

# Säule „Leistungseffizienz“



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

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

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

---

# Table of Contents

Überblick und Einführung .....	1
Überblick .....	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 der 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 von 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 den 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 Erfassen von Datenverarbeitungsmetriken .....	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 .....	36
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 .....	53
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF03-BP03 Erfassen und Aufzeichnen von Metriken zur Datenspeicherleistung .....	59
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7

PERF03-BP04 Implementieren von Strategien zur Verbesserung der Abfrageleistung im Datenspeicher .....	62
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF03-BP05 Implementieren von Datenzugriffsmustern, die Caching nutzen .....	65
Implementierungsleitfaden .....	6
Ressourcen .....	7
Networking und Bereitstellung von Inhalten .....	69
PERF04-BP01 Verstehen der Auswirkungen des Netzwerks auf die Leistung .....	70
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP02 Evaluieren verfügbarer Netzwerk-Features .....	73
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP03 Auswählen von entsprechend dedizierter Konnektivität oder VPN für Ihre Workload .....	80
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP04 Lastausgleich verwenden, um den Datenverkehr auf mehrere Ressourcen zu verteilen .....	83
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP05 Auswählen leistungsfördernder Netzwerkprotokolle .....	88
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP06 Auswählen des Workload-Standortes entsprechend den Netzwerkanforderungen .....	92
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF04-BP07 Optimieren der Netzwerkkonfiguration basierend auf Metriken .....	97
Implementierungsleitfaden .....	6
Ressourcen .....	7
Prozess und Kultur .....	103
PERF05-BP01 Festlegen wichtiger Leistungskennzahlen (KPIs) zum Messen des Zustands und der Leistung des Workloads .....	105
Implementierungsleitfaden .....	6

Implementierungsschritte .....	6
Ressourcen .....	7
PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren .....	108
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP03 Definieren eines Prozesses zum Verbessern der Workload-Leistung .....	111
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP04 Durchführen von Lasttests für den Workload .....	113
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP05 Verwenden von Automatisierung zur proaktiven Behebung leistungsbezogener Probleme .....	115
Implementierungsleitfaden .....	6
Ressourcen .....	7
PERF05-BP06 Konstantes Aktualisieren des Workloads und der Services .....	118
Implementierungsleitfaden .....	6
Implementierungsschritte .....	6
Ressourcen .....	7
PERF05-BP07 Regelmäßiges Überprüfen von Metriken .....	120
Implementierungsleitfaden .....	6
Ressourcen .....	7
Fazit .....	123
Mitwirkende .....	124
Weitere Informationen .....	125
Dokumentversionen .....	126
AWS Glossary .....	128

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

Veröffentlichungsdatum: 27. Juni 2024 ([Dokumentversionen](#))

## Überblick

Thema dieses Whitepapers ist die Säule „Leistungseffizienz“ des [AWS Well-Architected Framework](#). Der Zweck dieses Dokuments besteht darin, Leitlinien bereitzustellen, die Kunden dabei unterstützen, Cloud-Ressourcen effizient zu nutzen, um ihre Geschäftsanforderungen zu erfüllen und diese Effizienz aufrechtzuerhalten, wenn sich die Nachfrage ändert und sich die Technologien weiterentwickeln.

## 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:

- Operative Exzellenz
- 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 auf AWS, die im Laufe der Zeit effizient für eine nachhaltige Leistung sorgen. Die Leitlinien 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](#)
- [Networking und Bereitstellung von Inhalten](#)
- [Prozess und Kultur](#)

Dieses Dokument richtet sich an Nutzer in technologischen Rollen, z. B. CTOs (Chief Technology Officers), Architekten, Entwickler und Mitglieder von Operations-Teams. Sie erfahren darin mehr über die bewährten Methoden und Strategien von AWS zum Entwickeln einer leistungsfähigen Cloud-Architektur.

# Leistungseffizienz

Die Säule der Leistungseffizienz befasst sich mit der effizienten Nutzung von Computing-Ressourcen entsprechend den Anforderungen sowie der Aufrechterhaltung dieser Effizienz, wenn sich die Nachfrage ändert und sich die Technologien weiterentwickeln.

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. Es gibt Technologien, wie etwa die NoSQL-Datenbanken, das Transcodieren von Medien sowie Machine Learning, die spezielles Fachwissen erfordern. 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.
- **Globale Verteilung in wenigen Minuten:** Durch die Bereitstellung Ihres Workloads in mehreren AWS-Regionen auf der ganzen Welt können Sie Ihren Kunden geringere Latenz und eine bessere Erfahrung bei minimalen Kosten bieten.
- **Nutzung von serverlosen Architekturen:** Aufgrund der in der Cloud verwendeten serverlosen Architekturen müssen Sie für herkömmliche Rechenaktivitäten keine physischen Server mehr ausführen und verwalten. Serverlose 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 der mechanischen Präferenz: Nutzen Sie den für Ihre Anforderungen optimalen Technologieansatz. Berücksichtigen Sie bei der Auswahl der passenden Datenbank oder des Speichers für Ihren 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](#)
- [Networking und Bereitstellung von Inhalten](#)
- [Prozess und Kultur](#)

Wählen Sie einen datenbasierten Ansatz, um eine leistungsstarke Architektur zu entwickeln. Sammeln Sie zu allen Aspekten der Architektur Daten, angefangen vom allgemeinen Design bis hin zur Auswahl und Konfiguration der Ressourcentypen.

Durch regelmäßiges Überprüfen Ihrer Auswahl stellen Sie die bestmögliche Nutzung der sich fortlaufend weiterentwickelnden AWS Cloud sicher. 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 einen bestimmten Workload variiert und Lösungen bestehen häufig aus einer Kombination mehrerer Ansätze. Well-Architected-Workloads nutzen mehrere Lösungen und ermöglichen verschiedene Funktionen 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 lokalen 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 Berücksichtigen der Kosten bei architektonischen Entscheidungen](#)
- [PERF01-BP04 Evaluieren, wie sich Kompromisse auf Kunden und Architektureffizienz auswirken](#)
- [PERF01-BP05 Verwenden von Richtlinien 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 des Workloads 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](#)
- [YouTube-Kanal: AWS](#)
- [AWS-Workshops](#)
- [AWS-Communitys](#)
- 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](#)
- [Mein Lernpfad zum AWS Solutions Architect](#)
- [AWS-Architekturzentrum](#)
- [AWS Partner Network](#)
- [Die AWS-Lösungsbibliothek](#)
- [AWS Knowledge Center](#)
- [Moderne Anwendungen in AWS entwickeln](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Neuerungen bei Amazon EC2](#)
- [AWS re:Invent 2022 – Betriebs- und Infrastrukturkosten mit Amazon ECS senken](#)
- [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 Samples](#)
- [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 den Workload wählen und Entscheidungen mit größerer Zuversicht treffen.

Risikostufe bei fehlender Befolgung dieser bewährten Methode: 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)
  - [AWS-Blog](#)
  - [AWS Training und -Zertifizierungen](#)
  - [YouTube-Kanal: AWS](#)
- 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 [AWS 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 2023 – Fortgeschrittene ereignisgesteuerte Muster mit Amazon EventBridge](#)
- [AWS re:Invent 2023 – Implementierung verteilter Designmuster in AWS](#)
- [AWS re:Invent 2023 – Anwendungsarchitektur als Code](#)

### Zugehörige Beispiele:

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

## PERF01-BP03 Berücksichtigen der Kosten bei architektonischen Entscheidungen

Berücksichtigen Sie die Kosten bei Ihren architektonischen Entscheidungen, um die Ressourcennutzung und Leistungseffizienz der Cloud-Workloads zu verbessern. Wenn Sie sich der Kostenauswirkungen des Cloud-Workloads 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 prüfen keine neuen Services und Features der 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 einem 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 den Cloud-Workload fest.
- Identifizieren Sie die wesentlichen Komponenten (wie Instances und Speicher), die die Kosten des Workloads erhöhen. Nutzen Sie Instrumentierungsservices wie [AWS Pricing Calculator](#) und [AWS Cost Explorer](#), um die wichtigsten Kostentreiber in Ihrem Workload zu identifizieren.
- Verstehen Sie die [Preismodelle](#) in der Cloud, z. B. On-Demand, Reserved Instances, Savings Plans und Spot Instances.
- Verwenden Sie [Best Practices zur Kostenoptimierung bei Well-Architected-Technologien](#), um diese Schlüsselkomponenten aus Kostengründen zu optimieren.
- Überwachen und analysieren Sie kontinuierlich die Kosten, um Möglichkeiten zur Kostenoptimierung im Workload zu identifizieren.
  - Verwenden 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.
  - Verwenden Sie [AWS Cost Anomaly Detection](#), um das automatisierte Erkennen von Kostenanomalien mit Ursachenanalyse zu erhalten.

## Ressourcen

Zugehörige Dokumente:

- [Was ist AWS-Fakturierung und Kostenmanagement?](#)
- [Kostenoptimierung mit AWS](#)
- [Auswahl einer AWS-Kostenverwaltungsstrategie](#)
- [Ein Leitfaden für Anfänger zur AWS-Kostenverwaltung](#)

- [Eine detaillierte Übersicht über das Cost Intelligence Dashboard](#)
- [AWS-Architekturzentrum](#)
- [Die AWS-Lösungsbibliothek](#)
- [AWS Knowledge Center](#)

Zugehörige Videos:

- [This is My Architecture](#)
- [AWS re:Invent 2023 – Neuerungen bei der AWS-Kostenoptimierung](#)
- [AWS re:Invent 2023 – Optimieren der Kosten und Leistung sowie Verfolgen der Fortschritte bei der Schadensbegrenzung](#)
- [AWS re:Invent 2023 – Bewährte Methoden zur Kostenoptimierung für AWS-Speicher](#)
- [AWS re:Invent 2023 – Optimieren der 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 des Workloads 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](#)
- [Amazon QuickSight-KPIs](#)
- [Amazon CloudWatch RUM](#)
- [X-Ray-Dokumentation](#)
- [Resilienzmuster und Kompromisse verstehen, um eine effiziente Architektur in der Cloud zu entwickeln](#)

Zugehörige Videos:

- [Optimieren von Anwendungen mithilfe von 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](#)
- [Amazon CloudWatch RUM Web Client](#)

## PERF01-BP05 Verwenden von Richtlinien und Referenzarchitekturen

Verwenden Sie interne Richtlinien und vorhandene Referenzarchitekturen bei der Auswahl von Services und Konfigurationen, um den 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 den 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 des Cloud-Workloads 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 den Workload in Sandbox-Umgebungen.
- Bleiben Sie über Industriestandards und AWS-Updates auf dem Laufenden, um sicherzustellen, dass die Richtlinien und Referenzarchitekturen zur Optimierung des Cloud-Workloads beitragen.

## Ressourcen

Zugehörige Dokumente:

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

Zugehörige Videos:

- [This is My Architecture](#)
- [AWS re:Invent 2022 – Beschleunigen der Wertschöpfung für Ihr Unternehmen mit SAP und der AWS-Referenzarchitektur](#)

Zugehörige Beispiele:

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

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

Führen Sie einen Benchmark-Vergleich für einen vorhandenen Workload durch, um sich ein Bild über dessen 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 Nutzung dieser bewährten Methode: Durch das Benchmarking der aktuellen Implementierung können Sie Leistungsverbesserungen messen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: 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 der 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 die Benutzererlebnis und Faktoren wie Reaktionszeit und Barrierefreiheit.
  - Identifizieren Sie ein Benchmarking-Tool, das für Ihren Workload geeignet ist. Sie können AWS-Services wie [Amazon CloudWatch](#) oder ein Drittanbieter-Tool verwenden, das mit Ihrem 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 das Workload anzupassen. Dies kann die Änderung von Services oder die Einführung neuer Funktionen beinhalten.
- Optimierung und Wiederholung:
  - Passen Sie die Ressourcenkonfigurationen und -zuweisungen auf der Grundlage Ihrer Benchmarks an.
  - Testen Sie Ihr 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](#)
- [Genomik-Workflows, Teil 5: automatisiertes Benchmarking](#)
- [Benchmarking und Optimierung der Endpunktbereitstellung in Amazon SageMaker JumpStart](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 — Benchmarking von AWS Lambda-Kaltstarts](#)
- [Benchmarking von zustandsbehafteten Services in der Cloud](#)
- [This is My Architecture: Expedia](#)
- [Optimieren von Anwendungen mithilfe von Amazon CloudWatch RUM](#)
- [Demo von Amazon CloudWatch Synthetics](#)

### Zugehörige Beispiele:

- [AWS-Beispiele](#)
- [AWS-SDK-Beispiele](#)
- [Verteilte Belastungstests](#)
- [Messen der Seitenladezeit mit Amazon CloudWatch Synthetics](#)
- [Amazon CloudWatch RUM Web Client](#)

## 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 im 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 den 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 Interessenvertretern 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 den 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](#)).
- Konsultieren Sie andere Teams, Architekturdiagramme und Ressourcen wie AWS Solution Architects, [AWS-Architekturzentrum](#) und [AWS Partner Network](#), um Ihnen bei der Auswahl der richtigen Architektur für Ihren Workload zu helfen.
- Definieren Sie Leistungsmetriken wie Durchsatz und Reaktionszeit, anhand derer Sie die Leistung des Workloads 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:

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

# Computer und Hardware

Die optimale Datenverarbeitungsoption für einen bestimmten Workload kann sich je nach Anwendungsdesign, Nutzungsmustern und Konfigurationseinstellungen unterscheiden. Architekturen verwenden mitunter unterschiedliche Datenverarbeitungslösungen für verschiedene Komponenten und bieten unterschiedliche Möglichkeiten zur Leistungsoptimierung. Die Wahl der falschen Datenverarbeitungslösung für eine Architektur kann die Leistungseffizienz schmälern.

In diesem Schwerpunktbereich finden Sie Anleitungen und bewährte Verfahren zur Identifizierung und Optimierung von Rechenoptionen für die Leistungseffizienz in der Cloud.

## Bewährte Methoden

- [PERF02-BP01 Auswählen der besten Datenverarbeitungsoptionen für den Workload](#)
- [PERF02-BP02 Verstehen verfügbarer Konfigurationen und Features für die Datenverarbeitung](#)
- [PERF02-BP03 Erfassen von Datenverarbeitungsmetriken](#)
- [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 den Workload

Wenn Sie die für den Workload am besten geeignete Computing-Option auswählen, können Sie die Leistung verbessern, unnötige Infrastrukturkosten reduzieren und den Betriebsaufwand für die Aufrechterhaltung des Workloads senken.

### Typische Anti-Muster:

- Sie verwenden dieselbe Option für die Datenverarbeitung, die on-premises verwendet wurde.
- Ihnen fehlt es an Bewusstsein für Cloud-Datenverarbeitungsoptionen, -funktionen 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 den Workload ressourceneffizienter gestalten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: 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 einem 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 die verschiedenen Datenverarbeitungsoptionen, die für Ihren Workload in AWS verfügbar sind (wie unter [PERF01-BP01 Informieren über verfügbare Cloud-Services und -Features](#) beschrieben). Hier finden Sie einige wichtige AWS-Datenverarbeitungsoptionen, ihre Eigenschaften und gängige Anwendungsfälle:

AWS service	Key characteristics	Common use cases
<a href="#">Amazon Elastic Compute Cloud (Amazon EC2)</a>	Has dedicated option for hardware, license requirements, large selection of different	Lift and shift migrations, monolithic application, hybrid

AWS service	Key characteristics	Common use cases
	instance families, processor types and compute accelerators	environments, enterprise applications
<a href="#">Amazon Elastic Container Service (Amazon ECS)</a> , <a href="#">Amazon Elastic Kubernetes Service (Amazon EKS)</a>	Easy deployment, consistent environments, scalable	Microservices, hybrid environments
<a href="#">AWS Lambda</a>	<a href="#">Serverlose Datenverarbeitung</a> service that runs code in response to events and automatically manages the underlying compute resources.	Microservices, event-driven applications
<a href="#">AWS Batch</a>	Efficiently and dynamically provisions and scales <a href="#">Amazon Elastic Container Service (Amazon ECS)</a> , <a href="#">Amazon Elastic Kubernetes Service (Amazon EKS)</a> , and <a href="#">AWS Fargate</a> compute resources, with an option to use On-Demand or Spot Instances based on your job requirements	HPC, train ML models
<a href="#">Amazon Lightsail</a>	Preconfigured Linux and Windows application for running small workloads	Simple web applications, custom 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 Computing-Ressourcen kontinuierlich auf der Grundlage realer Nutzungsmuster zu optimieren.

## Ressourcen

### Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon EC2-Instance-Typen](#)
- [Amazon EKS-Container: Amazon EKS-Worker-Knoten](#)
- [Amazon ECS-Container: Amazon ECS-Container-Instances](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)
- [Präskriptive Anleitung für Container](#)
- [Präskriptive Anleitung für Serverless](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – AWS Graviton: Das beste Preis-Leistungs-Verhältnis für Ihre AWS-Workloads](#)
- [AWS re:Invent 2023 – Neue generative KI-Funktionen von Amazon Elastic Compute Cloud in AMS](#)
- [AWS re:Invent 2023 – Neuerungen bei Amazon Elastic Compute Cloud](#)
- [AWS re:Invent 2023 – Intelligentes Sparen: Amazon Elastic Compute Cloud-Strategien zur Kostenoptimierung](#)
- [AWS re:Invent 2021 – Amazon Elastic Compute Cloud der nächsten Generation: Ausführliche Beschreibung des Nitro System](#)
- [AWS re:Invent 2019 – Optimieren von Leistung und Kosten für Ihr AWS-Compuntung](#)
- [AWS re:Invent 2019 – Amazon Elastic Compute Cloud-Grundlagen](#)

- [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 Ihr AWS-Computing](#)
- [Amazon EC2-Grundlagen](#)
- [ML-Modelle für Inferenz mit hoher Leistung und niedrigen Kosten bereitstellen](#)

Zugehörige Beispiele:

- [Migration der Webanwendung zu Containern](#)
- [Ausführen eines Serverless-„Hello World“](#)
- [Amazon EKS-Workshop](#)
- [Amazon EC2-Workshop](#)
- [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 den 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

1. Verstehen Sie die Workload-Anforderungen (wie CPU-Bedarf, Arbeitsspeicher und Latenz).
2. 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 die 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>• <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 <a href="#">Auto Scaling</a> Konfiguration zur Anpassung der Datenverarbeitungsressourcen an die Datenverkehrsmuster.
Dimensionierung	<ul style="list-style-type: none"><li>• Verwenden Sie <a href="#">Compute Optimizer</a> zum Erhalt von Machine-Learning-gestützten Empfehlungen dazu, welche Datenverarbeitungskonfiguration am besten Ihren Datenverarbeitungsmerkmalen entspricht.</li><li>• Verwenden Sie <a href="#">AWS Lambda Power Tuning</a> können Sie die beste Konfiguration für Ihre Lambda-Funktion auswählen.</li></ul>

Konfigurationsoption	Beispiele
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 Ihren 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](#)
- [Steuerung des Prozessorzustands für Ihre Amazon EC2-Instance](#)
- [Amazon EKS-Container: Amazon EKS-Worker-Knoten](#)
- [Amazon ECS-Container: 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 – Neue generative KI-Funktionen von Amazon EC2 in AWS Management Console](#)
- [AWS re:Invent 2023 – Neuerungen bei Amazon EC2](#)
- [AWS re:Invent 2023 – Intelligentes Sparen: Amazon EC2-Strategien zur Kostenoptimierung](#)
- [AWS re:Invent 2021 – Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro System](#)
- [AWS re:Invent 2019 – Amazon EC2-Grundlagen](#)

- [AWS re:INVENT 2022 – https://www.youtube.com/watch?v=5B4-s\\_ivn1o](https://www.youtube.com/watch?v=5B4-s_ivn1o)

Zugehörige Beispiele:

- [Compute Optimizer-Demo-Code](#)
- [Workshop zu Amazon EC2 Spot Instances](#)
- [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](#)
- [Amazon EKS-Workshop](#)

## PERF02-BP03 Erfassen von Datenverarbeitungsmetriken

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 im 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 die Erfassung von Metriken ein entscheidender Schritt zur Verbesserung von Sicherheit, Kosteneffizienz, Leistung und Nachhaltigkeit. AWS stellt eine Vielzahl von Leistungsmetriken bereit und nutzt dazu Überwachungsservices wie [Amazon CloudWatch](#), um Ihnen wertvolle Einblicke zu bieten. Metriken wie CPU-Nutzung, Arbeitsspeicherauslastung, Datenträger-E/A sowie eingehender und ausgehender Netzwerkverkehr können Einblick in die Nutzung bzw. in Leistungsengpässe bieten. Nutzen Sie diese Metriken im Rahmen eines datengestützten Ansatzes, der Ihnen die aktive Feinabstimmung und Optimierung der vom 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

1. Identifizieren Sie, welche Leistungsmetriken für den Workload relevant sind. Sie sollten Metriken zur Ressourcennutzung und zum Betrieb des Cloud-Workloads (wie Reaktionszeit und Durchsatz) erfassen.
  - a. [Amazon EC2-Standardmetriken](#)
  - b. [Amazon ECS-Standardmetriken](#)
  - c. [Amazon EKS-Standardmetriken](#)
  - d. [Lambda-Standardmetriken](#)
  - e. [Amazon EC2-Arbeitsspeicher- und -Datenträgermetriken](#)
2. Wählen Sie die richtige Protokollierungs- und Überwachungslösung für den Workload aus und richten Sie sie ein.
  - a. [AWS-native Beobachtbarkeit](#)
  - b. [AWS Distro for OpenTelemetry](#)
  - c. [Amazon Managed Service for Prometheus](#)
3. Definieren Sie den erforderlichen Filter und die erforderliche Aggregation für die Metriken auf der Grundlage Ihrer Workload-Anforderungen.
  - a. [Quantifizieren benutzerdefinierter Anwendungsmetriken mit Amazon CloudWatch Logs und Metrikfiltern](#)
  - b. [Erfassen benutzerdefinierter Metriken mit Amazon CloudWatch und strategischer Markierung](#)

4. Konfigurieren Sie Richtlinien zur Datenaufbewahrung für Ihre Metriken so, dass sie Ihren Sicherheits- und Betriebszielen entsprechen.
  - a. [Standard-Datenaufbewahrung für CloudWatch-Metriken](#)
  - b. [Standard-Datenaufbewahrung für CloudWatch Logs](#)
5. Erstellen Sie bei Bedarf Alarme und Benachrichtigungen für Ihre Metriken, damit Sie proaktiv auf leistungsbezogene Probleme reagieren können.
  - a. [Alarme für benutzerdefinierte Metriken mit der Amazon CloudWatch-Erkennung von Unregelmäßigkeiten erstellen](#)
  - b. [Metriken und Alarmen für bestimmte Webseiten mit Amazon CloudWatch RUM erstellen](#)
6. Verwenden Sie die Automatisierung, um die Kundendienstmitarbeiter für die Metrik- und Protokollaggregation einzusetzen.
  - a. [AWS Systems Manager-Automatisierung](#)
  - b. [OpenTelemetry Collector](#)

## Ressourcen

### Zugehörige Dokumente:

- [Überwachung und Beobachtbarkeit](#)
- [Bewährte Methoden: Implementierung der Beobachtbarkeit mit AWS](#)
- [Amazon CloudWatch-Dokumentation](#)
- [Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und On-Premises-Servern mit dem CloudWatch Agent](#)
- [Zugriff auf Amazon CloudWatch Logs für AWS Lambda](#)
- [Verwenden von CloudWatch Logs mit Container-Instances](#)
- [Veröffentlichen von benutzerdefinierten Metriken](#)
- [AWS Answers: Zentralisierte Protokollierung](#)
- [CloudWatch-Services, die AWS-Metriken veröffentlichen](#)
- [Überwachen von Amazon EKS auf 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 Beobachtbarkeitsstrategie](#)
- [AWS re:Invent 2023 – Nahtlose Beobachtbarkeit mit AWS Distro for OpenTelemetry](#)
- [Verwaltung der Anwendungsleistung in AWS](#)

Zugehörige Beispiele:

- [AWS für Linux-Workloads Immersion Day – Amazon CloudWatch](#)
- [Überwachung von Clustern und Containern in Amazon ECS](#)
- [Überwachung mit Amazon CloudWatch-Dashboards](#)
- [Amazon EKS-Workshop](#)

## PERF02-BP04 Konfigurieren und richtiges Dimensionieren von Datenverarbeitungsressourcen

Konfigurieren und passen Sie die Größe der Datenverarbeitungsressourcen so an, dass sie den Leistungsanforderungen des 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 den 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 zu hohe Bereitstellung von Cloud-Ressourcen kann zu zusätzlichen Kosten führen, während eine unzureichende Bereitstellung zu einer schlechten Leistung und einem negativen Kundenerlebnis führen kann. AWS bietet Tools wie [AWS Compute Optimizer](#) und [AWS Trusted Advisor](#), die historische Daten verwenden, um Empfehlungen zur richtigen Dimensionierung Ihrer Rechenressourcen abzugeben.

### Implementierungsschritte

- Wählen Sie eine Instance, die am besten zu Ihren Anforderungen passt:
  - [Wie wähle ich einen geeigneten Amazon EC2-Instance-Typ für meinen Workload aus?](#)
  - [Attributbasierte Auswahl des Instance-Typs für die Amazon EC2 Flotte](#)
  - [Erstellen Sie eine Auto Scaling-Gruppe unter Verwendung einer attributbasierten 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 des 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 kurz andauernde Workloads [Amazon CloudWatch-Instance-Metriken](#) wie die CPUUtilization um festzustellen, ob die Instance zu wenig oder zu stark ausgelastet ist.
  - 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 Ihres Workloads.

## Ressourcen

### Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon EC2-Instance-Typen](#)
- [Amazon ECS-Container: Amazon ECS-Container-Instances](#)
- [Amazon EKS-Container: Amazon EKS-Worker-Knoten](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)
- [Steuerung des Prozessorzustands für Ihre Amazon EC2-Instance](#)

### Zugehörige Videos:

- [Amazon EC2-Grundlagen](#)
- [AWS re:Invent 2023 – AWS Graviton: Das beste Preis-Leistungs-Verhältnis für Ihre AWS-Workloads](#)
- [AWS re:Invent 2023 – Neue generative KI-Funktionen von Amazon EC2 in AWS Management Console](#)
- [AWS re:Invent 2023 – Neuerungen bei Amazon EC2](#)
- [AWS re:Invent 2023 – Intelligentes Sparen: Amazon EC2-Strategien zur Kostenoptimierung](#)
- [AWS re:Invent 2021 – Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro System](#)
- [AWS re:Invent 2019 – Amazon EC2-Grundlagen](#)

### Zugehörige Beispiele:

- [AWS Compute Optimizer-Demo-Code](#)
- [Amazon EKS-Workshop](#)
- [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 den 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 Rechenressourcen 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.
- Vorausschauende 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 Merkmal 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.
<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.

- Das Skalieren wird häufig im Zusammenhang mit Datenverarbeitungsservices wie Amazon EC2-Instances oder AWS Lambda-Funktionen genannt. 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 des bereitgestellten Workloads 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 eine [benutzerdefinierte Metrik](#) für Ihre Skalierungsrichtlinie verwenden, falls erforderlich. Beachten Sie zur Auswahl der geeigneten Metriken die folgenden Hinweise zu Amazon EC2:

- Es sollte sich um eine gültige Nutzungsmetrik handeln, die beschreibt, wie stark eine Instance genutzt wird.
- Der Metrikwert muss proportional zur Anzahl der Instances in der Auto Scaling-Gruppe steigen oder sinken.
- Vergewissern Sie sich, dass Sie [dynamische Skalierung](#) anstelle von [manueller Skalierung](#) für Ihre Auto Scaling-Gruppe verwenden. Weiterhin empfehlen wir, dass Sie [Zielverfolgungs-Skalierungsrichtlinien](#) für Ihre dynamische Skalierung verwenden.
- Prüfen Sie, ob Workload-Bereitstellungen mit beiden Skalierungen (nach oben und unten) umgehen können. Beispielsweise können Sie [den Aktivitätsverlauf](#) verwenden, um eine Skalierungsaktivität für eine Auto Scaling-Gruppe zu verifizieren.
- Evaluieren Sie Ihren 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 Kapazität vermeiden. Weitere Details finden Sie unter [Vorausschauende Skalierung mit Amazon EC2 Auto Scaling](#).

## Ressourcen

### Zugehörige Dokumente:

- [Cloud Computing mit AWS](#)
- [Amazon EC2-Instance-Typen](#)
- [Amazon ECS-Container: Amazon ECS-Container-Instances](#)
- [Amazon EKS-Container: Amazon EKS-Worker-Knoten](#)
- [Funktionen: Lambda-Funktionskonfiguration](#)
- [Steuerung des Prozessorzustands für Ihre Amazon EC2-Instance](#)
- [Ausführliche Beschreibung von 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 – Neue generative KI-Funktionen von Amazon EC2 in der AWS-Managementkonsole](#)

- [AWS re:Invent 2023 – Neuerungen bei Amazon EC2](#)
- [AWS re:Invent 2023 – Intelligentes Sparen: Amazon EC2-Strategien zur Kostenoptimierung](#)
- [AWS re:Invent 2021 – Amazon EC2 der neuesten Generation: Ausführliche Beschreibung des Nitro System](#)
- [AWS re:Invent 2019 – Amazon EC2-Grundlagen](#)

Zugehörige Beispiele:

- [Amazon EC2 Auto Scaling-Gruppenbeispiele](#)
- [Amazon EKS-Workshop](#)
- [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 im Workload keine Benchmark einer universellen 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, wenn diese bewährte Methode nicht eingeführt wird: 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, Transcodierung 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 Ihren Workload abgestimmt ist, z. B. [AWS Trainium](#), [AWS Inferentia](#) oder [Amazon EC2 DL1](#). AWS-Inferentia-Instances wie Inf2-Instances [bieten eine um bis zu 50 % bessere Leistung/Watt als vergleichbare Amazon EC2-Instances](#).
- Erfassen Sie Nutzungsmetriken für Ihre beschleunigten Computing-Instances. Sie können z. B. den CloudWatch Agent verwenden, um Metriken wie `utilization_gpu` und `utilization_memory` für Ihre GPUs zu erfassen. Dies wird im [Artikel zum Erfassen von NVIDIA GPU-Metriken mit Amazon CloudWatch](#) genauer 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](#)
  - [Optimieren von E/A für die GPU-Leistungsoptimierung von Deep Learning-Training in Amazon SageMaker](#)
- Verwenden Sie die aktuellen leistungsstarken Bibliotheken und GPU-Treiber.
- Automatisieren Sie die Freigabe nicht genutzter GPU-Instances.

## Ressourcen

Zugehörige Dokumente:

- [Arbeiten mit GPUs in 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](#)

- [Accelerated Computing](#)
- [Amazon EC2 VT1-Instances](#)
- [Wie wähle ich einen geeigneten Amazon EC2 Instance-Typ für meinen Workload aus?](#)
- [Auswählen des besten KI-Beschleunigers und der Modellkompilierung für Computer Vision Inference mit Amazon SageMaker](#)

#### Zugehörige Videos:

- [AWS re:Invent 2021 – Auswählen von Amazon Elastic Compute Cloud-GPU-Instances für Deep Learning](#)
- [AWS re:Invent 2022 – \[NEUER LAUNCH\] Einführung von AWS-Inferentia2-basierten Amazon EC2-Inf2-Instances](#)
- [AWS re:Invent 2022 – Beschleunigung von Deep Learning und schnellere Innovationen mit AWS Trainium](#)
- [AWS re:Invent 2022 – Deep Learning in AWS mit NVIDIA: Vom Training bis zur Bereitstellung](#)

#### Zugehörige Beispiele:

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

# Datenverwaltung

Die optimale Datenverwaltungslösung für ein spezielles System richtet sich nach dem Datentyp (Block, Datei oder Objekt), den Zugriffsmustern (Zufallsprinzip oder sequenziell), dem erforderlichen Durchsatz, der Zugriffshäufigkeit (online, offline, Archiv), der Aktualisierungshäufigkeit (WORM, dynamisch) sowie den Einschränkungen hinsichtlich der Verfügbarkeit und Langlebigkeit. Gut geplante Workloads verwenden speziell entwickelte Datenspeicher, die verschiedene Features zur Leistungsverbesserung ermöglichen.

In diesem Schwerpunktbereich werden Anleitungen und bewährte Methoden zur Optimierung der Datenspeicherung, der Datenbewegungs- 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 den 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 unterbewerten, können Sie die effizienteste und leistungsfähigste Speichertechnologie ermitteln, die für Ihre Workload-Anforderungen geeignet ist.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: 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 Ihrem 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)

- Aufbewahrungsfrist der Daten
- Erfahren Sie mehr über die verschiedenen Datenspeicher (Speicher- und Datenbank-Services), die für Ihren 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
Object storage	<a href="#">Amazon S3</a>	Unlimited scalability, high availability, and multiple options for accessibility. Transferring and accessing objects in and out of Amazon S3 can use a service, such as <a href="#">Transfer Acceleration</a> or <a href="#">Zugriffspunkte</a> , to support your location, security needs, and access patterns.
Archiving storage	<a href="#">Amazon S3 Glacier</a>	Built for data archiving.
Streaming storage	<a href="#">Amazon Kinesis</a> <a href="#">Amazon Managed Streaming for Apache Kafka (Amazon MSK)</a>	Efficient ingestion and storage of streaming data.
Shared file system	<a href="#">Amazon Elastic File System (Amazon EFS)</a>	Bereitstellbares Dateisystem, auf das mehrere Arten von Datenverarbeitungslösungen zugreifen können.
Shared file system	<a href="#">Amazon FSx</a>	Built on the latest AWS compute solutions to support four commonly used file systems: NetApp ONTAP, OpenZFS, Windows File Server, and Lustre. Amazon

Typ	AWS-Services	Schlüsselmerkmale
		FSx <a href="#">variieren</a> vary per file system and should be considered when selecting the right file system for your workload needs.
Block storage	<a href="#">Amazon Elastic Block Store (Amazon EBS)</a>	Scalable, high-performance block-storage service designed for Amazon Elastic Compute Cloud (Amazon EC2). Amazon EBS includes SSD-backed storage for transactional, IOPS-intensive workloads and HDD-backed storage for throughput-intensive workloads.
Relational database	<a href="#">Amazon Aurora</a> , <a href="#">Amazon RDS</a> , <a href="#">Amazon Redshift</a> .	Designed to support ACID (atomicity, consistency, isolation, durability) transactions, and maintain referential integrity and strong data consistency. Many traditional applications, enterprise resource planning (ERP), customer relationship management (CRM), and ecommerce use relational databases to store their data.

Typ	AWS-Services	Schlüsselmerkmale
Key-value database	<a href="#">Amazon DynamoDB</a>	Optimized for common access patterns, typically to store and retrieve large volumes of data. High-traffic web apps, ecommerce systems, and gaming applications are typical use-cases for key-value databases.
Document database	<a href="#">Amazon DocumentDB</a>	Designed to store semi-structured data as JSON-like documents. These databases help developers build and update applications such as content management, catalogs, and user profiles quickly.
In-memory database	<a href="#">Amazon ElastiCache</a> , <a href="#">Amazon MemoryDB für Redis</a>	Used for applications that require real-time access to data, lowest latency and highest throughput. You may use in-memory databases for application caching, session management, gaming leaderboards, low latency ML feature store, microservices messaging system, and a high-throughput streaming mechanism

Typ	AWS-Services	Schlüsselmerkmale
Graph database	<a href="#">Amazon Neptune</a>	Used for applications that must navigate and query millions of relationships between highly connected graph datasets with millisecond latency at large scale. Many companies use graph databases for fraud detection, social networking, and recommendation engines.
Time Series database	<a href="#">Amazon Timestream</a>	Used to efficiently collect, synthesize, and derive insights from data that changes over time. IoT applications, DevOps, and industrial telemetry can utilize time-series databases.
Wide column	<a href="#">Amazon Keyspaces (für Apache Cassandra)</a>	Uses tables, rows, and columns, but unlike a relational database, the names and format of the columns can vary from row to row in the same table. You typically see a wide column store in high scale industrial apps for equipment maintenance, fleet management, and route optimization.

Typ	AWS-Services	Schlüsselmerkmale
Ledger	<a href="#">Amazon Quantum Ledger Database (Amazon QLDB)</a>	Provides a centralized and trusted authority to maintain a scalable, immutable, and cryptographically verifiable record of transactions for every application. We see ledger databases used for systems of record, supply chain, registrations, and even banking transactions.

- Wenn Sie eine Datenplattform aufbauen, nutzen Sie [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 Ihren Workload berücksichtigen müssen, lauten wie folgt:

Question	Things to consider
How is the data structured?	<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 for Redis</a> oder <a href="#">Amazon MemoryDB for Redis</a></li> </ul>
What level of referential integrity is required?	<ul style="list-style-type: none"> <li>• Bei Fremdschlüsseleinschränkungen können relationale Datenbanken wie <a href="#">Amazon RDS</a> und <a href="#">Aurora</a> diese Integritäts Ebene bieten.</li> <li>• Üblicherweise würden Sie innerhalb eines NoSQL-Datenmodells Ihre Daten in ein einzelnes Dokument oder eine Sammlung von Dokumenten denormalisieren, die in</li> </ul>

Question	Things to consider
	<p>einer einzelnen Anfrage abgerufen werden können, anstatt Daten in Dokumenten oder Tabellen zusammenzufügen.</p>
<p>Is ACID (atomicity, consistency, isolation, durability) compliance required?</p>	<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 strikte Konsistenz für eine <a href="#">NoSQL-Datenbank</a> erforderlich ist, können Sie strikt konsistente Lesevorgänge mithilfe von <a href="#">DynamoDB</a> verwenden.</li> </ul>
<p>How will the storage requirements change over time? How does this impact scalability?</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>What is the proportion of read queries in relation to write queries? Would caching be likely to improve performance?</p>	<ul style="list-style-type: none"> <li>• Leseintensive Workloads können 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 wie <a href="#">Amazon RDS</a> ausgelagert werden.</li> </ul>

Question	Things to consider
Does storage and modification (OLTP - Online Transaction Processing) or retrieval and reporting (OLAP - Online Analytical Processing) have a higher priority?	<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 Analysen mithilfe von <a href="#">Athena</a> oder <a href="#">Amazon QuickSight</a>.</li></ul>
What level of durability does the data require?	<ul style="list-style-type: none"><li>• Aurora repliziert Ihre Daten automatisch in drei Availability Zones innerhalb von 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>

Question	Things to consider
Is there a desire to move away from commercial database engines or licensing costs?	<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>
What is the operational expectation for the database? Is moving to managed services a primary concern?	<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>
How is the database currently accessed? Is it only application access, or are there business intelligence (BI) users and other connected off-the-shelf applications?	<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:

- [Amazon EBS-Volume-Typen](#)
- [Amazon EC2-Speicher](#)
- [Amazon EFS: Leistung von Amazon EFS](#)

- [Leistung von Amazon FSx for Lustre](#)
- [Leistung von Amazon FSx for Windows File Server](#)
- [Amazon S3 Glacier: S3 Glacier-Dokumentation](#)
- [Amazon S3: Überlegungen zu Anfragerate und Leistung](#)
- [Cloud-Speicher mit AWS](#)
- [Amazon EBS-E/A-Merkmale](#)
- [Cloud-Datenbanken mit AWS](#)
- [AWS-Datenbank-Caching](#)
- [DynamoDB Accelerator](#)
- [Bewährte Methoden für Amazon Aurora](#)
- [Leistung von Amazon Redshift](#)
- [Die besten 10 Leistungstipps für Amazon Athena](#)
- [Bewährte Methoden für Amazon Redshift Spectrum](#)
- [Bewährte Methoden für Amazon DynamoDB](#)
- [Wählen Sie zwischen Amazon EC2 und Amazon RDS](#)
- [Bewährte Methoden für die Implementierung von Amazon ElastiCache](#)

#### Zugehörige Videos:

- [AWS re:Invent 2023 – Steigerung der Effizienz von Amazon Elastic Block Store und der allgemeinen Kosteneffizienz](#)
- [AWS re:Invent 2023 – Optimierung der Speicherkosten und der Leistung mit Amazon Simple Storage Service](#)
- [AWS re:Invent 2023 – Aufbau und Optimierung eines Data Lake in Amazon Simple Storage Service](#)
- [AWS re:Invent 2022: Erstellen von modernen Datenarchitekturen in AWS](#)
- [AWS re:Invent 2022 – Aufbau von Data-Mesh-Architekturen in AWS](#)
- [AWS re:Invent 2023 – Vertiefung in Amazon Aurora und seine Innovationen](#)
- [AWS re:Invent 2023: Fortschrittliche Datenmodellierung mit Amazon DynamoDB](#)
- [AWS re:Invent 2022: Modernisierung von Apps mit speziell entwickelten Datenbanken](#)
- [Vertiefung in Amazon DynamoDB: Fortschrittliche Entwurfsmuster](#)

Zugehörige Beispiele:

- [AWS-Workshop „Speziell entwickelte Datenbanken“](#)
- [Datenbanken für Entwickler](#)
- [AWS Immersion Day in moderne Datenarchitekturen](#)
- [Erstellung eines Data Mesh in 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 für die Datenbankmodernisierung](#)
- [Amazon Neptune-Beispiele](#)

## PERF03-BP02 Bewerten verfügbarer Konfigurationsoptionen für den Datenspeicher

Machen Sie sich mit den verschiedenen Funktionen und Konfigurationsoptionen vertraut, die für Ihre Datenspeicher verfügbar sind, und bewerten Sie sie, um Speicherplatz und Leistung für Ihren 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, wenn diese bewährte Methode nicht eingeführt wird: mittel

## Implementierungsleitfaden

Für einen 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:

Configuration option	Examples
Offloading reads (like read replicas and 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 Amazon ElastiCache for Redis-Cluster 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</li> </ul>

Configuration option	Examples
	<p>das Hinzufügen von Read Replicas, um die Lesevorgänge des Workloads auszulagern.</p> <ul style="list-style-type: none"><li>• Serverless-Datenbanken wie DynamoDB skalieren automatisch. Stellen Sie sicher, dass Sie ausreichend Read Capacity Units (RCU) bereitstellen, um den Workload zu verarbeiten.</li></ul>

Configuration option	Examples
Scaling writes (like partition key sharding or introducing a queue)	<ul style="list-style-type: none"><li>• Bei relationalen Datenbanken können Sie die Größe der Instance erhöhen, um einen erhöhten 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>

Configuration option	Examples
Policies to manage the lifecycle of your datasets	<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 von <a href="#">Amazon S3-Storage-Lens</a>-Metriken 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>
Connection management and 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](#)
- [Amazon EBS-Volumen-Typen](#)
- [Amazon EC2-Speicher](#)
- [Amazon EFS: Leistung von Amazon EFS](#)
- [Leistung von Amazon FSx for Lustre](#)
- [Leistung von Amazon FSx for Windows File Server](#)
- [Amazon S3 Glacier: S3 Glacier-Dokumentation](#)
- [Amazon S3: Überlegungen zu Anfragerate und Leistung](#)
- [Amazon EBS-E/A-Merkmale](#)
- [Cloud-Datenbanken mit AWS](#)
- [AWS-Datenbank-Caching](#)
- [DynamoDB Accelerator](#)
- [Bewährte Methoden für Amazon Aurora](#)
- [Leistung von Amazon Redshift](#)
- [Die besten 10 Leistungstipps für Amazon Athena](#)
- [Bewährte Methoden für Amazon Redshift Spectrum](#)
- [Bewährte Methoden für Amazon DynamoDB](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Steigerung der Effizienz von Amazon Elastic Block Store und der allgemeinen Kosteneffizienz](#)
- [AWS re:Invent 2023 – Optimierung der Speicherkosten und der Leistung mit Amazon Simple Storage Service](#)

- [AWS re:Invent 2023 – Aufbau und Optimierung eines Data Lake in Amazon Simple Storage Service](#)
- [AWS re:Invent 2023 – Neuerungen bei AWS-Dateispeicher](#)
- [AWS re:Invent 2023 – Vertiefung in Amazon DynamoDB](#)

Zugehörige Beispiele:

- [AWS-Workshop „Speziell entwickelte Datenbanken“](#)
- [Datenbanken für Entwickler](#)
- [AWS Immersion Day in moderne Datenarchitekturen](#)
- [Amazon EBS – automatische Skalierung](#)
- [Amazon S3-Beispiele](#)
- [Amazon DynamoDB-Beispiele](#)
- [Beispiele von AWS-Datenbankmigration](#)
- [Workshop für die Datenbankmodernisierung](#)
- [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 Ihres Workloads.
- 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 Ihres Workloads 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 der Workload ist und wie die Datenverwaltungslösung genutzt wird. Nutzen Sie diese Metriken im Rahmen eines datengestützten Ansatzes, der Ihnen die Feinabstimmung und Optimierung der vom Workload genutzten Ressourcen ermöglicht.

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

## Implementierungsschritte

1. Identifizieren Sie die wichtigsten Leistungsmetriken, die der Datenspeicher verfolgen soll.
  - a. [Metriken und Dimensionen von Amazon S3](#)
  - b. [Überwachungsmetriken für innerhalb einer Amazon RDS-Instance](#)
  - c. [Überwachen der DB-Last mit Performance Insights auf Amazon RDS](#)

- d. [Überblick über Erweiterte Überwachung](#)
  - e. [Metriken und Dimensionen von DynamoDB](#)
  - f. [Überwachen von DynamoDB Accelerator](#)
  - g. [Überwachen von Amazon MemoryDB for Redis mit Amazon CloudWatch](#)
  - h. [Welche Metriken sollte ich überwachen?](#)
  - i. [Überwachen der Amazon Redshift-Cluster-Leistung](#)
  - j. [Metriken und Dimensionen von Timestream](#)
  - k. [Amazon CloudWatch-Metriken für Amazon Aurora](#)
  - l. [Protokollieren und Überwachen von Amazon Keyspaces \(for Apache Cassandra\)](#)
  - m. [Überwachen von Amazon Neptune-Ressourcen](#)
2. Verwenden Sie eine zugelassene Protokollierungs- und Überwachungslösung, um diese Metriken zu erfassen. [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.
3. Prüfen Sie, ob die Datenspeicherüberwachung von einer Machine-Learning-Lösung profitieren kann, die Leistungsanomalien erkennt.
- a. [Amazon DevOps Guru für Amazon RDS](#) ermöglicht einen Einblick in Leistungsprobleme und bietet Empfehlungen für Korrekturmaßnahmen.
4. Konfigurieren Sie die Datenaufbewahrung in Ihrer Überwachungs- und Protokollierungslösung so, dass sie Ihren Sicherheits- und Betriebszielen entspricht.
- a. [Standard-Datenaufbewahrung für CloudWatch-Metriken](#)
  - b. [Standard-Datenaufbewahrung für CloudWatch Logs](#)

## Ressourcen

Zugehörige Dokumente:

- [AWS-Datenbank-Caching](#)
- [Die besten 10 Leistungstipps für Amazon Athena](#)
- [Bewährte Methoden für Amazon Aurora](#)
- [DynamoDB Accelerator](#)

- [Bewährte Methoden für Amazon DynamoDB](#)
- [Bewährte Methoden für Amazon Redshift Spectrum](#)
- [Amazon Redshift-Leistung](#)
- [Cloud-Datenbanken mit AWS](#)
- [Amazon RDS Performance Insights](#)

#### Zugehörige Videos:

- [AWS re:Invent 2022 – Leistungsüberwachung mit Amazon RDS und Aurora, mit Autodesk](#)
- [Überwachung und Optimierung der Datenbankleistung mit Amazon DevOps Guru für Amazon RDS](#)
- [AWS re:Invent 2023 – Neuerungen bei AWS-Dateispeicher](#)
- [AWS re:Invent 2023 – Ausführliche Beschreibung von Amazon DynamoDB](#)
- [AWS re:Invent 2023 – Aufbau und Optimierung eines Data Lake in Amazon S3](#)
- [AWS re:Invent 2023 – Neuerungen bei AWS-Dateispeicher](#)
- [AWS re:Invent 2023 – Ausführliche Beschreibung von Amazon DynamoDB](#)
- [Bewährte Methoden für die Überwachung von Redis-Workloads auf Amazon ElastiCache](#)

#### Zugehörige Beispiele:

- [Framework zur AWS-Datensatzerfassung und Sammlung von Metriken](#)
- [Workshop zur Überwachung von Amazon RDS](#)
- [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 des gesamten Cloud-Workloads 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:
  - Verwenden eines [spaltenförmigen 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.
    - [Partitionierung von Daten in Athena](#)
    - [Partitionen und Datenverteilung](#)

- Datenindizierung für die gemeinsamen Spalten in der Abfrage.
- Verwenden Sie materialisierte Ansichten für häufige Abfragen.
  - [Verstehen von materialisierten Ansichten](#)
  - [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 das Ausführen von Statistiken.
- Experimentieren und testen Sie Strategien in einer Nicht-Produktionsumgebung.

## Ressourcen

### Zugehörige Dokumente:

- [Bewährte Methoden für Amazon Aurora](#)
- [Amazon Redshift-Leistung](#)
- [Die besten 10 Leistungstipps für Amazon Athena](#)
- [AWS-Datenbank-Caching](#)
- [Bewährte Methoden für die Implementierung von Amazon ElastiCache](#)
- [Partitionierung von Daten in Athena](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Bewährte Methoden zur Kostenoptimierung für AWS-Speicher](#)
- [AWS re:Invent 2022 – Leistungsüberwachung mit Amazon RDS und Aurora, mit Autodesk](#)
- [Optimieren von Amazon Athena-Abfragen mit neuen Tools zur Abfrageanalyse](#)

### Zugehörige Beispiele:

- [Amazon S3 Select – Abfragen von Daten ohne Server oder Datenbanken](#)
- [AWS-Workshop „Speziell entwickelte Datenbanken“](#)

# 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, die Remoteanrufe tätigt (als clientseitiges Caching bezeichnet) oder indem Sie einen schnellen sekundären Service zum Speichern der Daten verwenden ( 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

Zum Lösen des Problems doppelter Daten zwischen Clients kann ein schneller externer Service oder 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

1. 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 des API-Cachings zur Verbesserung der Reaktionsfähigkeit](#)
2. Identifizieren Sie die geeignete Caching-Strategie, die am besten zu Ihrem Zugriffsmuster passt.
  - [Strategien für Zwischenspeicher](#)
  - [AWS-Caching-Lösungen](#)

3. Folgen Sie den [bewährten Methoden für das Caching](#) für Ihren Datenspeicher.
4. 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.
5. 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 for Redis](#)
6. 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?](#)
  - [Bewährte Methoden für die Überwachung von Redis-Workloads auf Amazon ElastiCache](#)
  - [Bewährte Methoden für die Überwachung mit Amazon ElastiCache for Redis unter Verwendung von Amazon CloudWatch](#)
7. Implementieren Sie [die Datenreplikation](#), um Lesevorgänge auf mehrere Instances auszulagern und die Leseleistung und Verfügbarkeit von Daten zu verbessern.

## Ressourcen

### Zugehörige Dokumente:

- [Verwenden von Amazon ElastiCache Well-Architected Lense](#)
- [Bewährte Methoden für die Überwachung mit Amazon ElastiCache for Redis unter Verwendung von Amazon CloudWatch](#)
- [Welche Metriken sollte ich überwachen?](#)
- [Whitepaper „Skalierbare Leistung mit Amazon ElastiCache“](#)
- [Caching-Herausforderungen und -Strategien](#)

### Zugehörige Videos:

- [Lernpfad zu Amazon ElastiCache](#)
- [Erfolgreiches Design mit bewährten Methoden für Amazon ElastiCache](#)
- [AWS re:Invent 2020 – Erfolgreiches Design mit bewährten Methoden für Amazon ElastiCache](#)

- [AWS re:Invent 2023 – \[LAUNCH\] Einführung von Amazon ElastiCache Serverless](#)
- [AWS re:Invent 2022 – 5 hervorragende Methoden, um die Datenebene mit Redis neu zu gestalten](#)
- [AWS re:Invent 2021 – Ausführliche Beschreibung von Amazon ElastiCache for Redis](#)

Zugehörige Beispiele:

- [Boosting MySQL database performance with Amazon ElastiCache for Redis \(Steigern der MySQL-Datenbankleistung mit Amazon ElastiCache for Redis\)](#)

# Networking und Bereitstellung von Inhalten

Welche Networking-Lösung für einen 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 lokalen 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 Produktfunktionen wie Enhanced Networking, für Amazon EC2-Netzwerk optimierte Instances, Amazon S3 Transfer Acceleration sowie den dynamischen Amazon CloudFront-Service. Zur Verbesserung der Latenz und der Stabilität des Netzwerks finden Sie in AWS zudem Networking-Funktionen 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 Verfahren für die Entwicklung, Konfiguration und den Betrieb effizienter Networking- 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 Auswählen des Workload-Standortes entsprechend 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 Anforderungen der Benutzer 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 Routing-Regeln 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:

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

Merkmale	Tools und Metriken
Grundlegende Netzwerkmerkmale	<ul style="list-style-type: none"> <li>• <a href="#">VPC Flow Logs</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="#">Amazon API Gateway-Metriken</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="#">AWS 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>

#### 4. Benchmarks für die Netzwerkleistung festlegen und testen:

- a. [Benchmark](#)- Netzwerkdurchsatz, da einige Faktoren die Amazon EC2-Netzwerkleistung beeinflussen können, wenn sich Instances in derselben VPC befinden. Messen Sie die Netzwerkbandbreite zwischen Amazon EC2-Linux-Instances in der gleichen VPC.
- b. 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](#)
- [Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter \(ENA\) in 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 zu Cloud-Netzwerken](#)
- [AWS re:Invent 2019 – Konnektivität mit AWS und hybriden AWS-Netzwerkarchitekturen](#)
- [AWS re:Invent 2019 – Optimieren der Netzwerkleistung für 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 Netzwerkfehlkonfigurationen 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 Produkt-Features verfügen (wie das [Amazon S3 Transfer Acceleration](#) -Feature) zur Optimierung des Netzwerkverkehrs.

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 Netzwerke und Kernnetzwerke mit [Amazon CloudWatch Logs-Metriken](#). 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 Availability Zones sowie innerhalb jeder Availability Zone an, indem Sie [AWS Network Manager](#) verwenden, um Erkenntnisse bezüglich des Zusammenhangs zwischen der Leistung Ihrer Anwendung und der Leistung des zugrunde liegenden AWS-Netzwerks zu erhalten.
- Verwenden Sie ein vorhandenes Konfigurationsmanagementdatenbank-Tool (CMDB-Tool) oder einen Service wie [AWS Config](#) um eine Bestandsaufnahme Ihrer Workload und deren Konfiguration zu erstellen.
- Wenn es sich um einen bestehenden 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.

- Bei einer neuen Workload sollten Sie [Lasttests](#) durchführen, um Leistungsengpässe zu identifizieren.
- 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	Mithilfe des <a href="#">Network Access Analyzer</a> Pfade oder Routen identifizieren.
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>Nutzen 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="#">Routing auf der Grundlage der geografischen Nähe</a> und <a href="#">IP-basiertes Routing</a> und trägt damit zur Leistungsverbesserung der Workload für eine globale Zielgruppe bei. Ermitteln Sie, welche Routing-Option Ihre Workload-Leistung optimieren würde. Prüfen Sie dazu Ihren Workload-Datenverkehr und den Benutzerstandort bei der globalen Verteilung Ihrer Workload.</p>

Verbesserungsmöglichkeit	Lösung
Speicherressourcen-Features	<p><a href="#">Amazon S3 Transfer Acceleration</a> ist ein Feature, mit dessen Hilfe externe Benutzer beim Hochladen von Daten in Amazon S3 von den Netzwerkoptimierungen von CloudFront profitieren können. Dies erleichtert die Übertragung großer Datenmengen von Remote-Standorten ohne spezielle Konnektivität zur AWS Cloud.</p> <p><a href="#">Multi-Region-Zugriffspunkte in Amazon S3</a> replizieren Inhalte in mehreren Regionen und vereinfachen die Workload durch die Bereitstellung eines Zugriffspunkts. Bei Verwendung eines Multi-Region-Zugriffspunkts 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><a href="#">Elastic Network Interfaces (ENA, Elastic-Network-Schnittstellen)</a>, die von Amazon EC2-Instances, Containern und Lambda-Funktionen verwendet werden, sind pro Fluss begrenzt. Prüfen Sie Ihre Platzierungsgruppen, um Ihren <a href="#">EC2-Netzwerkdurchsatz zu optimieren</a>.. 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 ethtool -Befehl ist im ENA-Treiber enthalten und stellt zusätzliche netzwerkbezogene Metriken zur Verfügung, die als <a href="#">benutzerdefinierte Metriken</a> in CloudWatch veröffentlicht werden können.</p> <p><a href="#">Amazon Elastic Network Adapters (ENA)</a> ermöglichen eine weitere Optimierung, da sie einen besseren Durchsatz für Ihre Instances innerhalb einer <a href="#">Cluster-Placement-Gruppe</a> bieten.</p> <p><a href="#">Elastic Fabric Adapter (EFA)</a> ist eine Netzwerkschnittstelle für Amazon EC2-Instances, mit der Sie Workloads, die ein hohes Maß an Kommunikation zwischen Knoten erfordern, in AWS bedarfsgesteuert ausführen können.</p> <p><a href="#">Amazon EBS-optimierte Instances</a> verwenden einen optimierten Konfigurations-Stack und</p>

Verbesserungsmöglichkeit	Lösung
	stellen zusätzliche dedizierte Kapazität zur Erhöhung der Amazon EBS-I/O bereit.

## Ressourcen

### Zugehörige Dokumente:

- [Application Load Balancer](#)
- [EC2: Enhanced Networking unter Linux](#)
- [EC2: Enhanced Networking unter Windows](#)
- [EC2: Platzierungsgruppen](#)
- [Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter \(ENA\) in 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 – Optimieren der Netzwerkleistung für 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 Netzwerkfehlerkonfigurationen 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 Ihren hybriden 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 entsprechend den Bandbreitenanforderungen. [AWS 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 AWS Direct Connect darauf gewartet wird, dass eine resiliente physische Netzwerkkonnektivität hergestellt wird.

Wenn Ihre Bandbreitenanforderungen hoch sind, könnten Sie mehrere AWS 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 AWS Direct Connect und VPN.

## Implementierungsschritte

1. Schätzen Sie die Anforderungen an Bandbreite und Latenz für Ihre bestehenden Anwendungen ab.
  - a. Für bestehende Workloads, die auf AWS umgestellt werden, nutzen Sie die Daten aus Ihren internen Systemen zur Überwachung des Netzwerks.
  - b. 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.
2. Wählen Sie eine dedizierte Verbindung oder ein VPN als Konnektivitätsoption aus. Je nach den Anforderungen des Workloads (Verschlüsselung, Bandbreite und Traffic-Bedarf) können Sie entweder AWS Direct Connect oder [AWS VPN auswählen](#) (oder beides). Das folgende Diagramm kann Ihnen bei der Wahl der geeigneten Verbindungsart helfen.
  - a. [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 Ihr 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).
  - b. Das 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.
3. Folgen Sie der AWS-Dokumentation, um eine geeignete Verbindungsoption auszuwählen:
  - a. Wenn Sie sich für die Verwendung von AWS Direct Connect entscheiden, wählen Sie die entsprechende Bandbreite für Ihre Konnektivität aus.

- b. Wenn Sie ein AWS Site-to-Site VPN über mehrere Standorte hinweg nutzen, um eine Verbindung zu einer AWS-Region herzustellen, sollten Sie eine [beschleunigte Site-to-Site VPN-Verbindung verwenden](#), um die Netzwerkleistung verbessern zu können.
  - c. Wenn Ihr Netzwerkdesign aus einer IPSec-VPN-Verbindung über [AWS Direct Connect](#) besteht, sollten Sie erwägen, Private IP VPN zu verwenden, um die Sicherheit zu verbessern und eine Segmentierung zu erzielen. [AWS Site-to-Site Private IP VPN](#) wird auf der virtuellen Transitschnittstelle bereitgestellt.
  - d. [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](#), unter Umgehung von AWS-Regionen, gesendet werden.
4. 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.
  5. 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](#)
- [Erstellen einer skalierbaren und sicheren Multi-VPC-AWS-Netzwerkinfrastruktur](#)
- [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 – Optimierung der Leistung mit Amazon CloudFront](#)
- [AWS re:Invent 2019 – Konnektivität mit AWS und hybriden AWS-Netzwerkarchitekturen](#)
- [AWS re:Invent 2020 Connect – AWS Transit Gateway Connect](#)

Zugehörige Beispiele:

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

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

Verteilen Sie den Datenverkehr auf mehrere Ressourcen oder Services, um von der Elastizität der Cloud zu profitieren. 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 Ihres Workloads.
- Sie nutzen die Funktionen des Load Balancers nicht zur Optimierung der Leistung.
- Der 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, automatische Skalierung sowie eine bessere Nutzung für Ihre Workload.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: hoch

## Implementierungsleitfaden

Load Balancer fungieren als Eingangspunkt für Ihren 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 der Lastenverteilung bereit. [Application Load Balancer](#) eignet sich optimal für die Lastenverteilung von HTTP- und HTTPS-Datenverkehr. Sie profitieren hierbei von einem fortschrittlichen Routing von Anforderungen, die es Ihnen ermöglicht, moderne Anwendungsarchitekturen mit Microservices und Containern bereitzustellen.

[Network Load Balancer](#) eignet sich optimal für die Lastenverteilung 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 Balancers flexibel zentral verwalten und CPU-intensive Arbeitsschritte für Ihren 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 Ihres Workloads berücksichtigen. Wenn Sie beispielsweise eine latenzempfindliche Anwendung haben, können Sie sich für Network Load Balancer mit einer extrem niedrigen Latenz entscheiden. Alternativ können Sie Ihr Workload auch näher an Ihre Kunden heranbringen, indem Sie Application Load Balancer in [AWS Local Zones](#) oder sogar [AWS Outposts](#) einsetzen.

Eine weitere Überlegung für latenzempfindliche Workloads ist das zonenübergreifende Load-Balancing. 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 Balancers können für containerisierte Workloads außerdem mit [Amazon ECS](#) und [Amazon EKS](#) integriert werden.

- [Amazon ECS – Service-Lastenverteilung](#)
- [Anwendungslastenverteilung auf Amazon EKS](#)
- [Netzwerklastenverteilung auf Amazon EKS](#)

## Implementierungsschritte

- Definieren Sie Ihre Anforderungen an die Lastenverteilung, einschließlich Datenverkehrsvolumen, Verfügbarkeit und Anwendungsskalierbarkeit.
- Wählen Sie den richtigen Load Balancer-Typ für Ihre Anwendung.
  - Verwenden Sie Application Load Balancer für HTTP/HTTPS Workloads.
  - Verwenden Sie Network Load Balancer für Nicht-HTTP-Workloads, die TCP oder UDP nutzen.
  - Verwenden Sie eine Kombination aus beiden ([ALB als Ziel von NLB](#)) aus, wenn Sie die Funktionen 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 Ihren HTTP-Workload mit [AWS PrivateLink](#) teilen möchten.

- Einen vollständigen Vergleich von Load Balancern finden Sie im [ELB-Produktvergleich](#).
- Verwenden Sie nach Möglichkeit SSL/TLS-Offloading.
- Konfigurieren Sie HTTPS/TLS-Listener, bei denen [Application Load Balancer](#) und [Network Load Balancer](#) mit [AWS Certificate Manager](#) integriert sind.
- Beachten Sie, dass einige Workloads aus Compliance-Gründen 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 zum Beispiel [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 Domänen mit Zonenfehlern zu vermeiden. In NLB ist dies standardmäßig deaktiviert. In [ALB können Sie die Option pro Gruppe](#) deaktivieren.
  - Verwenden Sie die aktivierte zonenübergreifende Verarbeitung für eine höhere Verfügbarkeit und Flexibilität. Standardmäßig ist die zonenübergreifende Verarbeitung für ALB aktiviert. In [NLB können Sie sie pro Gruppe](#) 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`, `target_processing_time` und `response_processing_time`.
- Die wichtigsten zu berücksichtigenden Elemente für ALB sind `connection_time` und `tls_handshake_time`.
- Bereiten Sie sich darauf vor, die Protokolle bei Bedarf abfragen zu können. Sie können Amazon Athena verwenden, um sowohl [ALB-Protokolle](#) als auch [NLB-Protokolle](#) abzufragen.
- Erstellen Sie Alarme für leistungsbezogene Metriken wie [TargetResponseTime für ALB](#).

## Ressourcen

### Zugehörige Dokumente:

- [ELB-Produktvergleich](#)
- [Globale AWS-Infrastruktur](#)
- [Verbesserung der Leistung und Senkung der Kosten durch Availability Zone-Affinität](#)
- [Schritt für Schritt zur Protokollanalyse mit Amazon Athena](#)
- [Abfragen von Application Load Balancer-Protokollen](#)
- [Überwachen Ihrer Application Load Balancers](#)
- [Überwachen Ihrer Network Load Balancer](#)
- [Verwenden Sie Elastic Load Balancing, um den Datenverkehr über die Instances in Ihrer Auto Scaling-Gruppe zu verteilen.](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Welche Vorteile bietet die Vernetzung für Ihre Anwendung?](#)
- [AWS re:Inforce 20 – So verbessern Sie mit Elastic Load Balancing Ihren Sicherheitsstatus im großen Umfang](#)
- [AWS re:Invent 2018 – Elastic Load Balancing: Ein ausführliche Beschreibung und bewährte Methoden](#)
- [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 und AWS CloudFormation-Beispiele für die Protokollanalyse mit 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 Ihre Dateiübertragung 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 Ihres Workloads 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 des Workloads 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 eine große Menge an Daten zwischen den Knoten übertragen werden muss.

## Implementierungsschritte

1. Verwenden Sie die [AWS Global Accelerator](#)- und [AWS Transfer Family](#)- Services, um den Durchsatz Ihrer Anwendungen für die Onlineübertragung von Dateien zu verbessern. Der AWS Global Accelerator-Service hilft Ihnen, die Latenz zwischen Ihren Client-Geräten und Ihrem 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.
2. 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 Time (RTT) können zur Messung der Netzwerklatenz verwendet werden. Wenn Ihr 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.
3. 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 Ihren Workload führen. Ziehen Sie für solche Workloads das SSL/TLS-Offloading auf [Elastic Load Balancing](#) in Betracht, um die Leistung des Workloads 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.
4. 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 Ihres Workloads zu verbessern. Der NLB verteilt den eingehenden UDP-Datenverkehr auf mehrere Ziele, sodass Sie Ihren Workload horizontal skalieren, die Kapazität erhöhen und den Overhead eines einzelnen Ziels reduzieren können.

5. Für Ihre HPC-Workloads (High Performance Computing) sollten Sie die [Elastic Network Adapter \(ENA\) Express-Funktionalität](#) 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 (25 Perzentil) für den Netzwerkverkehr zwischen EC2-Instances bietet.
6. 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](#)
- [Aktivieren von Enhanced Networking-Funktionen mit dem Elastic Network Adapter \(ENA\) in Linux-Instances](#)
- [Network Load Balancer](#)
- [Netzwerkprodukte mit AWS](#)
- [Umstellung auf latenzbasiertes Routing in Amazon Route 53](#)
- [VPC-Endpunkte](#)

### Zugehörige Videos:

- [AWS re:Invent 2022 – Skalierung der Netzwerkleistung auf Amazon Elastic Compute Cloud-Instances der nächsten Generation](#)
- [AWS re:Invent 2022 – Grundlagen der Anwendungsnetzwerke](#)

Zugehörige Beispiele:

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

## PERF04-BP06 Auswählen des Workload-Standortes entsprechend 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: Das Benutzererlebnis wird stark von der Latenz zwischen dem/der Benutzer:in und Ihrer Anwendung beeinflusst. Durch die Verwendung geeigneter AWS-Regionen und des privaten globalen AWS-Netzwerks können Sie die Latenz reduzieren und Remote-Benutzern ein besseres Erlebnis bieten.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: mittel

### Implementierungsleitfaden

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

Amazon EC2 verfügt über Platzierungsgruppen für das Netzwerk. Eine Platzierungsgruppe ist eine logische Gruppierung von Instances, um die Latenz zu verringern. Die Verwendung von

Platzierungsgruppen mit unterstützten Instance-Typen und einem Elastic Network Adapter (ENA) ermöglicht die Verarbeitung von Workloads in einem Netzwerk mit 25 Gbit/s, reduziertem Jitter und geringer Latenz. 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 globales AWS-Netzwerk bereitgestellt, z. B. [Amazon CloudFront](#). Diese Edge-Standorte verfügen in der Regel über Services wie ein Content Delivery Network (CDN) und Domain Name System (DNS). Durch die Platzierung am Edge können die Workloads mit geringer Latenz auf Anforderungen zu Inhalten oder zur DNS-Auflösung 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 Cache-Steuerung für DNS und HTTP/HTTPS richtig, um aus diesen Ansätzen den größtmöglichen Nutzen zu ziehen.

## Implementierungsschritte

- Erfassen Sie Informationen über den an den Netzwerkschnittstellen ein- und ausgehenden IP-Datenverkehr.
  - [Protokollierung von IP-Datenverkehr mithilfe von VPC Flow Logs](#)
  - [Wie Sie die Client-IP-Adresse in AWS Global Accelerator beibehalten](#)
- Analysieren Sie die Netzwerkzugriffsmuster in Ihrem Workload, um zu ermitteln, wie die Benutzer Ihre Anwendung verwenden.
  - Verwenden Sie Überwachungstools wie [Amazon CloudWatch](#) und [AWS CloudTrail](#), um Daten zu Netzwerkaktivitäten zu erfassen.
  - Analysen Sie die Daten, um das Netzwerkzugriffsmuster zu identifizieren.
- Wählen Sie Regionen für Ihre Workload-Bereitstellung auf der Grundlage der folgenden zentralen Elemente aus:
  - Dem Speicherort Ihrer Daten: Für datenintensive Anwendungen (wie Big Data oder Machine Learning) sollte der Anwendungscode so nahe wie möglich zu den Daten ausgeführt werden.
  - Den Standorten Ihrer Benutzer:innen: Wählen Sie bei nutzerorientierten Anwendungen eine Region (oder Regionen) möglichst nahe an den Benutzer:innen Ihres Workloads.
  - Anderen einschränkenden Faktoren: Denken Sie dabei etwa an Kosten und Compliance, wie in [Überlegungen bei der Auswahl einer Region für Ihren Workload](#) beschrieben.

- Verwenden Sie [AWS Local Zones](#) für Workloads wie Video-Rending. 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.
- Anwendungen wie hochauflösendes Live-Video-Streaming, High-Fidelity-Audio und Augmented Reality oder Virtual Reality (AR/VR) erfordern extrem niedrige Latenzen für 5G-Geräte. Ziehen Sie für solche Anwendungen [AWS Wavelength](#) in Betracht. AWS Wavelength bettet AWS-Computing- und Speicher-Services in 5G-Netzwerke ein und bietet eine mobile Edge-Computing-Infrastruktur für die Entwicklung, Bereitstellung und Skalierung von Anwendungen mit extrem niedriger Latenz.
- Verwenden Sie lokale Zwischenspeicherung oder [AWS-Zwischenspeicherungslösungen](#) für häufig genutzte Ressourcen zur Verbesserung der Leistung, zur Verringerung von Datenverschiebungen und zur Reduzierung der Umweltauswirkungen.

Service	When to use
<a href="#">Amazon CloudFront</a>	Verwenden Sie dies für die Zwischenspeicherung statischer Inhalte wie Bilder, Skripts und Videos sowie dynamischer Inhalte wie API-Antworten oder Webanwendungen.
<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 Nutzern Ihres Workloads auszuführen:

Service	When to use
<a href="#">Lambda@edge</a>	Verwenden Sie dies für rechenintensive Anwendungen, die initiiert werden, wenn sich Objekte nicht im Zwischenspeicher befinden.

Service	When to use
<a href="#">Amazon CloudFront-Funktionen</a>	Verwenden Sie diese für einfache Anwendungsfälle wie HTTP(s)-Anfragen oder Antwortmanipulationen, die von kurzlebigen Funktionen initiiert 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 Netzwerk-Services profitieren, die statische Anycast-IP-Adressen und eine TCP-Terminierung an Edge-Standorten bieten. [AWS Global Accelerator](#) kann die Leistung Ihrer Anwendungen um bis zu 60 % verbessern und bietet ein schnelles Failover für Architekturen mit mehreren Regionen. AWS Global Accelerator stellt Ihnen statische Anycast-IP-Adressen zur Verfügung, die als fester Zugangspunkt für Ihre Anwendungen dienen, die in einer oder mehreren AWS-Regionen gehostet werden. Diese IP-Adressen sorgen dafür, dass Datenverkehr so nah wie möglich an Ihren Benutzern in das globale AWS-Netzwerk eingebunden wird. 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 liegt. Prüfen Sie die Verwendung von AWS Global Accelerator, um die Leistung Ihrer TCP/UDP-Workloads zu verbessern und einen schnellen Failover für Architekturen mit mehreren Regionen zu ermöglichen.

## Ressourcen

Zugehörige bewährte Methoden:

- [COST07-BP02 Implementieren von Regionen auf Basis der Kosten](#)
- [COST08-BP03 Implementieren von Services zur Senkung der Datenübertragungskosten](#)
- [REL10-BP01 Bereitstellen des Workloads an mehreren Standorten](#)
- [REL10-BP02 Auswählen der geeigneten Standorte für Ihre Multi-Standort-Bereitstellung](#)
- [SUS01-BP01 Auswählen der Region auf Grundlage von Unternehmensanforderungen und Nachhaltigkeitszielen](#)

- [SUS02-BP04 Optimieren der geografischen Platzierung von Workloads auf der Grundlage ihrer Netzwerkanforderungen](#)
- [SUS04-BP07 Minimieren von Datenübertragungen zwischen Netzwerken](#)

#### Zugehörige Dokumente:

- [Globale AWS-Infrastruktur](#)
- [AWS Local Zones und AWS Outposts Outposts: Die 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:

- [Erklärungsvideo zu AWS Local Zones](#)
- [AWS Outposts: Übersicht und Funktionsweise](#)
- [AWS re:Invent 2023 – Eine Migrationsstrategie für Edge- und On-Premises-Workloads](#)
- [AWS re:Invent 2021 – AWS Outposts: Das AWS AWS Erlebnis on-premises](#)
- [AWS re:Invent 2020: AWS Wavelength: Apps mit ultraniedriger Latenz am 5G-Edge ausführen](#)
- [AWS re:Invent 2022 – AWS Local Zones: Entwickeln von Anwendungen für einen verteilten Edge](#)
- [AWS re:Invent 2021 – Entwicklung von Websites mit niedriger Latenz mit Amazon CloudFront](#)
- [AWS re:Invent 2022 – Verbessern der Leistung und Verfügbarkeit mit AWS Global Accelerator](#)
- [AWS re:Invent 2022 – Aufbau Ihres globalen Wide Area Networks mit AWS](#)
- [AWS re:Invent 2020: Globales Datenverkehrsmanagement mit Amazon Route 53](#)

Zugehörige Beispiele:

- [AWS Global Accelerator Workshop für benutzerdefiniertes 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 Netzwerk-Services.
- 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 Messwerte 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 Logs</a>	Nutzen Sie VPC Flow Logs, um detaillierte Informationen über den Datenverkehr zu und von den Netzwerkschnittstellen in Ihren VPCs zu protokollieren. Mit VPC Flow Logs können Sie restriktive oder freizügige Regeln für Sicherheitsgruppen diagnostizieren und die Richtung des Datenverkehrs zu und von den Netzwerkschnittstellen ermitteln.
<a href="#">AWS Transit Gateway Flow Logs</a>	Nutzen Sie AWS Transit Gateway Flow Logs, um Informationen über den IP-Datenverkehr zu und von Ihren Transit-Gateways zu erfassen.

AWS-Tool	Aktion
<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 Domäne oder Subdomäne bzw. die Route 53-Edge-Standorte, die auf DNS-Abfragen geantwortet haben, nachvollziehen.
<a href="#">Reachability Analyzer</a>	Reachability Analyzer hilft Ihnen, die Erreichbarkeit des Netzwerks zu analysieren und zu debuggen. Reachability Analyzer ist ein Konfigurationsanalyse-Tool, mit dem Sie die Konnektivität zwischen einer Quelle und einer Zielressource in Ihren VPCs testen können. Mit diesem Tool können Sie überprüfen, ob Ihre Netzwerkkonfiguration der geplanten Konnektivität entspricht.
<a href="#">Network Access Analyzer</a>	Network Access Analyzer hilft Ihnen, den Netzwerkzugriff auf Ihre Ressourcen zu verstehen. 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 und belegen, dass Ihr AWS-Netzwerk Ihre Compliance-Anforderungen erfüllt.

AWS-Tool	Aktion
<a href="#">Amazon CloudWatch</a>	<p>Nutzen Sie <a href="#">Amazon CloudWatch</a> und aktivieren Sie geeignete Metriken für Netzwerkooptionen. Stellen Sie sicher, dass Sie die richtige Netzwerk-Metrik für Ihren 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>Mithilfe von AWS Network Manager können Sie die Echtzeit- und vergangene Leistung des <a href="#">AWS Global Network</a> für betriebliche und planerische Zwecke 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.
- Beurteilen 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.
- Untersuchen 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?](#)
- [Was ist Reachability Analyzer?](#)
- [Was ist Network Access Analyzer?](#)
- [CloudWatch-Metriken für Ihre VPCs](#)
- [Optimize performance and reduce costs for network analytics with VPC Flow Logs in Apache Parquet format \(Optimieren der Leistung und Reduzieren der Kosten für die Netzwerk-Analytik mit VPC Flow Logs im Apache Parquet-Format\)](#)
- [Monitoring your global and core networks with Amazon CloudWatch metrics \(Überwachen von globalen und Kernnetzwerken mit Amazon CloudWatch-Metriken\)](#)
- [Continuously monitor network traffic and resources \(Kontinuierliches Überwachen von Netzwerkdatenverkehr und -ressourcen\)](#)

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

- [AWS re:Invent 2020 – Überwachen des Netzwerkdatenverkehrs und Fehlerbehebung](#)

Zugehörige Beispiele:

- [Workshops zu AWS-Netzwerken](#)
- [Überwachung des AWS-Netzwerks](#)
- [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 Praktiken, die Sie übernehmen können, um effiziente und leistungsstarke Cloud-Workloads besser zu betreiben. Dieser Schwerpunktbereich bietet bewährte Methoden zur Schaffung 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 Verfahren 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 Metriken und deren Überwachung 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 Kennzahlen wie den aggregierten kumulativen Kosten pro Anfrage können Sie Möglichkeiten zur Kostensenkung ermitteln. Erwägen Sie sorgfältig, wie Kennzahlen 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 gestartet 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 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-Lösungen (SaaS) verwenden. Erwägen Sie die Verwendung von [AWS Marketplace](#) -Lösungen und [Spot Instances](#). Sie können zu einer kostengünstigen Lastgenerierung beitragen.
- **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 Überprüfungsvorgang:** 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-Workloads.

#### Bewährte Methoden

- [PERF05-BP01 Festlegen wichtiger Leistungskennzahlen \(KPIs\) zum Messen des Zustands und der Leistung des Workloads](#)
- [PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren](#)
- [PERF05-BP03 Definieren eines Prozesses zum Verbessern der Workload-Leistung](#)
- [PERF05-BP04 Durchführen von Lasttests für den Workload](#)
- [PERF05-BP05 Verwenden von Automatisierung zur proaktiven Behebung leistungsbezogener Probleme](#)
- [PERF05-BP06 Konstantes Aktualisieren des Workloads und der Services](#)
- [PERF05-BP07 Regelmäßiges Überprüfen von Metriken](#)

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

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

Typische Anti-Muster:

- Sie überwachen nur Metriken auf Systemebene, um Erkenntnisse über Ihren 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 des Workloads widerspiegeln, hilft Teams dabei, ihre Prioritäten auszurichten und erfolgreiche Geschäftsergebnisse zu definieren. Das Teilen dieser Metriken mit allen Abteilungen bietet Sichtbarkeit und die Ausrichtung an Grenzwerten, Erwartungen und Geschäftsauswirkungen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: 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 ein 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

- **Stakeholder identifizieren:** Identifizieren und dokumentieren Sie wichtige Stakeholder im Unternehmen, einschließlich Entwicklungs- und Betriebsteams.
- **Ziele definieren:** Arbeiten Sie mit diesen Stakeholdern zusammen, um die Ziele Ihres Workloads zu definieren und zu dokumentieren. Berücksichtigen Sie die kritischen Leistungsaspekte Ihrer Workloads, wie Durchsatz, Reaktionszeit und Kosten, sowie Geschäftsziele wie die Benutzerzufriedenheit.
- **Bewährte Methoden der Branche überprüfen:** Sehen Sie sich in der Branche bewährte Methoden an, um relevante KPIs zu identifizieren, die auf Ihre Workload-Ziele abgestimmt sind.
- **Metriken identifizieren:** Identifizieren Sie Metriken, die mit Ihren Workload-Zielen übereinstimmen 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:innen.
- **KPIs definieren und dokumentieren:** 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.
- **Überwachung implementieren:** Verwenden Sie Überwachungstools wie [Amazon CloudWatch](#) oder [AWS Config](#), um Metriken zu erfassen und KPIs zu messen.
- **KPIs visuell kommunizieren:** Verwenden Sie Dashboard-Tools wie [Amazon QuickSight](#), um KPIs anzuzeigen und Stakeholdern mitzuteilen.
- **Analysieren und optimieren:** Überprüfen und analysieren Sie regelmäßig KPIs, um Bereiche Ihres Workloads zu identifizieren, die verbessert werden müssen. Arbeiten Sie mit den Stakeholdern zusammen, um diese Verbesserungen umzusetzen.

- Wiederaufgreifen und nachbessern: Überprüfen Sie regelmäßig 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](#)
- [Die Bedeutung von Key Performance Indicators \(KPIs\) für groß angelegte Cloud-Migrationen](#)
- [Wie Sie mit dem KPI-Dashboard Ihre KPIs zur Kostenoptimierung nachverfolgen](#)
- [X-Ray-Dokumentation](#)
- [Verwendung von Amazon CloudWatch-Dashboards](#)
- [Amazon QuickSight-KPIs](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Kosten- und Leistungsoptimierung sowie Fortschrittsverfolgung bei der Schadensbegrenzung](#)
- [AWS re:Invent 2023 – Verwaltung von Ereignissen im Ressourcenlebenszyklus im großen Maßstab mit AWS Health](#)
- [AWS re:Invent 2023 – Leistung und Effizienz bei Pinterest: Optimierung der neuer 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-Beobachtbarkeit und -Überwachung von Anwendungen mit AWS](#)
- [AWS re:Invent 2023 – Skalierung in AWS für die ersten 10 Millionen Benutzer:innen](#)
- [AWS re:Invent 2022 – Wie Amazon bessere Metriken für eine höhere Website-Leistung verwendet](#)
- [Erstellung einer effektiven Metrikenstrategie für Ihr Unternehmen | AWS-Ereignisse](#)

### Zugehörige Beispiele:

- [Erstellen eines Dashboards mit Amazon QuickSight](#)

## 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, wenn diese bewährte Methode nicht eingeführt wird: 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	Where to use
<a href="#">Amazon CloudWatch Real-User Monitoring (RUM)</a>	To capture application performance metrics from real user client-side and frontend sessions.
<a href="#">AWS X-Ray</a>	To trace traffic through the application layers and identify latency between components and dependencies. Use X-Ray service maps to see relationships and latency between workload components.
<a href="#">Amazon Relational Database Service Performance Insights</a>	To view database performance metrics and identify performance improvements.
<a href="#">Amazon RDS Erweiterte Überwachung</a>	To view database OS performance metrics.
<a href="#">Amazon DevOps Guru</a>	To detect abnormal operating patterns so you can identify operational issues before they impact your customers.

- 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](#) ein, um browserbasierte Benutzeraktivitäten programmgesteuert mit Linux-Cron-Aufträgen oder Ratenausdrücken nachzuahmen und im Zeitverlauf konsistente Metriken zu erhalten.
  - Verwenden Sie die Lösung [AWS Distributed Load Testing](#), 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 Ihren Workload zu testen.

## Ressourcen

### Zugehörige Dokumente:

- [Neuheiten im Bereich AWS-Beobachtbarkeit bei der re:Invent 2023](#)
- [Amazon Builders' Library](#)
- [X-Ray-Dokumentation](#)
- [Amazon CloudWatch RUM](#)
- [Amazon DevOps Guru](#)

### 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 Beobachtbarkeitsstrategie](#)
- [AWS Summit SF 2022 – Full-Stack-Beobachtbarkeit und -Überwachung von Anwendungen mit 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 – Wie Amazon bessere Metriken für eine höhere Website-Leistung verwendet](#)
- [Visuelle Überwachung von Anwendungen mit Amazon CloudWatch Synthetics](#)

### Zugehörige Beispiele:

- [Messen der Seitenladezeit mit Amazon CloudWatch Synthetics](#)
- [Amazon CloudWatch RUM Web Client](#)
- [X-Ray SDK for Python](#)
- [Verteilte Lasttests in 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 können Sie die gesammelten Daten langfristig in die Gestaltung Ihres Workloads einfließen lassen.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: mittel

### Implementierungsleitfaden

Für Ihren Workload gibt es einige wesentliche Einschränkungen. Dokumentieren Sie diese, damit Sie besser einschätzen können, durch welche Art von Innovation die Leistung Ihres Workloads 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 Ihren 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

- KPIs identifizieren: Identifizieren Sie Ihre Workload-Leistungs-KPIs, wie unter [PERF05-BP01 Festlegen wichtiger Leistungskennzahlen \(KPIs\) zum Messen des Zustands und der Leistung des Workloads](#) beschrieben, um eine Baseline für Ihren Workload zu erstellen.

- Überwachung implementieren: Verwenden Sie [AWS-Tools zur Beobachtbarkeit](#), um Leistungsmetriken zu erfassen und KPIs zu messen.
- Analysen durchführen: Führen Sie eine eingehende Analyse durch, um leistungsschwache Bereiche (wie Konfiguration und Anwendungscode) in Ihrem Workload zu identifizieren, wie beschrieben unter [PERF05-BP02 Verwenden von Überwachungslösungen, um Bereiche mit kritischem Leistungsbedarf zu identifizieren](#). Verwenden Sie Analyse- und Leistungs-Tools, um die Strategie zur Leistungsverbesserung zu identifizieren.
- Verbesserungen bestätigen: Verwenden Sie Sandbox- oder Vorproduktionsumgebungen, um die Wirksamkeit von Verbesserungsstrategien zu überprüfen.
- Änderungen implementieren: Implementieren Sie die Änderungen in der Produktion und überwachen Sie kontinuierlich die Leistung des Workloads. Dokumentieren Sie die Verbesserungen und teilen Sie die Änderungen den Stakeholdern mit.
- Wiederaufgreifen und Verfeinern: Überprüfen Sie regelmäßig Ihren Leistungsverbesserungsprozess, um Verbesserungsmöglichkeiten 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 – Kosten- und Leistungsoptimierung sowie Fortschrittsverfolgung bei der Schadensbegrenzung](#)
- [AWS re:Invent 2022 – AWS-Optimierung: Umsetzbare Schritte für sofortige Ergebnisse](#)
- [AWS re:Invent 2022 — Optimierung Ihrer AWS-Workloads mit Anleitungen für bewährte Methoden](#)

### Zugehörige Beispiele:

- [AWS Github](#)

## PERF05-BP04 Durchführen von Lasttests für den Workload

Führen Sie für den Workload Lasttests durch, um sicherzustellen, dass er 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 Lasttests durch, ohne die [Amazon EC2-Testrichtlinien](#) zu lesen oder ein Formular für die Einreichung simulierter Ereignisse abzusenden. 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, wenn diese bewährte Methode nicht eingeführt wird: niedrig

### Implementierungsleitfaden

Lasttests in der Cloud sind ein Prozess zur Messung der Leistung eines Cloud-Workloads 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 Ihres Workloads zu bewerten, mit einer realistischen Last umzugehen. Verwenden Sie für Lasttests synthetische oder bereinigte Daten und entfernen Sie sensible oder personenbezogene Informationen. Führen Sie automatisch Lasttests als Teil Ihrer Bereitstellungs-Pipeline durch und vergleichen Sie die Ergebnisse mit vordefinierten KPIs und Schwellenwerten. Dieser Prozess hilft Ihnen dabei, die erforderliche Leistung weiterhin zu erreichen.

## Implementierungsschritte

- **Testziele definieren:** Identifizieren Sie die Leistungsaspekte Ihres Workloads, das Sie bewerten möchten, z. B. Durchsatz und Reaktionszeit.
- **Testtool auswählen:** Wählen und konfigurieren Sie das Lasttest-Tool, das zu Ihrem Workload passt.
- **Umgebung einrichten:** Richten Sie die Testumgebung basierend auf Ihrer Produktionsumgebung ein. Mithilfe von AWS-Services können Sie Umgebungen im Produktionsmaßstab ausführen und damit Ihre Architektur testen.
- **Überwachung implementieren:** Verwenden Sie Überwachungstools wie Amazon CloudWatch, um Metriken für alle Ressourcen in Ihrer Architektur zu erfassen. Sie können auch benutzerdefinierte Metriken erfassen und veröffentlichen.
- **Szenarien definieren:** Definieren Sie die Szenarien und Parameter der Lasttests (wie Testdauer und Anzahl der Benutzer:innen).
- **Lasttests durchführen:** Führen Sie Testszenarien in großem Maßstab durch. Testen Sie Ihren Workload mithilfe der AWS Cloud, um zu ermitteln, an welcher Stelle er nicht skalierbar ist oder ob die Skalierung nichtlinear erfolgt. 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 Verbesserungsmöglichkeiten zu identifizieren.
- **Erkenntnisse dokumentieren und teilen:** Erkenntnisse und Empfehlungen dokumentieren und kommunizieren. Teilen Sie diese Informationen mit Stakeholdern, um ihnen zu helfen, fundierte Entscheidungen über Strategien zur Leistungsoptimierung zu treffen.
- **Kontinuierlich iterieren:** Lasttests sollten in regelmäßigen Abständen durchgeführt werden, insbesondere nach einem Systemwechsel oder Update.

## Ressourcen

Zugehörige Dokumente:

- [Amazon CloudWatch RUM](#)
- [Amazon CloudWatch Synthetics](#)
- [Verteilte Lasttests in AWS](#)

### Zugehörige Videos:

- [AWS Summit ANZ 2023: Mit AWS Distributed Load Testing zuversichtlich in die Zukunft starten](#)
- [AWS re:Invent 2022: Skalierung in Scaling AWS für Ihre ersten 10 Millionen Benutzer:innen](#)
- [Lösen mit AWS-Solutions: Verteilte Lasttests](#)
- [AWS re:Invent 2021 – Optimierung von Anwendungen durch Endbenutzereinsichten mit Amazon CloudWatch RUM](#)
- [Demo von Amazon CloudWatch Synthetics](#)

### Zugehörige Beispiele:

- [Verteilte Lasttests in 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 Betriebsteam 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, wenn diese bewährte Methode nicht eingeführt wird: 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 Ihr 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

- Mängelbeseitigungsworkflow identifizieren: Identifizieren und verstehen Sie das Leistungsproblem, das automatisch behoben werden kann. Verwenden Sie AWS-Überwachungslösungen wie [Amazon CloudWatch](#) oder AWS X-Ray, um die Ursache des Problems besser zu verstehen.
- Automatisierungsprozess definieren: Erstellen Sie einen schrittweisen Prozess zur Mängelbeseitigung, mit dem das Problem automatisch behoben werden kann.
- Initiationsereignis konfigurieren: 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.
- Mängelbeseitigung automatisieren: Verwenden Sie AWS-Services und Technologien, um den Mängelbeseitigungsprozess zu automatisieren. [AWS Systems Manager Automation](#) bietet beispielsweise 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.
- Workflow testen: Testen Sie den automatisierten Prozess zur Mängelbeseitigung in einer Vorproduktionsumgebung.
- Workflow implementieren: Implementieren Sie die automatisierten Prozess zur Mängelbeseitigung in der Produktionsumgebung.
- Playbook entwickeln: 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: Beurteilen Sie regelmäßig die Effektivität des automatisierten Mängelbeseitigungs-Workflows. 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](#)
- [Verwendung von Alarmen und Alarmaktionen in CloudWatch](#)
- [Aufbau einer Cloud-Automatisierungspraxis für Operational Excellence: Bewährte Methoden von AWS Managed Services](#)
- [Automatisieren Ihrer Amazon Redshift-Leistungsoptimierung mit automatischer Tabellenoptimierung](#)

### Zugehörige Videos:

- [AWS re:Invent 2023 – Strategien für die automatisierte Skalierung, Mängelbeseitigung und intelligente Selbstreparatur](#)
- [AWS re:Invent 2023 – \[LAUNCH\] Anwendungsüberwachung für moderne Workloads](#)
- [AWS re:Invent 2023 – Implementierung der Anwendungsbeobachtbarkeit](#)
- [AWS re:Invent 2021 – Intelligente Automatisierung des Cloud-Betriebs](#)
- [AWS re:Invent 2022 – Einrichtung skalierbarer Kontrollen in Ihrer AWS-Umgebung](#)
- [AWS re:Invent 2022 – Automatisierung der Patch-Verwaltung und -Compliance mit AWS](#)
- [AWS re:Invent 2022 – Wie Amazon bessere Metriken für eine höhere Website-Leistung verwendet](#)
- [AWS re:Invent 2023 – Ballast abwerfen: Diagnose und Lösung von Leistungsproblemen mit Amazon RDS](#)
- [AWS re:Invent 2021 – {Neuer Launch} Automatische Erkennung und Behebung von Problemen mit Amazon DevOps Guru](#)

- [AWS re:Invent 2023 – Zentralisierung Ihrer Abläufe](#)

Zugehörige Beispiele:

- [CloudWatch Logs Konfigurieren von Alarmen](#)

## PERF05-BP06 Konstantes Aktualisieren des Workloads und der Services

Erhalten Sie aktuelle Informationen zu neuen Cloud-Services und -Funktionen, um effiziente Funktionen zu übernehmen, Probleme zu beseitigen und die allgemeine Leistungseffizienz des 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 Ihrem Workload kompatibel sind.

Vorteile der Nutzung dieser bewährten Methode: Wenn Sie einen Prozess einrichten, um aktuelle Informationen zu neuen Services und Angeboten zu erhalten, können Sie neue Funktionen und Kapazitäten nutzen, Probleme lösen und die Workload-Leistung verbessern.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: 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 Funktionen und Services, die für Ihren 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 des Workloads zu messen.

## Implementierungsschritte

- Inventarisierung Ihrer Workload: Inventarisieren Sie Ihre Workload-Software und -Architektur und identifizieren Sie Komponenten, die aktualisiert werden müssen.
- Identifizierung Ihrer Aktualisierungsquellen: Identifizieren Sie Nachrichten und Aktualisierungsquellen im Zusammenhang mit Ihren Workload-Komponenten. Sie können beispielsweise den [Neuigkeiten im AWS-Blog](#) für die Produkte abonnieren, 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, um neue Services und Features für Ihr Workload zu bewerten.
  - Sie können [AWS Systems Manager Inventory](#) verwenden, um Betriebssystem (BS)-, Anwendungs- und Instance-Metadaten von Ihren Amazon EC2-Instances zu sammeln und so schnell zu erfassen, welche Instances die Software und die Konfigurationen ausführen, die Ihre Softwarerichtlinie erfordert, und welche Instances aktualisiert werden müssen.
- Bewertung der neuen Aktualisierung: Erfahren Sie, wie die Komponenten Ihres Workloads aktualisiert werden. Nutzen Sie die Agilität in der Cloud, um schnell zu testen, wie neue Features Ihr 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, um AMIs, 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 Bewertung von Updates und neuen Services. Geben Sie Ihren Eigentümern ausreichend Zeit und Raum zum Forschen, Testen, Experimentieren und zur Validierung von Aktualisierungen und neuen Services. Nutzen Sie die dokumentierten geschäftlichen Anforderungen und KPIs, um zu ermitteln, welche Aktualisierungen positive geschäftliche Auswirkungen haben werden.

## Ressourcen

Zugehörige Dokumente:

- [AWS-Blog](#)
- [Neuerungen bei AWS](#)
- [Implementierung aktueller Images mit automatisierten EC2 Image Builder Pipelines](#)

Zugehörige Videos:

- [AWS re:Inforce 2022 – Automatisierung der Patch-Verwaltung und -Compliance mit AWS](#)
- [All Things Patch: AWS Systems Manager | AWS-Veranstaltungen](#)

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 Kennzahlen, 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 sicherzustellen, dass sie Probleme ordnungsgemäß identifizieren, beheben oder verhindern. Metriken können auch veralten, wenn sie für einen längeren Zeitraum im Alarmstatus bleiben.

Risikostufe, wenn diese bewährte Methode nicht eingeführt wird: mittel

## Implementierungsleitfaden

Verbessern Sie kontinuierlich die Erfassung und Überwachung von Metriken. Bewerten Sie beim Reagieren auf Vorfälle oder Ereignisse diejenigen Kennzahlen, die hilfreich für die Behebung des Problems waren, und überlegen Sie, welche derzeit noch nicht verfolgten Kennzahlen 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 Kennzahlen, die hilfreich für die Behebung des Problems waren, und überlegen Sie, welche derzeit noch nicht verfolgten Kennzahlen 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 Metriken 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](#)
- [Erfassen von Metriken und Protokollen aus Amazon EC2-Instances und On-Premises-Servern mit dem CloudWatch Agent](#)
- [Metrikabfrage mit CloudWatch Metrics Insights](#)
- [Überwachung, Protokollierung und Leistung von AWS Partner Network-Partnern](#)
- [X-Ray-Dokumentation](#)

### Zugehörige Videos:

- [AWS re:Invent 2022 – Einrichtung skalierbarer Kontrollen in Ihrer AWS-Umgebung](#)
- [AWS re:Invent 2022 – Wie Amazon bessere Metriken für eine höhere Website-Leistung verwendet](#)
- [AWS re:Invent 2023 – Aufbau einer effektiven Beobachtbarkeitsstrategie](#)
- [AWS Summit SF 2022 – Full-Stack-Beobachtbarkeit und -Überwachung von Anwendungen mit AWS](#)
- [AWS re:Invent 2023 – Ballast abwerfen: Diagnose und Lösung von Leistungsproblemen mit Amazon RDS](#)

### Zugehörige Beispiele:

- [Erstellen eines Dashboards mit Amazon QuickSight](#)
- [CloudWatch-Dashboards](#)

## Fazit

Um eine hohe Leistungseffizienz zu erzielen und aufrechtzuerhalten, ist ein datengesteuerter Ansatz erforderlich. Zur Leistungsoptimierung empfiehlt es sich daher, die Verwendung entsprechender 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 unterstützt Sie beim Aufbau von Architekturen mit effizienter Leistung, die einen Mehrwert für Ihr Unternehmen schaffen. Verwenden Sie die in diesem Whitepaper beschriebenen Tools und Verfahren, damit Ihre Arbeit erfolgreich ist.

# Mitwirkende

Dieses Dokument ist unter der Mitarbeit folgender Personen und Organisationen entstanden:

- 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="#">Umfangreiches Update und Neustrukturierung</a>	<p>Die Säule wurde neu strukturiert und umfasst nun fünf (statt acht) Bereiche für bewährte Methoden. Der Inhalt wurde in die fünf Bereiche zusammengefasst und aktualisiert.</p> <p>Die neuen Bereiche für bewährte Methoden sind <a href="#">Auswahl der Architektur</a>, <a href="#">Computer und Hardware</a>, <a href="#">Datenverwaltung</a>, <a href="#">Networking und Bereitstellung von Inhalten</a> und <a href="#">Prozess und Kultur</a>.</p>	October 3, 2023
<a href="#">Kleineres Update</a>	Entfernung von nicht-inklusive Sprache.	April 13, 2023
<a href="#">Updates für das neue Framework</a>	Bewährte Methoden mit verbindlichen Anleitungen aktualisiert und neue bewährte Methoden hinzugefügt.	April 10, 2023
<a href="#">Whitepaper aktualisiert</a>	Bewährte Methoden mit neuen Implementierungsanleitungen aktualisiert.	December 15, 2022

<a href="#">Whitepaper aktualisiert</a>	Weitere bewährte Methoden und Verbesserungspläne hinzugefügt.	October 20, 2022
<a href="#">Kleineres Update</a>	Nicht inklusive Sprache entfernt.	April 22, 2022
<a href="#">Kleineres Update</a>	Säule „Nachhaltigkeit“ wurde zur Einführung hinzugefügt.	December 2, 2021
<a href="#">Kleinere Updates</a>	Aktualisierte Links.	March 10, 2021
<a href="#">Kleinere Updates</a>	Der AWS Lambda-Timeout wurde auf 900 Sekunden geändert und der Name von Amazon Keyspaces (for Apache Cassandra) wurde korrigiert.	October 5, 2020
<a href="#">Kleineres Update</a>	Fehlerhafter Link behoben.	July 15, 2020
<a href="#">Updates für das neue Framework</a>	Umfangreiche Überprüfung und Aktualisierung der Inhalte	July 8, 2020
<a href="#">Whitepaper aktualisiert</a>	Kleine Grammatikkorrekturen	July 1, 2018
<a href="#">Whitepaper aktualisiert</a>	Aktualisierung des Whitepapers mit Änderungen in AWS	November 1, 2017
<a href="#">Erstveröffentlichung</a>	Säule „Leistungseffizienz“ – AWS-Well-Architected-Framework veröffentlicht.	November 1, 2016

# AWS Glossary

For the latest AWS terminology, see the [AWS glossary](#) in the AWS-Glossar Reference.