wenn Sie die statischen IP-Adressen von NLB zusammen mit dem HTTP-Header-basierten Routing von ALB verwenden möchten oder wenn Sie Ihre HTTP-Workload für einen [AWS PrivateLink](#) bereitstellen wollen.
  - Einen vollständigen Vergleich der Load Balancer finden Sie unter [ELB-Produktvergleich](#).
- Verwenden Sie nach Möglichkeit SSL/TLS-Offloading.
  - Konfigurieren Sie HTTPS/TLS-Listeners mit [Application Load Balancer](#) und [Network Load Balancer](#), die in [AWS Certificate Manager](#) integriert sind.

- Beachten Sie, dass einige Workloads aus Compliancegründen eine Ende-zu-Ende-Verschlüsselung benötigen können. In diesem Fall ist es erforderlich, die Verschlüsselung an den Zielen zuzulassen.
- Bewährte Methoden für die Sicherheit finden Sie unter [SEC09-BP02 Erzwingen einer Verschlüsselung bei der Übertragung](#).
- Wählen Sie den richtigen Routing-Algorithmus (nur ALB) aus.
  - Der Routing-Algorithmus kann einen entscheidenden Einfluss darauf haben, wie gut Ihre Backend-Ziele ausgelastet sind und wie sie die Leistung beeinflussen. ALB bietet beispielsweise [zwei Optionen für Routing-Algorithmen](#):
    - Am wenigsten ausstehende Anfragen: Verwenden Sie diese Option, um eine bessere Verteilung der Last auf Ihre Backend-Ziele zu erreichen, wenn die Anfragen für Ihre Anwendung unterschiedlich komplex sind oder Ihre Ziele unterschiedliche Kapazitäten für die Verarbeitung haben.
    - Round Robin: Verwenden Sie diese Option, wenn die Anfragen und Ziele ähnlich sind oder wenn Sie die Anfragen gleichmäßig auf die Ziele verteilen müssen.
- Ziehen Sie eine zonenübergreifende Verarbeitung oder Zonenisolierung in Betracht.
  - Verwenden Sie die deaktivierte zonenübergreifende Isolierung (Zonenisolierung), um die Latenz zu verbessern und Domains mit Zonenfehlern zu vermeiden. Sie ist standardmäßig in NLB deaktiviert und Sie können sie in [ALB pro Zielgruppe ausschalten](#).
  - Verwenden Sie die aktivierte zonenübergreifende Verarbeitung für eine höhere Verfügbarkeit und Flexibilität. Standardmäßig ist Cross-Zone für ALB aktiviert und [Sie können sie in NLB pro Zielgruppe aktivieren](#).
- Aktivieren Sie HTTP-Keep-Alives für Ihre HTTP-Workloads (nur ALB). Mit diesem Feature kann der Load Balancer Backend-Verbindungen wiederverwenden, bis die Keep-Alive-Zeit abgelaufen ist, wodurch sich Ihre HTTP-Anfrage- und Reaktionszeiten verbessern und die Auslastung der Ressourcen auf Ihren Backend-Zielen reduziert wird. Details zu dieser Funktion für Apache und Nginx finden Sie unter [Was sind die optimalen Einstellungen für die Verwendung von Apache oder NGINX als Backend-Server für ELB?](#)
- Aktivieren Sie die Überwachung für Ihren Load Balancer.
  - Aktivieren Sie die Zugriffsprotokolle für Ihren [Application Load Balancer](#) und [Network Load Balancer](#).
  - Die wichtigsten zu berücksichtigenden Elemente für ALB sind `request_processing_time`, `request_processing_time` und `response_processing_time`.
  - Die wichtigsten Elemente für NLB sind `connection_time` und `tls_handshake_time`.

- Bereiten Sie sich darauf vor, die Protokolle bei Bedarf abfragen zu können. Sie können mittels Amazon Athena sowohl [ALB-Protokolle](#) als auch [NLB-Protokolle](#) abfragen.
- Erstellen Sie Warnungen für leistungsbezogene Metriken wie [TargetResponseTime für ALB](#).

## Ressourcen

### Zugehörige Dokumente:

- [ELB-Produktvergleich](#)
- [AWS Global Infrastructure](#)
- [Verbesserung der Leistung und Senkung der Kosten durch Availability Zone-Affinität](#)
- [Step by step for Log Analysis with Amazon Athena](#)
- [Abfragen von Application-Load-Balancer-Protokollen](#)
- [Überwachen Ihrer Application Load Balancer](#)
- [Monitor your Network Load Balancer](#)
- [Um den Datenverkehr über die Instances in Ihrer Auto-Scaling-Gruppe zu verteilen, verwenden Sie Elastic-Load-Balancing](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Welche Vorteile bietet die Vernetzung für Ihre Anwendung?](#)
- [AWS re:Inforce 20: How to use Elastic Load Balancing to enhance your security posture at scale](#)
- [AWS re:Invent 2018: Elastic Load Balancing: Deep Dive and Best Practices](#)
- [AWS re:Invent 2021 – So wählen Sie den richtigen Load Balancer für Ihre AWS-Workloads aus](#)
- [AWS re:Invent 2019: Holen Sie das Beste aus Elastic Load Balancing für verschiedene Workloads heraus](#)

### Zugehörige Beispiele:

- [Gateway Load Balancer](#)
- [CDK and CloudFormation samples for Log Analysis with Amazon Athena](#)

## PERF04-BP05 Auswählen leistungsfördernder Netzwerkprotokolle

Treffen Sie Entscheidungen über Protokolle für die Kommunikation zwischen Systemen und Netzwerken auf Grundlage der Auswirkungen, die sich für die Leistung der Workload ergeben.

In Bezug auf die Erzielung eines höheren Durchsatzes besteht eine Beziehung zwischen der Latenz und der Bandbreite. Wenn Ihr File Transfer über TCP (Transmission Control Protocol) erfolgt, verringern höhere Latenzen höchstwahrscheinlich den gesamten Durchsatz. Es gibt verschiedene Ansätze, dies mit der TCP-Optimierung und optimierten Übertragungsprotokollen zu lösen. Eine Lösung besteht jedoch in der Verwendung des User Datagram Protocol (UDP).

Typische Anti-Muster:

- Sie verwenden TCP unabhängig von den Leistungsanforderungen für alle Workloads.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie sicherstellen, dass ein geeignetes Protokoll für die Kommunikation zwischen Benutzern und Workload-Komponenten verwendet wird, können Sie das Benutzererlebnis für Ihre Anwendungen insgesamt verbessern. Das verbindungslose UDP ermöglicht zwar beispielsweise eine hohe Geschwindigkeit, bietet aber weder eine erneute Übertragung noch hohe Zuverlässigkeit. TCP ist ein Protokoll mit vollem Funktionsumfang, bringt jedoch einen größeren Overhead für die Verarbeitung der Pakete mit sich.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

### Implementierungsleitfaden

Wenn Sie in der Lage sind, verschiedene Protokolle für Ihre Anwendung auszuwählen, und Sie über Fachwissen in diesem Bereich verfügen, optimieren Sie Ihre Anwendungs- und Endbenutzererfahrung, indem Sie ein anderes Protokoll verwenden. Beachten Sie, dass dieser Ansatz mit erheblichen Schwierigkeiten verbunden ist und nur versucht werden sollte, wenn Sie Ihre Anwendung zuvor auf andere Weise optimiert haben.

Um die Leistung Ihrer Workload zu verbessern, sollten Sie in erster Linie die Anforderungen an die Latenz und den Durchsatz kennen und dann Netzwerkprotokolle auswählen, die die Leistung optimieren.

Wann sollten Sie TCP verwenden

TCP bietet eine zuverlässige Zustellung von Daten und kann für die Kommunikation zwischen Workload-Komponenten verwendet werden, bei denen die Zuverlässigkeit und die garantierte

Zustellung von Daten wichtig sind. Viele webbasierte Anwendungen verlassen sich auf TCP-basierte Protokolle wie HTTP und HTTPS, um TCP-Sockets für die Kommunikation zwischen Anwendungskomponenten zu öffnen. E-Mail- und Dateidatenübertragung sind gängige Anwendungen, die auch TCP verwenden, da es sich um einen einfachen und zuverlässigen Übertragungsmechanismus zwischen Anwendungskomponenten handelt. Die Verwendung von TLS mit TCP kann zu einem gewissen Overhead bei der Kommunikation führen, was eine erhöhte Latenz und einen verringerten Durchsatz zur Folge haben kann. Sie bietet jedoch den Vorteil der Sicherheit. Der Overhead entsteht vor allem durch den zusätzlichen Aufwand des Handshake-Prozesses, der mehrere Roundtrips in Anspruch nehmen kann. Sobald der Handshake abgeschlossen ist, ist der Overhead für die Ver- und Entschlüsselung der Daten relativ gering.

Wann sollten Sie UDP verwenden

UDP ist ein verbindungsloses Protokoll und eignet sich daher für Anwendungen, die eine schnelle, effiziente Übertragung benötigen, wie z. B. die Protokollierung, die Überwachung und VoIP-Daten. Ziehen Sie die Verwendung von UDP auch in Betracht, wenn Sie Workload-Komponenten haben, die auf kleine Abfragen von einer großen Anzahl von Clients reagieren, um eine optimale Leistung der Workload zu gewährleisten. Datagram Transport Layer Security (DTLS) ist die UDP-Entsprechung von Transport Layer Security (TLS). Bei der Verwendung von DTLS mit UDP entsteht der Overhead durch die Verschlüsselung und Entschlüsselung der Daten, da der Handshake-Prozess vereinfacht ist. DTLS fügt den UDP-Paketen außerdem einen geringen Overhead hinzu, da es zusätzliche Felder zur Angabe der Sicherheitsparameter und zur Erkennung von Manipulationen umfasst.

Wann sollten Sie SRD verwenden

Scalable Reliable Datagram (SRD) ist ein Netzwerktransportprotokoll, das für Workloads mit hohem Durchsatz optimiert ist, da es in der Lage ist, den Datenverkehr über mehrere Pfade zu verteilen und sich schnell von Paketverlusten oder Verbindungsfehlern zu erholen. SRD eignet sich daher am besten für HPC-Workloads (High Performance Computing), die einen hohen Durchsatz und eine geringe Latenz bei der Kommunikation zwischen Computing-Knoten erfordern. Dazu gehören z. B. parallele Verarbeitungsaufgaben wie Simulationen, Modellierung und Datenanalyse, bei denen große Mengen an Daten zwischen den Knoten übertragen werden müssen.

## Implementierungsschritte

- Verwenden Sie die Services [AWS Global Accelerator](#) und [AWS Transfer Family](#), um den Durchsatz Ihrer Anwendungen für den File Transfer online zu verbessern. Der AWS Global Accelerator-Service hilft Ihnen, die Latenz zwischen Ihren Client-Geräten und Ihrer Workload auf AWS zu verringern. Mit AWS Transfer Family können Sie TCP-basierte Protokolle wie Secure Shell

File Transfer Protocol (SFTP) und File Transfer Protocol over SSL (FTPS) verwenden, um Ihre Dateiübertragungen zu AWS-Speicherservices sicher zu skalieren und zu verwalten.

- Bestimmen Sie anhand der Netzwerklatenz, ob TCP für die Kommunikation zwischen Workload-Komponenten geeignet ist. Wenn die Netzwerklatenz zwischen Ihrer Client-Anwendung und dem Server hoch ist, kann der TCP-Drei-Wege-Handshake einige Zeit in Anspruch nehmen, was sich auf die Reaktionsfähigkeit Ihrer Anwendung auswirkt. Metriken wie Time to First Byte (TTFB) und Round-Trip-Zeit (RTT) können zur Messung der Netzwerklatenz verwendet werden. Wenn Ihre Workload dynamische Inhalte für Benutzer bereitstellt, sollten Sie die Verwendung von [Amazon CloudFront](#) in Betracht ziehen. So wird eine dauerhafte Verbindung zu jeder Quelle für dynamische Inhalte hergestellt, um die Zeit für den Verbindungsaufbau zu vermeiden, die sonst jede Client-Anfrage verlangsamen würde.
- Die Verwendung von TLS mit TCP oder UDP kann aufgrund der Auswirkungen der Ver- und Entschlüsselung zu einer erhöhten Latenz und einem reduzierten Durchsatz für Ihre Workload führen. Ziehen Sie für solche Workloads das SSL/TLS-Offloading auf [Elastic Load Balancing](#) in Betracht, um die Leistung der Workload zu verbessern, indem Sie den Load Balancer die SSL/TLS-Verschlüsselung und -Entschlüsselung übernehmen lassen, anstatt dies den Backend-Instances zu überlassen. Dies kann dazu beitragen, die CPU-Auslastung der Backend-Instances zu reduzieren, was die Leistung verbessern und die Kapazität erhöhen kann.
- Verwenden Sie den [Network Load Balancer \(NLB\)](#), um Services bereitzustellen, die auf dem UDP-Protokoll basieren (wie die Authentifizierung und Autorisierung, die Protokollierung, DNS, IoT und das Streamen von Medien), um die Leistung und Zuverlässigkeit Ihrer Workload zu verbessern. Der NLB verteilt den eingehenden UDP-Datenverkehr auf mehrere Ziele, sodass Sie Ihre Workload horizontal skalieren, die Kapazität erhöhen und den Overhead eines einzelnen Ziels reduzieren können.
- Für Ihre HPC-Workloads (High Performance Computing) sollten Sie die Verwendung der Funktionalität [Elastic Network Adapter \(ENA\) Express](#) in Betracht ziehen, die das SRD-Protokoll nutzt, um die Leistung des Netzwerks zu verbessern, indem sie eine höhere Bandbreite für einen einzelnen Datenfluss (25 Gbit/s) und eine niedrigere Latenz (99,9 Perzentil) für den Netzwerkverkehr zwischen EC2-Instances bietet.
- Verwenden Sie den [Application Load Balancer \(ALB\)](#), um Ihren gRPC-Datenverkehr (Remote Procedure Calls) zwischen Workload-Komponenten oder zwischen gRPC-Clients und -Services zu routen und ein Load-Balancing durchzuführen. gRPC verwendet das TCP-basierte HTTP/2-Protokoll für den Transport und bietet Vorteile in Bezug auf die Leistung, wie z. B. einen geringeren Netzwerk-Footprint, Komprimierung, effiziente binäre Serialisierung, Unterstützung zahlreicher Sprachen und bidirektionales Streaming.

## Ressourcen

### Zugehörige Dokumente:

- [Routing von UDP-Verkehr nach Kubernetes](#)
- [Application Load Balancer](#)
- [EC2: Enhanced Networking unter Linux](#)
- [EC2: Enhanced Networking unter Windows](#)
- [EC2: Platzierungsgruppen](#)
- [Enabling Enhanced Networking with the Elastic Network Adapter \(ENA\) on Linux Instances](#)
- [Network Load Balancer](#)
- [Netzwerkprodukte mit AWS](#)
- [Umstellung auf latenzbasiertes Routing in Amazon Route 53](#)
- [VPC-Endpunkte](#)

### Zugehörige Videos:

- [AWS re:Invent 2022 – Scaling network performance on next-gen Amazon Elastic Compute Cloud instances](#)
- [AWS re:Invent 2022 – Grundlagen der Anwendungsnetzwerke](#)

### Zugehörige Beispiele:

- [AWS Transit Gateway und skalierbare Sicherheitslösungen](#)
- [Workshops zu AWS-Netzwerken](#)

## PERF04-BP06 Wählen Sie den Standort Ihres Workloads basierend auf den Netzwerkanforderungen

Evaluieren Sie Optionen für die Platzierung von Ressourcen, um die Latenz im Netzwerk zu verringern und den Durchsatz zu verbessern und so ein optimales Benutzererlebnis durch kürzere Seitenlade- und Datentransferzeiten zu gewährleisten.

### Typische Anti-Muster:

- Sie konsolidieren alle Workload-Ressourcen an einem geografischen Standort.
- Sie haben sich für die Region entschieden, die Ihrem Standort, aber nicht dem Workload-Endbenutzer, am nächsten liegt.

Vorteile der Nutzung dieser bewährten Methode: Die Benutzererfahrung wird stark von der Latenz zwischen dem Benutzer und Ihrer Anwendung beeinflusst. Durch die Verwendung eines geeigneten AWS-Regionen und AWS privaten globalen Netzwerks können Sie die Latenz reduzieren und Remote-Benutzern ein besseres Benutzererlebnis bieten.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Ressourcen, wie EC2 Amazon-Instances, werden in Availability Zones innerhalb [AWS-Regionen](#), [AWS Local Zones](#) oder [AWS Wavelength](#)Zonen platziert. [AWS Outposts](#) Die Auswahl dieses Standorts beeinflusst die Latenz des Netzwerks und den Durchsatz vom Standort des Benutzers aus. Edge-Services wie [Amazon CloudFront](#) [AWS Global Accelerator](#) können auch zur Verbesserung der Netzwerkleistung verwendet werden, indem sie entweder Inhalte an Edge-Standorten zwischenspeichern oder Benutzern einen optimalen Pfad zur Arbeitslast über das AWS globale Netzwerk bieten.

Amazon EC2 bietet Platzierungsgruppen zum Netzwerken an. Eine Platzierungsgruppe ist eine logische Gruppierung von Instances, um die Latenz zu verringern. Durch die Verwendung von Platzierungsgruppen mit unterstützten Instance-Typen und einem Elastic Network Adapter (ENA) können Workloads an einem 25-Gbit/s-Netzwerk mit geringer Latenz und reduziertem Jitter teilnehmen. Platzierungsgruppen werden für Workloads empfohlen, für die eine niedrige Netzwerklatenz bzw. ein hoher Durchsatz von Vorteil sind.

[Latenzempfindliche Dienste werden an Edge-Standorten über ein AWS globales Netzwerk wie Amazon bereitgestellt. CloudFront](#) Diese Edge-Standorte bieten in der Regel Dienste wie Content Delivery Network (CDN) und Domain Name System (DNS). Da sich diese Dienste an der Peripherie befinden, können Workloads mit geringer Latenz auf Anfragen nach Inhalten oder Lösungen reagieren. Es sind auch geografische Services wie das Geo-Targeting von Inhalten (Bereitstellung unterschiedlicher Inhalte gemäß dem Standort von Endbenutzern) oder die latenzbasierte Weiterleitung von Endbenutzern zur nächsten Region (minimale Latenz) verfügbar.

Verwenden Sie Edge-Services, um die Latenz zu reduzieren und das Caching von Inhalten zu ermöglichen. Konfigurieren Sie die Cachesteuerung DNS sowohl für als auch für HTTP/ korrekt HTTPS, um den größtmöglichen Nutzen aus diesen Ansätzen zu ziehen.

## Implementierungsschritte

- Erfassen Sie Informationen über den an den Netzwerkschnittstellen ein- und ausgehenden IP-Datenverkehr.
  - [Protokollierung von IP-Verkehr mithilfe von VPC Flow Logs](#)
  - [Wie wird die Client-IP-Adresse aufbewahrt in AWS Global Accelerator](#)
- Analysieren Sie die Netzwerkzugriffsmuster in Ihrer Workload, um zu ermitteln, wie die Benutzer Ihre Anwendung verwenden.
  - Verwenden Sie Überwachungstools wie [Amazon CloudWatch](#) und [AWS CloudTrail](#), um Daten zu Netzwerkaktivitäten zu sammeln.
  - Analysen Sie die Daten, um das Netzwerkzugriffsmuster zu identifizieren.
- Wählen Sie Regionen für Ihre Workload-Bereitstellung auf der Grundlage der folgenden zentralen Elemente aus:
  - Standort Ihrer Daten: Für datenintensive Anwendungen (wie etwa Big Data oder Machine Learning) sollte der Anwendungscode so nahe wie möglich zu den Daten ausgeführt werden.
  - Standort Ihrer Benutzer: Wählen Sie für benutzerseitige Anwendungen eine Region (oder Regionen) in der Nähe der Benutzer der Workload.
  - Weitere Einschränkungen: Berücksichtigen Sie Einschränkungen wie Kosten und Compliance, wie unter [Relevante Aspekte bei der Wahl einer Region für Ihre Workloads](#) erläutert.
- Verwenden Sie [AWS Local Zones](#), um Workloads wie Video-Rendering auszuführen. Mit Local Zones können Sie von allen Vorteilen profitieren, die sich durch die Platzierung der Datenverarbeitungs- und Speicherressourcen in der Nähe Ihrer Endbenutzer ergeben.
- Verwenden Sie [AWS Outposts](#) für Workloads, die On-Premises verarbeitet werden müssen und die Sie nahtlos mit Ihren restlichen Workloads in AWS ausführen möchten.
- ultra-low-latency Für Anwendungen wie hochauflösendes Live-Videostreaming, High-Fidelity-Audio und Augmented Reality oder Virtual Reality (AR/VR) sind 5G-Geräte erforderlich. Beachten Sie bei solchen Anwendungen Folgendes: [AWS Wavelength](#) AWS Wavelength bettet AWS Rechen- und Speicherdienste in 5G-Netzwerke ein und bietet so eine mobile Edge-Computing-Infrastruktur für die Entwicklung, Bereitstellung und Skalierung von ultra-low-latency Anwendungen.

- Verwenden Sie lokale Zwischenspeicherung oder [AWS -Zwischenspeicherung](#) für häufig genutzte Ressourcen zur Verbesserung der Leistung, zur Verringerung von Datenverschiebungen und zur Reduzierung der Umweltauswirkungen.

Service	Wann sollte dies verwendet werden?
<a href="#">Amazon CloudFront</a>	Wird verwendet, um statische Inhalte wie Bilder, Skripte und Videos sowie dynamische Inhalte wie API Antworten oder Webanwendungen zwischenzuspeichern.
<a href="#">Amazon ElastiCache</a>	Verwenden Sie dies für die Zwischenspeicherung von Inhalten für Webanwendungen.
<a href="#">DynamoDB Accelerator</a>	Verwenden Sie dies für die Add-in-Speicher-Beschleunigung für Ihre DynamoDB-Tabellen.

- Nutzen Sie Services, die Ihnen dabei helfen können, Code näher an den Benutzern Ihrer Workload auszuführen:

Service	Wann sollte dies verwendet werden?
<a href="#">Lambda@Edge</a>	Verwenden Sie dies für rechenintensive Anwendungen, die initiiert werden, wenn sich Objekte nicht im Zwischenspeicher befinden.
<a href="#">CloudFront Amazon-Funktionen</a>	Wird für einfache Anwendungsfälle wie HTTP Anfragen oder Antwortmanipulationen verwendet, die durch kurzlebige Funktionen ausgelöst werden können.
<a href="#">AWS IoT Greengrass</a>	Verwenden Sie dies für die Ausführung lokaler Rechenoperationen, Messaging sowie die Datenzwischenspeicherung für verbundene Geräte.

- Einige Anwendungen benötigen feste Zugangspunkte oder eine höhere Leistung. Bei diesen müssen First-Byte-Latenz der Jitter verringert und der Durchsatz erhöht werden. Diese Anwendungen können von Netzwerkdiensten profitieren, die statische Anycast-IP-Adressen und TCP Terminierung an Edge-Standorten bereitstellen. [AWS Global Accelerator](#) kann die Leistung Ihrer Anwendungen um bis zu 60% verbessern und ein schnelles Failover für Architekturen mit mehreren Regionen ermöglichen. AWS Global Accelerator stellt Ihnen statische Anycast-IP-Adressen zur Verfügung, die als fester Einstiegspunkt für Ihre in einer oder mehreren gehosteten Anwendungen dienen. AWS-Regionen Diese IP-Adressen ermöglichen es, dass der Datenverkehr so nah wie möglich an Ihren Benutzern in das AWS globale Netzwerk gelangt. AWS Global Accelerator reduziert die Zeit für den anfänglichen Verbindungsaufbau, indem eine TCP Verbindung zwischen dem Client und dem AWS Edge-Standort hergestellt wird, der dem Client am nächsten ist. Prüfen Sie die Verwendung von AWS Global Accelerator , um die Leistung Ihrer TCP UDP /-Workloads zu verbessern und ein schnelles Failover für Architekturen mit mehreren Regionen zu ermöglichen.

## Ressourcen

Zugehörige bewährte Methoden:

- [COST07-BP02 Implementieren Sie Regionen auf der Grundlage der Kosten](#)
- [COST08-BP03 Implementieren Sie Dienste zur Senkung der Datenübertragungskosten](#)
- [REL10-BP01 Stellen Sie den Workload an mehreren Standorten bereit](#)
- [REL10-BP02 Wählen Sie die geeigneten Standorte für Ihre Bereitstellung an mehreren Standorten aus](#)
- [SUS01-BP01 Wählen Sie die Region auf der Grundlage von Geschäftsanforderungen und Nachhaltigkeitszielen](#)
- [SUS02-BP04 Optimieren Sie die geografische Verteilung von Workloads auf der Grundlage ihrer Netzwerkanforderungen](#)
- [SUS04-BP07 Minimierung der Datenbewegung zwischen Netzwerken](#)

Zugehörige Dokumente:

- [AWS Globale Infrastruktur](#)
- [AWS Local Zones und AWS Outposts Auswahl der richtigen Technologie für Ihren Edge-Workload](#)
- [Platzierungsgruppen](#)

- [AWS Local Zones](#)
- [AWS Outposts](#)
- [AWS Wavelength](#)
- [Amazon CloudFront](#)
- [AWS Global Accelerator](#)
- [AWS Direct Connect](#)
- [AWS Site-to-Site VPN](#)
- [Amazon Route 53](#)

#### Zugehörige Videos:

- [AWS Erklärvideo zu Local Zones](#)
- [AWS Outposts: Overview and How it Works](#)
- [AWS re:Invent 2023 — Eine Migrationsstrategie für Edge-Workloads und lokale Workloads](#)
- [AWS re:Invent 2021 — AWS Outposts: Das Erlebnis vor Ort umsetzen AWS](#)
- [AWS re:Invent 2020: AWS Wavelength: Führen Sie Apps mit extrem niedriger Latenz am 5G-Edge aus](#)
- [AWS re:Invent 2022 — AWS Local Zones: Entwicklung von Anwendungen für eine verteilte Kante](#)
- [AWS re:Invent 2021 — Websites mit niedriger Latenz mit Amazon erstellen CloudFront](#)
- [AWS re:Invent 2022 — Verbessern Sie Leistung und Verfügbarkeit mit AWS Global Accelerator](#)
- [AWS re:Invent 2022 — Bauen Sie Ihr globales Wide Area Network auf mit AWS](#)
- [AWS re:Invent 2020: Globales Verkehrsmanagement mit Amazon Route 53](#)

#### Zugehörige Beispiele:

- [AWS Global Accelerator Workshop zu kundenspezifischem Routing](#)
- [Verarbeitung von Rewrites und Redirects mit Edge-Funktionen](#)

# PERF04-BP07 Optimieren der Netzwerkkonfiguration basierend auf Metriken

Treffen Sie anhand der erfassten und analysierten Daten fundierte Entscheidungen zum Optimieren Ihrer Netzwerkkonfiguration.

Typische Anti-Muster:

- Sie gehen davon aus, dass alle leistungsbezogenen Probleme auf Anwendungen zurückzuführen sind.
- Sie testen die Netzwerkleistung ausschließlich an einem Standort nahe der Stelle, an der Sie die Workload bereitgestellt haben.
- Sie verwenden Standardkonfigurationen für alle Netzwerkservices.
- Sie führen eine Überdimensionierung der Netzwerkressourcen durch, um eine ausreichende Kapazität zu gewährleisten.

Vorteile der Nutzung dieser bewährten Methode: Das Sammeln der erforderlichen Metriken Ihres AWS-Netzwerks und die Implementierung von Tools zur Überwachung des Netzwerks bieten Ihnen die Möglichkeit, die Leistung des Netzwerks zu ermitteln und die Netzwerkkonfigurationen zu optimieren.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Niedrig

## Implementierungsleitfaden

Die Überwachung des Datenverkehrs von und zu VPCs, Subnetzen oder Netzwerkschnittstellen ist für das Verständnis der Nutzung von AWS-Netzwerkressourcen und zur Optimierung von Netzwerkkonfigurationen entscheidend. Mit den folgenden AWS-Networking-Tools können Sie Informationen über die Nutzung des Datenverkehrs, den Netzwerkzugriff und die Protokolle genauer untersuchen.

## Implementierungsschritte

- Identifizieren Sie die wichtigsten Leistungsmetriken wie Latenz oder Paketverlust, die erfasst werden müssen. AWS bietet mehrere Tools, die Ihnen bei der Erfassung dieser Metriken helfen können. Mit den folgenden Tools können Sie Informationen über die Nutzung des Datenverkehrs, den Netzwerkzugriff und die Protokolle genauer untersuchen:

AWS-Tool	Aktion
<a href="#">Amazon VPC IP Address Manager.</a>	Nutzen Sie IPAM, um Ihre IP-Adressen für Ihre AWS- und On-Premises-Workloads zu planen, nachzuverfolgen und zu überwachen. Dies ist eine bewährte Methode zur Optimierung der Nutzung und Zuweisung von IP-Adressen.
<a href="#">VPC Flow-Protokolle</a>	Nutzen Sie VPC Flow-Protokolle, um detaillierte Informationen über den Datenverkehr zu und von den Netzwerkschnittstellen in Ihren VPCs zu protokollieren. Mit VPC Flow-Protokollen können Sie restriktive oder zu freizügige Regeln für Sicherheitsgruppen diagnostizieren und die Richtung des Datenverkehrs zu und von den Netzwerkschnittstellen ermitteln.
<a href="#">AWS Transit Gateway-Flow-Protokolle</a>	Nutzen Sie AWS Transit Gateway-Flow-Protokolle, um Informationen über den IP-Datenverkehr zu und von Ihren Transit-Gateways zu erfassen.
<a href="#">DNS-Abfrageprotokollierung</a>	Protokollieren Sie Informationen über von Route 53 empfangene öffentliche oder private DNS-Abfragen. Mit DNS-Protokollen können Sie DNS-Konfigurationen optimieren, indem Sie die angefragte Domain oder Subdomain bzw. die Route-53-Edge-Standorte, die auf DNS-Abfragen geantwortet haben, nachvollziehen.

AWS-Tool	Aktion
<a href="#">Reachability Analyzer</a>	<p>Reachability Analyzer hilft Ihnen, die Erreichbarkeit von Netzwerken zu analysieren und zu debuggen. Reachability Analyzer ist ein Tool zur Konfigurationsanalyse, mit dem Sie Konnektivitätstests zwischen einer Quellressource und einer Zielressource in Ihren VPCs durchführen können. Mit diesem Tool können Sie überprüfen, ob Ihre Netzwerkkonfiguration der geplanten Konnektivität entspricht.</p>
<a href="#">Network Access Analyzer</a>	<p>Mit Network Access Analyzer können Sie den Netzwerkzugriff auf die eigenen Ressourcen analysieren. Mit Network Access Analyzer können Sie Ihre Anforderungen an den Netzwerkzugriff spezifizieren und potenzielle Netzwerkpfade identifizieren, die Ihren Anforderungen nicht entsprechen. Indem Sie Ihre entsprechende Netzwerkkonfiguration optimieren, können Sie den Zustand Ihres Netzwerks nachvollziehen und überprüfen sowie belegen, dass Ihr AWS-Netzwerk Ihre Complianceanforderungen erfüllt.</p>

AWS-Tool	Aktion
<a href="#">Amazon CloudWatch</a>	<p>Verwenden Sie <a href="#">Amazon CloudWatch</a> und aktivieren Sie geeignete Metriken für Netzwerkoptionen. Stellen Sie sicher, dass Sie die richtige Netzwerk-Metrik für Ihre Workload auswählen. Sie können zum Beispiel Metriken für die VPC-Netzwerkadressennutzung, VPC-NAT-Gateways, AWS Transit Gateway, VPN-Tunnel, AWS Network Firewall, Elastic Load Balancing und AWS Direct Connect aktivieren. Die kontinuierliche Überwachung von Metriken ist eine gute Vorgehensweise, um den Status und die Nutzung Ihres Netzwerks zu beobachten und nachzuvollziehen. Sie hilft Ihnen, die Netzwerkkonfiguration auf der Basis Ihrer Beobachtungen zu optimieren.</p>
<a href="#">AWS Network Manager</a>	<p>Mit AWS Network Manager können Sie die Echtzeit- und historische Leistung des <a href="#">globalen AWS-Netzwerks</a> für Betriebs- und Planungszwecke überwachen. Network Manager bietet aggregierte Netzwerklatenz zwischen AWS-Regionen und Availability Zones sowie innerhalb jeder Availability Zone, wodurch Sie besser verstehen können, wie Ihre Anwendungsleistung mit der Leistung des zugrunde liegenden AWS-Netzwerks zusammenhängt.</p>
<a href="#">Amazon CloudWatch RUM</a>	<p>Verwenden Sie Amazon CloudWatch RUM, um die Metriken zu erfassen, die Ihnen die Erkenntnisse liefern, mit denen Sie die Benutzererfahrung identifizieren, verstehen und verbessern können.</p>

- Identifizieren Sie mithilfe von VPC und AWS Transit Gateway Flow Logs Top-Talker und Muster des Anwendungsdatenverkehrs.
- Bewerten und optimieren Sie Ihre aktuelle Netzwerkarchitektur, einschließlich VPCs, Subnetze und Routing. Sie können beispielsweise bewerten, wie unterschiedliches VPC-Peering oder AWS Transit Gateway Ihnen helfen können, das Netzwerk in Ihrer Architektur zu verbessern.
- Bewerten Sie die Routingpfade in Ihrem Netzwerk, um sicherzustellen, dass immer der kürzeste Pfad zwischen Zielen verwendet wird. Network Access Analyzer kann Ihnen dabei helfen.

## Ressourcen

### Zugehörige Dokumente:

- [Öffentliche DNS-Abfrageprotokollierung](#)
- [Was ist IPAM?](#)
- [What is Reachability Analyzer?](#)
- [What is Network Access Analyzer?](#)
- [CloudWatch-Metriken für Ihre VPCs](#)
- [Die Leistung optimieren und die Kosten für Netzwerkanalytik mit VPC-Ablaufprotokollen im Apache-Parquet-Format senken](#)
- [Überwachen von globalen und Kernnetzwerken mit Amazon CloudWatch-Metriken](#)
- [Kontinuierliches Überwachen von Netzwerkdatenverkehr und -ressourcen](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Leitfaden für Entwickler von Cloud-Netzwerken](#)
- [AWS re:Invent 2023 – Sind Sie bereit für Neues? Gestaltung von Netzwerken für Wachstum und Flexibilität](#)
- [AWS re:Invent 2023 – Erweiterte VPC-Designs und neue Funktionen](#)
- [AWS re:Invent 2022 – Ausführliche Beschreibung der AWS-Netzwerkinfrastruktur](#)
- [AWS re:Invent 2020 – Bewährte Methoden für Netzwerke und Tipps für das AWSWell-Architected Framework](#)
- [AWS re:Invent 2020 – Überwachen des Netzwerkdatenverkehrs und Fehlerbehebung](#)

## Zugehörige Beispiele:

- [Workshops zu AWS-Netzwerken](#)
- [AWS Network Monitoring](#)
- [Beobachten und Diagnostizieren Ihres Netzwerks in AWS](#)
- [Finden und Beheben von Netzwerkfehlerkonfigurationen in AWS](#)

# Prozess und Kultur

Bei der Architektur von Workloads gibt es Prinzipien und Methoden, die Sie übernehmen können, um effiziente und leistungsstarke Cloud-Workloads besser zu betreiben. Dieser Schwerpunktbereich bietet bewährte Methoden zur Einführung einer Kultur, die die Leistungseffizienz von Cloud-Workloads fördert.

Beachten Sie beim Aufbau dieser Kultur die folgenden Schlüsselprinzipien:

- **Infrastruktur als Code:** Definieren Sie Ihre Infrastruktur beispielsweise mithilfe von AWS CloudFormation-Vorlagen als Code. Mit Vorlagen können Sie Ihre Infrastruktur zusammen mit Ihrem Anwendungscode und Ihren Konfigurationen per Quellcodeüberwachung verwalten. Dies ermöglicht es Ihnen, dieselben Methoden wie bei der Softwareentwicklung auch auf Ihre Infrastruktur anzuwenden, um von einer schnellen Iteration zu profitieren.
- **Bereitstellungspipeline:** Nutzen Sie zur Bereitstellung der Infrastruktur eine CI/CD-Pipeline (Continuous Integration/Continuous Deployment) wie etwa ein Quellcode-Repository, Build-Systeme sowie automatisierte Bereitstellungs- und Testverfahren. Dies lässt eine reproduzierbare, konsistente und kostengünstige Iteration zu.
- **Gut definierte Metriken:** Richten Sie Ihre Metriken und Überwachungsfunktionen so ein, dass wichtige Leistungskennzahlen (KPIs) erfasst werden. Wir empfehlen die Verwendung technischer und geschäftlicher Metriken. Anhand von wichtigen Metriken für Websites oder mobile Apps wird die Zeit bis zum ersten Byte oder Rendering erfasst. Zu den weiteren allgemein anwendbaren Metriken zählen die Thread-Anzahl, die Garbage Collection-Rate sowie Wartezustände. Anhand von geschäftlichen Metriken wie den aggregierten kumulativen Kosten pro Anforderung können Sie Möglichkeiten zur Kostensenkung ermitteln. Erwägen Sie sorgfältig, wie Metriken interpretiert werden sollen. Sie können beispielsweise anstelle von Durchschnittswerten Maximalwerte oder das 99. Perzentil wählen.
- **Automatische Leistungstests:** Sorgen Sie im Rahmen der Bereitstellung dafür, dass nach dem erfolgreichen Absolvieren der schnelleren Ausführungstests automatisch Leistungstests initiiert werden. Durch die Automatisierung sollte eine neue Umgebung mit entsprechenden Anfangsbedingungen, z. B. Testdaten, entstehen, in der anschließend einige Benchmark- und Lasttests ausgeführt werden. Die Ergebnisse dieser Tests sollten mit dem Build in Verbindung gebracht werden, um Leistungsänderungen verfolgen zu können. Für langwierige Tests können Sie diesen Teil der Pipeline gegenüber dem restlichen Build asynchron ausführen. Sie haben auch die Möglichkeit, Leistungstests über Nacht mit Amazon EC2 Spot Instances auszuführen.

- **Lastgenerierung:** Erstellen Sie eine Reihe von Testskripts zum Replizieren synthetischer oder vorab aufgezeichneter Benutzerreisen. Diese Skripts sollten idempotent und nicht gekoppelt sein. Um gültige Ergebnisse zu erzielen, sind möglicherweise zusätzliche vorbereitende Skripts erforderlich. Die Testskripts sollten das Nutzungsverhalten in der Produktion möglichst authentisch replizieren. Zur Lastgenerierung können Sie Software- oder Software as a Service (SaaS)-Lösungen verwenden. Erwägen Sie die Verwendung von [AWS Marketplace](#)-Lösungen und [Spot Instances](#). Dies können kostengünstige Ansätze zum Generieren der Last sein.
- **Leistungstransparenz:** Wichtige Metriken sollten für das ganze Team sichtbar sein. Dies gilt insbesondere für die Metriken der einzelnen Build-Versionen. Damit lassen sich wichtige positive oder negative Trends erkennen. Wichtig sind auch Metriken zur Anzahl der Fehler oder Ausnahmen, um sicherzustellen, dass das System funktioniert.
- **Visualisierung:** Nutzen Sie Visualisierungstechniken, mit denen Leistungsprobleme, Hotspots, Wartezustände oder niedrige Auslastungen klar aufgezeigt werden. Zeigen Sie Leistungsmetriken in Architekturdiagrammen an. Aufrufgrafiken oder Code können die Problemerkennung beschleunigen.
- **Regelmäßiger Prüfungsprozess:** Wenn Architekturen eine schlechte Leistung aufweisen, liegt dies normalerweise daran, dass ein Prozess zur Überprüfung der Leistung fehlt oder fehlerhaft ist. Falls Sie derartige Probleme mit Ihrer Architektur haben, können Sie jederzeit ein Leistungsprüfverfahren implementieren und somit iterative Verbesserungen fördern.
- **Fortlaufende Optimierung:** Schaffen Sie eine Kultur fortlaufender Optimierung der Leistungseffizienz Ihrer Cloud-Workload.

## Bewährte Methoden

- [PERF05-BP01 Festlegen wichtiger Leistungskennzahlen \(KPIs\) zum Messen des Zustands und der Leistung der Workload](#)
- [PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren](#)
- [PERF05-BP03 Definieren eines Prozesses zum Verbessern der Workload-Leistung](#)
- [PERF05-BP04 Belastungstest Ihr Workload](#)
- [PERF05-BP05 Verwenden von Automatisierung zur proaktiven Behebung leistungsbezogener Probleme](#)
- [PERF05-BP06 Behalten Sie Ihren Workload und Ihre Services up-to-date](#)
- [PERF05-BP07 Regelmäßiges Überprüfen von Metriken](#)

## PERF05-BP01 Festlegen wichtiger Leistungskennzahlen (KPIs) zum Messen des Zustands und der Leistung der Workload

Identifizieren Sie die KPIs, die die Workload-Leistung quantitativ und qualitativ messen. Mithilfe von KPIs können Sie den Zustand und die Leistung einer Workload im Zusammenhang mit einem Geschäftsziel messen.

Typische Anti-Muster:

- Sie überwachen nur Metriken auf Systemebene, um Erkenntnisse über Ihre Workload zu gewinnen, und verstehen den geschäftlichen Einfluss dieser Metriken nicht.
- Sie gehen davon aus, dass Ihre KPIs bereits als standardmäßige Metrikdaten veröffentlicht und geteilt werden.
- Sie definieren keinen quantitativen, messbaren KPI.
- Sie richten KPIs nicht an Geschäftszielen oder -strategien aus.

Vorteile der Nutzung dieser bewährten Methode: Die Identifizierung spezifischer KPIs, die den Zustand und die Leistung der Workload widerspiegeln, hilft Teams dabei, sich auf ihre Prioritäten zu konzentrieren und erfolgreiche Geschäftsergebnisse zu definieren. Das Teilen dieser Metriken mit allen Abteilungen bietet Sichtbarkeit und die Ausrichtung an Grenzwerten, Erwartungen und Geschäftsauswirkungen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

### Implementierungsleitfaden

KPIs helfen Business- und Entwicklungsteams, das Messen von Zielen und Strategien abzustimmen und festzustellen, wie diese Faktoren gemeinsam zu Geschäftsergebnissen beitragen. Beispielsweise könnte eine Website-Workload die Ladezeit der Seite als Indikator für die Gesamtleistung heranziehen. Diese Metrik wäre einer von mehreren Datenpunkten, mit denen das Benutzererlebnis gemessen wird. Zusätzlich zum Ermitteln der Grenzwerte für Seitenladezeiten sollten Sie das gewünschte Resultat dokumentieren bzw. das Geschäftsrisiko, wenn die ideale Leistung nicht erreicht wird. Die lange Ladezeit einer Seite betrifft Ihre Endbenutzer direkt, verringert die Bewertung ihres Benutzererlebnisses und kann zu einem Verlust von Kunden führen. Kombinieren Sie beim Definieren Ihrer KPI-Grenzwerte die Benchmarks der Branche und die Erwartungen Ihrer Endbenutzer. Beispielsweise, wenn die aktuelle Benchmark der Branche das Laden einer Webseite innerhalb von zwei Sekunden ist, Ihre Endbenutzer aber erwarten, dass eine Webseite innerhalb

von einer Sekunde geladen wird, sollten Sie beim Einrichten des KPI beide Datenpunkte in Betracht ziehen.

Ihr Team muss Ihre Workload-KPIs mithilfe von detaillierten Echtzeitdaten und historischen Daten als Referenz evaluieren und Dashboards erstellen, die Metrikberechnungen für Ihre KPI-Daten durchführen, um Einblicke in Betrieb und Auslastung zu erhalten. KPIs sollten dokumentiert werden und Grenzwerte enthalten, die Geschäftsziele und -strategien unterstützen, und sie sollten den Metriken zugeordnet sein, die überwacht werden. KPIs sollten erneut aufgegriffen werden, wenn sich Geschäftsziele, Strategien oder Anforderungen von Endbenutzern ändern.

## Implementierungsschritte

- **Identifizieren von Stakeholdern:** Identifizieren und dokumentieren Sie wichtige Geschäfts-Stakeholder, einschließlich Entwicklungs- und Betriebsteams.
- **Definition von Zielen:** Arbeiten Sie mit diesen Stakeholdern zusammen, um die Ziele Ihrer Workload zu definieren und zu dokumentieren. Berücksichtigen Sie die kritischen Leistungsaspekte Ihrer Workloads, wie Durchsatz, Reaktionszeit und Kosten, sowie Geschäftsziele wie die Benutzerzufriedenheit.
- **Ansehen von in der Branche bewährter Methoden:** Sehen Sie sich in der Branche bewährte Methoden an, um relevante KPIs zu identifizieren, die auf Ihre Workload-Ziele abgestimmt sind.
- **Identifizieren von Metriken:** Identifizieren Sie Metriken, die auf Ihre Workload-Ziele abgestimmt sind und Ihnen helfen können, Leistung und Geschäftsziele zu messen. Richten Sie KPIs basierend auf diesen Metriken ein. Beispiele für Metriken sind Messungen wie die durchschnittliche Reaktionszeit oder die Anzahl gleichzeitiger Benutzer.
- **Definition und Dokumentation von KPIs:** Verwenden Sie in der Branche bewährte Methoden und Ihre Workload-Ziele, um Ziele für Ihren Workload-KPI festzulegen. Verwenden Sie diese Informationen, um KPI-Schwellenwerte für Schweregrad oder Alarmebene festzulegen. Identifizieren und dokumentieren Sie das Risiko und die Auswirkungen, wenn ein KPI nicht erreicht wird.
- **Implementieren der Überwachung:** Verwenden Sie Überwachungstools wie [Amazon CloudWatch](#) oder [AWS Config](#), um Metriken zu sammeln und KPIs zu messen.
- **Visuelle Kommunikation von KPIs:** Verwenden Sie Dashboard-Tools wie [Amazon QuickSight](#), um KPIs zu visualisieren und mit den Stakeholdern zu kommunizieren.
- **Analysieren und Optimieren:** Überprüfen und analysieren Sie regelmäßig die KPIs, um Bereiche Ihres Workloads zu identifizieren, die verbessert werden müssen. Arbeiten Sie mit den Stakeholdern zusammen, um diese Verbesserungen umzusetzen.

- Wiederaufgreifen und verfeinern: Überprüfen Sie regelmäßig die Metriken und KPIs, um ihre Effektivität zu bewerten, insbesondere wenn sich die Geschäftsziele oder die Workload-Leistung ändern.

## Ressourcen

### Zugehörige Dokumente:

- [CloudWatch-Dokumentation](#)
- [Überwachung, Protokollierung und Leistung von AWS Partners](#)
- [AWS-Tools zur Beobachtbarkeit](#)
- [The Importance of Key Performance Indicators \(KPIs\) for Large-Scale Cloud Migrations](#)
- [How to track your cost optimization KPIs with the KPI Dashboard](#)
- [X-Ray-Dokumentation](#)
- [Verwenden von Amazon-CloudWatch-Dashboards](#)
- [Quick-KPIs](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Optimieren der Kosten und Leistung sowie Verfolgen der Fortschritte bei der Schadensbegrenzung](#)
- [AWS re:Invent 2023 – Manage resource lifecycle events at scale with AWS Health](#)
- [AWS re:Invent 2023 – Performance & efficiency at Pinterest: Optimizing the latest instances](#)
- [AWS re:Invent 2022 – AWS-Optimierung: Umsetzbare Schritte für sofortige Ergebnisse](#)
- [AWS re:Invent 2023 – Aufbau einer effektiven Beobachtbarkeitsstrategie](#)
- [AWS Summit SF 2022 – Full-stack observability and application monitoring with AWS](#)
- [AWS re:Invent 2023 – Scaling on AWS for the first 10 million users](#)
- [AWS re:Invent 2022 – How Amazon uses better metrics for improved website performance](#)
- [Creating an Effective Metrics Strategy for Your Business | AWS Events](#)

### Zugehörige Beispiele:

- [Erstellen eines Dashboards mit Quick](#)

## PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren

Ermitteln Sie die Bereiche, in denen sich durch Steigern der Workload-Leistung positive Auswirkungen auf die Effizienz oder den Kundenkomfort realisieren lassen. Beispiel: Eine Website mit zahlreichen Kundeninteraktionen kann von der Nutzung von Edge-Services profitieren, indem Inhalte näher bei den Kunden bereitgestellt werden.

Typische Anti-Muster:

- Sie gehen davon aus, dass standardmäßige Datenverarbeitungsmetriken wie CPU-Auslastung oder Arbeitsspeicherdruck ausreichen, um Leistungsprobleme zu erfassen.
- Sie verwenden nur die Standardmetriken, die von der Überwachungssoftware Ihrer Wahl aufgezeichnet wurden.
- Sie überprüfen Metriken nur dann, wenn ein Problem vorliegt.

Vorteile der Nutzung dieser bewährten Methode: Das eingehende Verständnis kritischer Bereiche hilft Workload-Eigentümern dabei, KPIs zu überwachen und Verbesserungen mit größeren Auswirkungen zu priorisieren.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Hoch

### Implementierungsleitfaden

Richten Sie durchgehende Nachverfolgung ein, um Datenverkehrsmuster, Latenz und kritische Leistungsbereiche zu identifizieren. Überwachen Sie Ihre Datenzugriffsmuster auf langsame Abfragen oder schlecht fragmentierte und partitionierte Daten. Identifizieren Sie problematische Workload-Bereiche mithilfe von Lasttests oder -überwachung.

Erhöhen Sie die Leistungseffizienz durch eingehendes Verständnis Ihrer Architektur, der Datenverkehrs- und der Datenzugriffsmuster und identifizieren Sie Ihre Latenz- und Verarbeitungszeiten. Identifizieren Sie potenzielle Engpässe, die sich bei zunehmenden Workloads auf den Kundenkomfort auswirken könnten. Nachdem Sie diese Bereiche untersucht haben, sollten Sie prüfen, welche Lösung Sie nutzen können, um diese Leistungsprobleme zu beseitigen.

## Implementierungsschritte

- Richten Sie durchgehende Überwachung ein, um alle Workload-Komponenten und -Metriken zu erfassen. Hier finden Sie Beispiele für Überwachungslösungen in AWS.

Service	Aktion
<a href="#">Amazon CloudWatch Real-User Monitoring (RUM)</a>	Zum Erfassen von Metriken zur Anwendungleistung aus realen clientseitigen und Frontend-Sitzungen.
<a href="#">AWS X-Ray</a>	Zum Verfolgen des Datenverkehrs durch die Anwendungsebenen und zum Identifizieren der Latenz zwischen Komponenten und Abhängigkeiten. Verwenden Sie X-Ray-Service-Zuordnungen, um Beziehungen und Latenz zwischen Workload-Komponenten zu erkennen.
<a href="#">Amazon Relational Database Service – Performance Insights</a>	Zum Anzeigen von Metriken zur Datenbankleistung und zum Identifizieren von Möglichkeiten zur Leistungsverbesserung.
<a href="#">Amazon RDS Enhanced Monitoring</a>	Zum Anzeigen von Datenbank-BS-Leistungsmetriken.
<a href="#">Amazon DevOps Guru</a>	Zum Erkennen ungewöhnlicher Betriebsmuster, damit Sie betriebliche Probleme identifizieren können, bevor sie sich auf Ihre Kunden auswirken.

- Führen Sie Tests durch, um Metriken zu generieren sowie Datenverkehrsmuster, Engpässe und kritische Leistungsbereiche zu identifizieren. Hier finden Sie einige Beispiele zum Durchführen von Tests:
  - Richten Sie [CloudWatch Synthetic Canaries](#) zur programmgesteuerten Nachahmung browserbasierter Benutzeraktivitäten mit Linux-Cron-Aufträgen oder Rate-Ausdrücken und zum Erhalt konsistenter Metriken in Zeitverlauf ein.

- Verwenden Sie die Lösung für [verteilte Lasttests auf AWS](#), um Spitzendatenverkehr zu generieren oder Workloads mit der erwarteten Wachstumsrate zu testen.
- Evaluieren Sie die Metriken und die Telemetriedaten, um Ihre kritischen Leistungsbereiche zu identifizieren. Prüfen Sie diese Bereiche zusammen mit Ihrem Team und besprechen Sie Überwachung und Lösung zur Vermeidung von Engpässen.
- Experimentieren Sie mit Leistungsverbesserungen und messen Sie diese Änderungen anhand von Daten. Beispielsweise können Sie [CloudWatch Evidently](#) verwenden, um neue Verbesserungen und Leistungsauswirkungen auf Ihrer Workload zu testen.

## Ressourcen

### Zugehörige Dokumente:

- [What's new in AWS Observability at re:Invent 2023](#)
- [Amazon Builders' Library](#)
- [X-Ray-Dokumentation](#)
- [Amazon CloudWatch RUM](#)
- [Amazon DevOps Guru](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – \[LAUNCH\] Application monitoring for modern workloads](#)
- [AWS re:Invent 2023 – Implementierung der Anwendungsbeobachtbarkeit](#)
- [AWS re:Invent 2023 – Aufbau einer effektiven Beobachtbarkeitsstrategie](#)
- [AWS Summit SF 2022 – Full-stack observability and application monitoring with AWS](#)
- [AWS re:Invent 2022 – AWS-Optimierung: Umsetzbare Schritte für sofortige Ergebnisse](#)
- [AWS re:Invent 2022 – Die Amazon Builders' Library: 25 Jahre operative Exzellenz von Amazon](#)
- [AWS re:Invent 2022 – How Amazon uses better metrics for improved website performance](#)
- [Visual Monitoring of Applications with Amazon CloudWatch Synthetics](#)

### Zugehörige Beispiele:

- [Messen der Seitenladezeit mit Amazon CloudWatch Synthetics](#)

- [Webclient von Amazon CloudWatch RUM](#)
- [X-Ray-SDK für Python](#)
- [Distributed Load Testing auf AWS](#)

## PERF05-BP03 Definieren eines Prozesses zum Verbessern der Workload-Leistung

Definieren Sie einen Prozess, mit dem sich neu verfügbare Services, Designmuster, Ressourcentypen und Konfigurationen bewerten lassen. Führen Sie beispielsweise vorhandene Leistungstests für neue Instance-Angebote durch, um zu ermitteln, welche Verbesserungen sich für Ihre Workload ergeben.

Typische Anti-Muster:

- Sie gehen davon aus, dass Ihre aktuelle Architektur statisch ist und im Laufe der Zeit nicht aktualisiert wird.
- Sie führen im Laufe der Zeit Änderungen an der Architektur ein, ohne sie begründen.

Vorteile der Nutzung dieser bewährten Methode: Durch einen definierten Prozess zum Ändern der Architektur erhalten Sie die Möglichkeit, die gesammelten Daten langfristig in die Gestaltung Ihrer Workload einfließen zu lassen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

### Implementierungsleitfaden

Für Ihre Workload gibt es einige wesentliche Einschränkungen. Dokumentieren Sie diese, damit Sie besser einschätzen können, durch welche Art von Innovation die Leistung Ihrer Workload gesteigert werden könnte. Ziehen Sie diese Informationen heran, wenn Sie von neuen verfügbaren Services oder Technologien erfahren, um Möglichkeiten zur Beseitigung von Einschränkungen oder Engpässen zu identifizieren.

Identifizieren Sie wesentliche Leistungseinschränkungen für Ihre Workload. Dokumentieren Sie die Leistungseinschränkungen Ihrer Workload, damit Sie besser einschätzen können, durch welche Art von Innovation die Leistung Ihrer Workload ggf. gesteigert werden kann.

## Implementierungsschritte

- Identifizieren von KPIs: Identifizieren Sie Ihre Workload-Leistungs-KPIs, wie unter [PERF05-BP01 Festlegen wichtiger Leistungskennzahlen \(KPIs\) zum Messen des Zustands und der Leistung der Workload](#) zur Ermittlung Ihrer Workload-Baseline beschrieben.
- Implementieren der Überwachung: Verwenden Sie [AWS-Beobachtbarkeitstools](#), um Leistungskennzahlen zu erfassen und KPIs zu messen.
- Durchführen einer Analyse: Führen Sie eine eingehende Analyse durch, um die Bereiche (wie Konfiguration und Anwendungscode) in Ihrer Workload zu identifizieren, die leistungsschwach sind, wie beschrieben unter [PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren](#). Verwenden Sie Analyse- und Leistungstools, um die Strategie zur Leistungsverbesserung zu identifizieren.
- Validieren von Verbesserungen: Verwenden Sie Sandbox- oder Vorproduktionsumgebungen, um die Effektivität von Verbesserungsstrategien zu überprüfen.
- Implementieren von Änderungen: Implementieren Sie die Änderungen in der Produktion und überwachen Sie kontinuierlich die Leistung der Workload. Dokumentieren Sie die Verbesserungen und teilen Sie die Änderungen den Stakeholdern mit.
- Wiederaufgreifen und verfeinern: Überprüfen Sie regelmäßig Ihren Prozess zur Leistungsverbesserung, um Bereiche mit Verbesserungspotenzial zu identifizieren.

## Ressourcen

### Zugehörige Dokumente:

- [AWS Blog](#)
- [Neuerungen bei AWS](#)
- [AWS Skill Builder](#)

### Zugehörige Videos:

- [AWS re:Invent 2022 – Bereitstellung nachhaltiger, leistungsstarker Architekturen](#)
- [AWS re:Invent 2023 – Optimieren der Kosten und Leistung sowie Verfolgen der Fortschritte bei der Schadensbegrenzung](#)
- [AWS re:Invent 2022 – AWS-Optimierung: Umsetzbare Schritte für sofortige Ergebnisse](#)

- [AWS re:Invent 2022 – Optimize your AWS workloads with best-practice guidance](#)

Zugehörige Beispiele:

- [AWS GitHub](#)

## PERF05-BP04 Belastungstest Ihr Workload

Führen Sie für die Workload Lasttests durch, um sicherzustellen, dass sie die Produktionslast bewältigen kann, und identifizieren Sie Leistungsengpässe.

Typische Anti-Muster:

- Sie führen Lasttests für einzelne Teile der Workload durch, aber nicht für die gesamte Workload.
- Sie führen Lasttests in einer Infrastruktur durch, die sich von Ihrer Produktionsumgebung unterscheidet.
- Sie führen Lasttests nur für die erwartete Last durch und nicht für noch größere Lasten, um mögliche künftige Probleme besser vorherzusehen.
- Sie führen Belastungstests durch, ohne die [Amazon EC2 Testing Policy zu lesen](#) und ein Formular zur Einreichung von simulierten Ereignissen einzureichen. Dies führt dazu, dass Ihr Test nicht ausgeführt werden kann, da er wie ein denial-of-service Ereignis aussieht.

Vorteile der Nutzung dieser bewährten Methode: Die Messung der Leistung im Rahmen eines Lasttests gibt Aufschluss darüber, wo bei zunehmender Last mit Auswirkungen zu rechnen ist. Auf diese Weise können Sie erforderliche Änderungen vorhersehen, bevor sie sich auf Ihre Workload auswirken.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Niedrig

## Implementierungsleitfaden

Lasttests in der Cloud sind ein Prozess zur Messung der Leistung einer Cloud-Workload unter realistischen Bedingungen mit erwarteter Benutzerlast. Dieser Prozess beinhaltet die Bereitstellung einer produktionsähnlichen Cloud-Umgebung, die Verwendung von Lasttest-Tools zur Lastgenerierung und die Analyse von Metriken, um die Fähigkeit Ihrer Workload zu bewerten, mit einer realistischen Last umzugehen. Verwenden Sie für Lasttests synthetische oder bereinigte Produktionsdaten und entfernen Sie sensible oder personenbezogene Informationen. Führen Sie

automatisch Lasttests als Teil Ihrer Bereitstellungs pipeline durch und vergleichen Sie die Ergebnisse mit vordefinierten Grenzwerten KPIs und Schwellenwerten. Dieser Prozess hilft Ihnen dabei, die erforderliche Leistung weiterhin zu erreichen.

## Implementierungsschritte

- **Testziele definieren:** Identifizieren Sie die Leistungsaspekte Ihrer Workload, die Sie bewerten möchten, wie Durchsatz und Reaktionszeit.
- **Testtool auswählen:** Wählen und konfigurieren Sie das Lasttest-Tool, das zu Ihrer Workload passt.
- **Umgebung einrichten:** Richten Sie die Testumgebung auf der Grundlage Ihrer Produktionsumgebung ein. Sie können AWS Services verwenden, um Produktionsumgebungen auszuführen und Ihre Architektur zu testen.
- **Implementieren Sie Überwachung:** Verwenden Sie Überwachungstools wie [Amazon CloudWatch](#), um Kennzahlen für alle Ressourcen in Ihrer Architektur zu sammeln. Sie können auch benutzerdefinierte Metriken erfassen und veröffentlichen.
- **Szenarien definieren:** Definieren Sie die Szenarien und Parameter der Lasttests (wie Testdauer und Anzahl der Benutzer).
- **Belastungstests durchführen:** Führen Sie Testszenarien in großem Maßstab durch. Nutzen Sie das AWS Cloud , um Ihren Workload zu testen und herauszufinden, wo er nicht skaliert werden kann oder ob er nicht linear skaliert wird. Nutzen Sie beispielsweise Spot Instances, um kostengünstig Lasten zu erzeugen und Engpässe zu identifizieren, bevor diese in der Produktionsumgebung auftreten.
- **Testergebnisse analysieren:** Analysieren Sie die Ergebnisse, um Leistungsengpässe und verbesserungswürdige Bereiche zu identifizieren.
- **Erkenntnisse dokumentieren und teilen:** Dokumentieren Sie Erkenntnisse und Empfehlungen und erstellen Sie Berichte darüber. Teilen Sie diese Informationen mit Stakeholdern, um ihnen zu helfen, fundierte Entscheidungen über Strategien zur Leistungsoptimierung zu treffen.
- **Kontinuierliche Iteration:** Lasttests sollten in regelmäßigen Abständen durchgeführt werden, insbesondere nach einer Systemänderung oder einem Systemupdate.

## Ressourcen

Zugehörige Dokumente:

- [Amazon CloudWatch RUM](#)

- [Amazon CloudWatch Synthetics](#)
- [Testen verteilter Lasten auf AWS](#)

Zugehörige Videos:

- [AWS Summit ANZ 2023: Mit AWS Distributed Load Testing mit Zuversicht beschleunigen](#)
- [AWS re:Invent 2022 — Weitere Skalierung AWS für Ihre ersten 10 Millionen Benutzer](#)
- [Problemlösung mit AWS Lösungen: Testen verteilter Lasten](#)
- [AWS re:Invent 2021 — Optimieren Sie Anwendungen mithilfe von Erkenntnissen für Endbenutzer mit Amazon CloudWatch RUM](#)
- [Demo von Amazon CloudWatch Synthetics](#)

Zugehörige Beispiele:

- [Verteilte Belastungstests auf AWS](#)

## PERF05-BP05 Verwenden von Automatisierung zur proaktiven Behebung leistungsbezogener Probleme

Verwenden Sie wichtige Leistungskennzahlen (KPIs) in Kombination mit Überwachungs- und Warnsystemen, um eine proaktive Behandlung leistungsbezogener Probleme zu ermöglichen.

Typische Anti-Muster:

- Sie geben dem Betriebspersonal nur die Möglichkeit, betriebliche Änderungen an der Workload vorzunehmen.
- Sie lassen alle Alarme ohne proaktive Behebung zum Operations-Team filtern.

Vorteile der Nutzung dieser bewährten Methode: Die proaktive Behebung von Alarmaktionen ermöglicht es dem Support-Personal, sich auf die Elemente zu konzentrieren, die nicht automatisch umsetzbar sind. Dies hilft dem Betriebspersonal, alle Alarme zu bewältigen, ohne überfordert zu werden, und sich stattdessen auf die kritischen Alarme zu konzentrieren.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Niedrig

## Implementierungsleitfaden

Verwenden Sie Alarme, um automatisierte Aktionen auszulösen und auf diese Weise Probleme nach Möglichkeit zu beheben. Leiten Sie den Alarm an die Personen weiter, die die richtigen Maßnahmen einleiten können, falls keine automatisierte Reaktion möglich ist. Beispielsweise können Sie ein System nutzen, das erwartete Werte wichtiger Leistungskennzahlen (KPIs) prognostiziert und bei Überschreiten bestimmter Schwellenwerte einen Alarm ausgibt. Denkbar ist auch ein Tool, das Bereitstellungen automatisch anhält oder zurücksetzt, wenn sich KPIs außerhalb der erwarteten Werte befinden.

Implementieren Sie Prozesse, die Ihnen Einblick in die Leistung gewähren, während Ihre Workload ausgeführt wird. Entwickeln Sie Dashboards für die Überwachung und legen Sie Leistungsnormen in Form von Grundwerten fest, um zu bestimmen, ob die Workload optimal funktioniert.

### Implementierungsschritte

- Identifizierung eines Fehlerbehebungs-Workflows: Identifizieren und verstehen Sie das Leistungsproblem, das automatisch behoben werden kann. Verwenden Sie Überwachungslösungen von AWS wie [Amazon CloudWatch](#) oder AWS X-Ray, damit Sie die Ursache des Problems besser verstehen.
- Definition des Automatisierungsprozesses: Erstellen Sie einen schrittweisen Prozess zur Behebung des Problems, mit dem das Problem automatisch behoben werden kann.
- Konfiguration des Initiierungsereignisses: Konfigurieren Sie das Ereignis so, dass der Prozess zur Mängelbeseitigung automatisch eingeleitet wird. Sie können beispielsweise einen Auslöser definieren, der eine Instance automatisch neu startet, wenn sie einen bestimmten Schwellenwert für die CPU-Auslastung erreicht.
- Automatisierung der Mängelbeseitigung: Nutzen Sie AWS-Services und -Technologien, um den Prozess zur Mängelbeseitigung zu automatisieren. Beispielsweise bietet [AWS Systems Manager Automation](#) eine sichere und skalierbare Möglichkeit, den Prozess zur Mängelbeseitigung zu automatisieren. Achten Sie darauf, die Selbstheilungslogik zu verwenden, um Änderungen rückgängig zu machen, wenn das Problem nicht gelöst wurde.
- Testen des Workflows: Testen Sie den automatisierten Prozess zur Mängelbeseitigung in einer Vorproduktionsumgebung.
- Implementieren des Workflows: Implementieren Sie die automatisierte Mängelbeseitigung in der Produktionsumgebung.
- Entwicklung eines Playbooks: Entwickeln und dokumentieren Sie ein Playbook, in dem die Schritte für den Mängelbeseitigungsplan beschrieben werden, einschließlich der Initiierungsereignisse, der

Mängelbeseitigungslogik und der ergriffenen Maßnahmen. Stellen Sie sicher, dass alle Stakeholder entsprechend geschult werden, damit sie effektiv auf automatisierte Mängelbeseitigungsereignisse reagieren können.

- Überprüfen und verfeinern: Bewerten Sie regelmäßig die Effektivität des automatisierten Mängelbeseitigungsworkflows. Passen Sie bei Bedarf die Initiierungsereignisse und die Mängelbeseitigungslogik an.

## Ressourcen

Zugehörige Dokumente:

- [CloudWatch-Dokumentation](#)
- [Überwachung, Protokollierung und Leistung von AWS Partner Network-Partnern](#)
- [X-Ray-Dokumentation](#)
- [Using Alarms and Alarm Actions in CloudWatch](#)
- [Build a Cloud Automation Practice for Operational Excellence: Best Practices from AWS Managed Services](#)
- [Automate your Amazon Redshift performance tuning with automatic table optimization](#)

Zugehörige Videos:

- [AWS re:Invent 2023 – Strategies for automated scaling, remediation, and smart self-healing](#)
- [AWS re:Invent 2023 – \[LAUNCH\] Application monitoring for modern workloads](#)
- [AWS re:Invent 2023 – Implementierung der Anwendungsbeobachtbarkeit](#)
- [AWS re:Invent 2021 – Intelligently automating cloud operations](#)
- [AWS re:Invent 2022 – Setting up controls at scale in your AWS environment](#)
- [AWS re:Invent 2022 – Automating patch management and compliance using AWS](#)
- [AWS re:Invent 2022 – How Amazon uses better metrics for improved website performance](#)
- [AWS re:Invent 2023 – Take a load off: Diagnose & resolve performance issues with Amazon RDS](#)
- [AWS re:Invent 2021 – {New Launch} Automatically detect and resolve issues with Amazon DevOps Guru](#)
- [AWS re:Invent 2023 – Centralize your operations](#)

Zugehörige Beispiele:

- [CloudWatch Logs – Konfigurieren von Alarmen](#)

## PERF05-BP06 Behalten Sie Ihren Workload und Ihre Services up-to-date

Bleiben Sie up-to-date auf der Suche nach neuen Cloud-Diensten und -Funktionen, um effiziente Funktionen einzuführen, Probleme zu beheben und die allgemeine Leistungseffizienz Ihres Workloads zu verbessern.

Typische Anti-Muster:

- Sie gehen davon aus, dass Ihre aktuelle Architektur statisch ist und im Laufe der Zeit nicht aktualisiert wird.
- Sie haben keine Systeme oder regelmäßigen Besprechungen zur Prüfung, ob aktualisierte Software und Pakete mit Ihrer Workload kompatibel sind.

Vorteile der Einführung dieser bewährten Methode: Durch die Einrichtung eines Prozesses up-to-date zur Beibehaltung neuer Services und Angebote können Sie neue Funktionen und Fähigkeiten einführen, Probleme lösen und die Workload-Leistung verbessern.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Niedrig

### Implementierungsleitfaden

Evaluieren Sie Möglichkeiten zur Verbesserung der Leistung, wenn neue Services, Entwurfsmuster und Produktfunktionen verfügbar sind. Ermitteln Sie anhand von Bewertungen, internen Diskussionen oder externen Analysen, wie sich diese neuen Optionen positiv auf die Leistung oder Effizienz der Workload auswirken können. Definieren Sie einen Prozess zum Bewerten von Updates, neuen Features und Services, die für Ihre Workload relevant sind. Erstellen Sie beispielsweise Machbarkeitsstudien, die auf neuen Technologien aufbauen, oder beraten Sie sich mit einer internen Gruppe. Führen Sie beim Ausprobieren neuer Ideen oder Services Leistungstests durch, um die Auswirkungen auf die Leistung der Workload zu messen.

## Implementierungsschritte

- Inventarisierung Ihrer Workload: Inventarisieren Sie Ihre Workload-Software und -Architektur und identifizieren Sie Komponenten, die aktualisiert werden müssen.
- Identifizierung von Quellen für Updates: Identifizieren Sie Quellen für Neuigkeiten und Updates im Zusammenhang mit Ihren Workload-Komponenten. Sie können beispielsweise den [AWS Blog What's New at](#) abonnieren, um die Produkte zu finden, die zu Ihrer Workload-Komponente passen. Sie können den RSS Feed abonnieren oder Ihre [E-Mail-Abonnements](#) verwalten.
- Definition eines Aktualisierungszeitplans: Definieren Sie einen Zeitplan zur Evaluierung neuer Services und Features für Ihre Workload.
  - Sie können [AWS Systems Manager Inventory](#) verwenden, um Betriebssystem- (OS), Anwendungs- und Instance-Metadaten von Ihren EC2 Amazon-Instances zu sammeln und so schnell zu verstehen, auf welchen Instances die Software und die Konfigurationen ausgeführt werden, die gemäß Ihrer Softwarerichtlinie erforderlich sind, und welche Instances aktualisiert werden müssen.
- Bewertung der neuen Aktualisierung: Erfahren Sie, wie die Komponenten Ihrer Workload aktualisiert werden. Nutzen Sie die Agilität in der Cloud, um schnell zu testen, wie neue Features Ihre Workload verbessern und so die Leistungseffizienz steigern können.
- Verwendung von Automatisierung: Verwenden Sie Automatisierung für den Aktualisierungsvorgang, um den Aufwand für die Bereitstellung neuer Features zu reduzieren und Fehler zu begrenzen, die durch manuelle Prozesse verursacht werden.
  - Sie können [CI/CD](#) verwenden AMIs, um Container-Images und andere Artefakte im Zusammenhang mit Ihrer Cloud-Anwendung automatisch zu aktualisieren.
  - Sie können Tools wie den [AWS Systems Manager Patch Manager](#) verwenden, um den Systemaktualisierungsprozess zu automatisieren und die Aktivitäten mit [AWS Systems Manager Maintenance Windows](#) zu planen.
- Dokumentation des Prozesses: Dokumentieren Sie Ihren Prozess zur Evakuierung von Aktualisierungen und neuen Services. Geben Sie Ihren Eigentümern ausreichend Zeit und Raum zum Forschen, Testen, Experimentieren und zur Validierung von Aktualisierungen und neuen Services. Schauen Sie sich die dokumentierten Geschäftsanforderungen KPIs an und finden Sie heraus, welche Aktualisierung sich positiv auf Ihr Unternehmen auswirken wird.

## Ressourcen

Zugehörige Dokumente:

- [AWS -Blog](#)
- [Was ist neu bei AWS](#)
- [Implementieren von up-to-date Images mit automatisierten EC2 Image Builder Builder-Pipelines](#)

Zugehörige Videos:

- [AWS re:INFORCE 2022 — Automatisierung von Patch-Management und Compliance mithilfe AWS](#)
- [All Things Patch: | Veranstaltungen AWS Systems ManagerAWS](#)

Zugehörige Beispiele:

- [Bestands- und Patch-Verwaltung](#)
- [Workshop zur Beobachtbarkeit](#)

## PERF05-BP07 Regelmäßiges Überprüfen von Metriken

Überprüfen Sie im Rahmen der routinemäßigen Wartungsmaßnahme oder als Reaktion auf Ereignisse oder Vorfälle, welche Metriken erfasst werden. Ermitteln Sie anhand dieser Überprüfung, welche Metriken für die Behebung von Problemen wesentlich waren und welche zusätzlichen Metriken, sofern nachverfolgt, helfen könnten, Probleme zu identifizieren, zu beheben oder zu verhindern.

Typische Anti-Muster:

- Sie lassen zu, dass Metriken für einen längeren Zeitraum im Alarmstatus bleiben.
- Sie erstellen Alarme, die von einem Automatisierungssystem nicht umsetzbar sind.

Vorteile der Nutzung dieser bewährten Methode: Überprüfen Sie kontinuierlich Metriken, die erfasst werden, um zu bestätigen, dass sie Probleme ordnungsgemäß identifizieren, beheben oder verhindern. Metriken können auch veralten, wenn sie für einen längeren Zeitraum im Alarmstatus bleiben.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: Mittel

## Implementierungsleitfaden

Verbessern Sie kontinuierlich die Erfassung und Überwachung von Metriken. Bewerten Sie beim Reagieren auf Vorfälle oder Ereignisse diejenigen Metriken, die hilfreich für die Behebung des Problems waren, und überlegen Sie, welche derzeit noch nicht verfolgten Metriken förderlich sein könnten. Verbessern Sie auf diese Weise die Qualität der erfassten Metriken, damit Sie zukünftige Probleme verhindern oder schneller beheben können.

Bewerten Sie beim Reagieren auf Vorfälle oder Ereignisse diejenigen Metriken, die hilfreich für die Behebung des Problems waren, und überlegen Sie, welche derzeit noch nicht verfolgten Metriken förderlich sein könnten. Verbessern Sie auf diese Weise die Qualität der erfassten Metriken, damit Sie zukünftige Probleme verhindern oder schneller beheben können.

### Implementierungsschritte

- **Metriken definieren:** Definieren Sie wichtige Leistungsmetriken zur Überwachung, die auf Ihr Workload-Ziel abgestimmt sind, einschließlich Metriken wie Reaktionszeit und Ressourcenauslastung.
- **Ausgangswert festlegen:** Legen Sie für jede Metrik einen Ausgangswert und einen Zielwert fest. Der Ausgangswert sollte Referenzpunkte zur Identifizierung von Abweichungen oder Anomalien enthalten.
- **Takt festlegen:** Legen Sie einen Takt zur Überprüfung wichtiger Kennzahlen fest (z. B. wöchentlich oder monatlich).
- **Leistungsprobleme identifizieren:** Beurteilen Sie bei jeder Überprüfung Trends und Abweichungen von den Ausgangswerten. Suchen Sie nach Leistungsengpässen oder Anomalien. Führen Sie bei identifizierten Problemen eine eingehende Ursachenanalyse durch, um den Hauptgrund für das Problem zu ermitteln.
- **Korrekturmaßnahmen identifizieren:** Identifizieren Sie Korrekturmaßnahmen mithilfe Ihrer Analysen. Dies kann die Parameteroptimierung, das Beheben von Fehlern und das Skalieren von Ressourcen beinhalten.
- **Ergebnisse dokumentieren:** Dokumentieren Sie Ihre Erkenntnisse, einschließlich identifizierter Probleme, Ursachen und Korrekturmaßnahmen.
- **Iterieren und verbessern:** Beurteilen und verbessern Sie kontinuierlich den Prozess zur Überprüfung der Metriken. Nutzen Sie die Erkenntnisse aus der vorherigen Überprüfung, um den Prozess im Laufe der Zeit zu verbessern.

# Ressourcen

## Zugehörige Dokumente:

- [CloudWatch-Dokumentation](#)
- [Sammeln von Metriken und Protokollen von Amazon-EC2-Instances und On-Premises-Servern mit dem CloudWatch-Agenten](#)
- [Query your metrics with CloudWatch Metrics Insights](#)
- [Überwachung, Protokollierung und Leistung von AWS Partner Network-Partnern](#)
- [X-Ray-Dokumentation](#)

## Zugehörige Videos:

- [AWS re:Invent 2022 - Setting up controls at scale in your AWS environment](#)
- [AWS re:Invent 2022 - How Amazon uses better metrics for improved website performance](#)
- [AWS re:Invent 2023 - Building an effective observability strategy](#)
- [AWS Summit SF 2022 - Full-stack observability and application monitoring with AWS](#)
- [AWS re:Invent 2023 - Take a load off: Diagnose & resolve performance issues with Amazon RDS](#)

## Zugehörige Beispiele:

- [Erstellen eines Dashboards mit Quick](#)
- [CloudWatch-Dashboards](#)

## Schlussfolgerung

Um eine hohe Leistungseffizienz zu erzielen und aufrechtzuerhalten, ist ein datengesteuerter Ansatz erforderlich. Zur Leistungsoptimierung empfiehlt sich daher, entsprechende Zugriffsmuster und Kompromisse zu erwägen. Mithilfe von Benchmarks und Lasttests können Sie die passenden Ressourcentypen und Konfigurationen ermitteln. Indem Sie Ihre Infrastruktur als Code verwalten, können Sie Ihre Architektur schnell und sicher ausbauen. Daten helfen Ihnen dabei, fundierte Entscheidungen zu treffen. Durch eine Kombination aus aktiver und passiver Überwachung stellen Sie für Ihre Architektur eine dauerhaft konsistente Leistung sicher.

AWS ist bestrebt, Sie beim Aufbau von Architekturen zu unterstützen, die effizient arbeiten und gleichzeitig einen geschäftlichen Nutzen bieten. Verwenden Sie die in diesem Whitepaper beschriebenen Tools und Verfahren, damit Ihre Arbeit erfolgreich ist.

# Mitwirkende

Folgende Personen und Organisationen haben zu diesem Dokument beigetragen:

- Sam Mokhtari, Senior Efficiency Lead Solutions Architect, Amazon Web Services
- Josh Hart, Solutions Architect, Amazon Web Services
- Richard Trabing, Solutions Architect, Amazon Web Services
- Brett Looney, Principal Solutions Architect, Amazon Web Services
- Nina Vogl, Principal Solutions Architect, Amazon Web Services
- Eric Pullen, Solutions Architect, Amazon Web Services
- Julien Lépine, Specialist SA Manager, Amazon Web Services
- Ronnen Slasky, Solutions Architect, Amazon Web Services

## Weitere Informationen

Weitere Informationen finden Sie in den folgenden Quellen:

- [AWS Well-Architected Framework](#)
- [AWS -Architekturzentrum](#)

# Dokumentversionen

Abonnieren Sie den RSS-Feed, um über Aktualisierungen des Whitepapers benachrichtigt zu werden.

Änderung	Beschreibung	Datum
<a href="#">Kleinere Aktualisierung bewährter Methoden</a>	PERF03-BP04 wurde mit neuen Serviceempfehlungen aktualisiert.	6. November 2024
<a href="#">Leitfäden zu bewährten Methoden aktualisiert</a>	Mehrere kleine Aktualisierungen in der gesamten Säule.	27. Juni 2024
<a href="#">Umfangreiche Aktualisierung und Umstrukturierung</a>	Säule wurde umstrukturiert und umfasst nun fünf Bereiche mit bewährten Methoden (statt acht). Der Inhalt wurde in die fünf Bereiche zusammengefasst und aktualisiert.  Neue Bereiche von bewährten Methoden sind <a href="#">Architektur</a> , <a href="#">Auswahl</a> , <a href="#">Datenverarbeitung und Hardware</a> , <a href="#">Datenmanagement</a> , <a href="#">Netzwerke und Inhaltsbereitstellung</a> sowie <a href="#">Prozesse und Kultur</a> .	3. Oktober 2023
<a href="#">Kleines Update</a>	Entfernung von nicht-inklusive Sprache	13. April 2023
<a href="#">Aktualisierungen für das neue Framework</a>	Bewährte Methoden mit verbindlichen Anleitungen aktualisiert und neue bewährte Methoden hinzugefügt.	10. April 2023

<a href="#">Whitepaper aktualisiert</a>	Bewährte Methoden mit neuen Implementierungsanleitungen aktualisiert.	15. Dezember 2022
<a href="#">Whitepaper aktualisiert</a>	Weitere bewährte Methoden und Verbesserungspläne hinzugefügt.	20. Oktober 2022
<a href="#">Kleines Update</a>	Nicht inklusive Sprache entfernt.	22. April 2022
<a href="#">Kleinere Updates</a>	Links aktualisiert.	10. März 2021
<a href="#">Kleinere Updates</a>	Das AWS Lambda-Timeout wurde in 900 Sekunden geändert und der Name von Amazon Keyspaces (für Apache Cassandra) wurde korrigiert.	5. Oktober 2020
<a href="#">Kleines Update</a>	Fehlerhafter Link behoben.	15. Juli 2020
<a href="#">Aktualisierungen für das neue Framework</a>	Umfangreiche Überprüfung und Aktualisierung der Inhalte	8. Juli 2020
<a href="#">Whitepaper aktualisiert</a>	Kleine Grammatikkorrekturen	1. Juli 2018
<a href="#">Whitepaper aktualisiert</a>	Aktualisierung des Whitepapers mit Änderungen in AWS	1. November 2017
<a href="#">Erste Veröffentlichung</a>	Säule „Leistungseffizienz“ – AWS-Well-Architected-Framework veröffentlicht.	1. November 2016

# Hinweise

Kunden sind dafür verantwortlich, Ihre eigene unabhängige Bewertung der Informationen in diesem Dokument vorzunehmen. Dieses Dokument: (a) dient nur zu Informationszwecken, (b) stellt aktuelle AWS Produktangebote und Praktiken dar, die ohne vorherige Ankündigung geändert werden können, und (c) stellt keine Verpflichtungen oder Zusicherungen von AWS und seinen verbundenen Unternehmen, Lieferanten oder Lizenzgebern dar. AWS Produkte oder Dienstleistungen werden „wie sie sind“ ohne ausdrückliche oder stillschweigende Garantien, Zusicherungen oder Bedingungen jeglicher Art bereitgestellt. Die Verantwortlichkeiten und Verbindlichkeiten AWS gegenüber seinen Kunden werden durch AWS Vereinbarungen geregelt, und dieses Dokument ist weder Teil einer Vereinbarung zwischen AWS und seinen Kunden noch ändert es diese.

© 2023, Amazon Web Services, Inc. oder Tochterfirmen. Alle Rechte vorbehalten.

# AWS Glossar

Die neueste AWS Terminologie finden Sie im [AWS Glossar](#) in der AWS-Glossar Referenz.