

# Das IT-Handbuch zur Optimierung des Cloud-Budgets:

Wie CPU-Technologie Kostenersparnisse und die Geschäftsentwicklung beeinflusst



In der heutigen Zeit, in der zur geschäftlichen Unsicherheit noch eine globale Unsicherheit kommt, ist es um so wichtiger, dass Sie Ihre technologischen Ressourcen optimal einsetzen. Jetzt, wo Unternehmen zunehmend in die Cloud migrieren, um die für die Wettbewerbsfähigkeit nötige Agilität und Resilienz zu erlangen, hat sich daran nichts geändert.

Bei der Optimierung der Cloud-Ausgaben herrscht jedoch ein mangelndes Verständnis für die Rolle der CPU-Technologie vor. Oft erkennen Unternehmen während (oder nach) der Migration ihrer Infrastruktur nicht, dass sie eine Hardware-Auswahl haben. Die gute Nachricht: Die führenden Cloud-Plattformen und CPU-Hersteller bieten eine Reihe von Tools an, mit denen Sie die Kosten des Cloud-Computings optimieren können. Wenn Sie mehr aus Ihren Cloud-Ressourcen herausholen, können Sie mehr Kunden bedienen, mehr Probleme lösen und die Anpassungsfähigkeit des gesamten Unternehmens erhöhen.

## Bewährte Praktiken für die Kostenoptimierung: die Grundlagen

Beginnen Sie mit der Prüfung Ihrer Umgebung, um sicherzustellen, dass Sie bewährte Praktiken verfolgen. Das Ziel besteht darin, den Geschäftswert zu maximieren und dabei unter Berücksichtigung der effektivsten und effizientesten Verwendung von Cloud-Ressourcen Kosten zu optimieren. Wenn es um die Kostenoptimierung geht, gibt es viele Tools und Techniken, die Unternehmen verwenden können. Doch auch Tools haben ihre Grenzen. Unserer Erfahrung nach gibt es mehrere hochrangige Prinzipien, die Unternehmen jeder Größe befolgen können, um sicherzustellen, dass Beste aus der Cloud herauszuholen.



Die führenden Cloud-Plattformen und CPU-Hersteller bieten eine Reihe von Tools, mit denen Sie die Kosten des Cloud-Computings optimieren können.

- ➔ **Kostentransparenz.** Dazu gehört das Wissen, was Sie im Einzelnen ausgeben und wie bestimmte Leistungen abgerechnet werden, und dass Sie die Möglichkeit haben, darzustellen, wie (oder warum) Sie einen bestimmten Betrag zum Erreichen ausgegeben haben, um ein Geschäftsziel zu erreichen. Zu den wichtigsten Funktionen gehören die Möglichkeit, gemeinsame Verantwortlichkeiten zu schaffen, häufige Kostenüberprüfungen durchzuführen, Trends zu analysieren und die Auswirkungen Ihrer Maßnahmen in Echtzeit zu visualisieren. Mit einer standardisierten Strategie zur Organisation Ihrer Ressourcen können Sie Ihre Kosten an der Betriebsstruktur des Unternehmens abbilden, um ein Showback/Chargeback-Modell zu schaffen. Außerdem können Sie Kostenkontrollen wie Budgetalarme und Kontingente verwenden, um die Kosten langfristig unter Kontrolle zu halten.
  
- ➔ **Optimierung der Ressourcenverwendung.** Verringern Sie den Abfall in Ihrer Umgebung, indem Sie den Verbrauch optimieren. Das Ziel besteht darin, eine Reihe von Normen zu implementieren, die innerhalb einer Umgebung ein geeignetes Gleichgewicht zwischen Kosten und Performance aufrechterhalten. Dies ist der Blickwinkel, durch den man prüft, ob es ungenutzte Ressourcen oder bessere Dienste zur Bereitstellung einer App gibt oder ob sogar der Start einer benutzerdefinierten VM-Form sinnvoller wäre. Die meisten Unternehmen, die erfolgreich Verschwendung vermeiden, optimieren die Ressourcennutzung auf dezentrale Weise. Denn individuelle Anwendungseigentümer können aufgrund ihrer genauen Workload-Kenntnisse in der Regel am besten beurteilen, welche Ressourcen abzuschalten oder zu verkleinern sind. Außerdem können Sie das „Recommender“-Tool von Google Cloud verwenden, um Probleme wie unzureichend oder übermäßig bereitgestellte VM-Instanzen oder inaktive Ressourcen zu erkennen.
  
- ➔ **Preiseffizienz.** Dazu gehören Fähigkeiten wie Rabatte für die kontinuierliche oder zugesicherte Nutzung, Pauschaltarife, Sekundenabrechnung oder andere Mengenrabatte, mit denen sie die Preise für einen bestimmten Dienst optimieren können. Diese Fähigkeiten werden am besten von zentralisierten Teams in Ihrem Unternehmen genutzt, z. B. von einem Cloud Center of Excellence (CCoE) oder einem FinOps-Team, das das Potenzial für Verschwendung senken und gleichzeitig die Abdeckung aller Geschäftsbereiche optimieren kann.



Mit einer einheitlichen Organisationsstrategie bzgl. Ressourcen können Sie Ihre Kosten an der Betriebsstruktur des Unternehmens abbilden.



# Optimieren von Datenverarbeitungskosten

Wenn Unternehmen in die Cloud migrieren, besteht der erste Schritt üblicherweise in der Einrichtung virtueller Maschinen (VMs), um Workloads auf der Cloud-Plattform auszuführen. Und der Schlüssel zur Optimierung der Datenverarbeitungskosten besteht in der Auswertung des Preis-Leistungs-Verhältnisses.

## Die Wahl der richtigen CPU für Ihre VM

Es mag überraschen zu hören, dass Sie bei der physischen CPU zur Ausführung Ihrer Cloud-Workloads tatsächlich die Wahl haben – nämlich durch die Auswahl der optimalen VM. Diese Wahl kann sich maßgeblich auf Performance und Kosten auswirken. Sie profitieren von der sofortigen Kosteneffizienz, wenn Sie eine CPU mit besserem Preis-Leistungs-Verhältnis wählen. Zum Beispiel bieten N2D-VMs mit AMD EPYC™ Prozessoren der 3. Generation 13 % geringere Kosten als vergleichbare VMs der N-Serie.<sup>1</sup> Darüber hinaus kann die Ausführung Ihrer Workloads auf einer leistungsfähigeren CPU die Laufzeit verkürzen, wodurch Sie laufende Rechenkosten sparen. Stellen Sie bei der Auswertung der VM-Typen sicher, dass die zugrundeliegende CPU Ihre Workload-Bedürfnisse abdeckt.

**Die „Allzwecklösung“** eignet sich für die meisten Unternehmensanwendungen, mittlere Datenbanken und Web- oder App-Server. Suchen Sie nach einem flexiblen und umfassenden Funktionspaket, das ein Gleichgewicht aus Anpassung, Performance und Gesamtbetriebskosten bietet. Google Cloud N2D-VMs für allgemeine Zwecke mit AMD EPYC™ Prozessoren der 3. Generation sind die größten verfügbaren Allzweck-VMs und bieten bis zu 224 vCPUs. Darüber hinaus bieten N2D-VMs mit AMD EPYC™ Prozessoren der 3. Generation ein bis zu 30 % besseres Preis-Leistungs-Verhältnis gegenüber AMD EPYC™ Prozessoren der 2. Generation.<sup>2</sup>

**Der Typ „Datenbank-orientiert“** wurde für anspruchsvolle Einsatzzwecke wie elektronische Designautomatisierung (EDA), High-Performance-Computing (HPC), Hochleistungs-Web- und Werbeserver, Media-Transkodierung, wissenschaftliche Modelle, KI/ML und AAA-Gaming entwickelt. Google Cloud C2D Rechen-optimierte VMs basieren auf AMD EPYC™ Prozessoren der 3. Generation. Jede VM unterstützt bis zu 112 vCPUs und 896 GiB Speicher und bietet außerdem eine fortschrittliche Netzwerkanbindung mit bis zu 100 Gbit/s sowie bis zu 3 TB lokalen SSD-Speicher.

**Der Typ „Skalierte Workloads“** unterstützt Cloud-Bereitstellungen effizient mit hochvariablen Workloads, die die Flexibilität einer nahtlosen Skalierung ohne Über- oder Unterbereitstellung erfordern. Google Cloud Tau T2D-VMs werden durch AMD EPYC™ Prozessoren der 3. Generation angetrieben. Tau-VMs bieten ein 42 % besseres Preis-Leistungs-Verhältnis gegenüber vergleichbaren Allzweck-Angeboten sowie eine 56 % höhere absolute Performance.<sup>3</sup> Die von T2D VMs bereitgestellte x86-Kompatibilität bietet Ihnen marktführende Performance-Steigerungen und Kostenersparnisse ohne, dass Anwendungen auf eine neue Prozessorarchitektur portiert werden müssen.

42%

*Das verbesserte Preis-Leistungs-Verhältnis von Tau-VMs im Vergleich zu ähnlichen Allzweck-Angeboten*

## Zahlen Sie nur für das, was Sie auch brauchen

1

**Dimensionieren Sie VMs richtig.** Die Workload-Anforderungen können sich im Laufe der Zeit ändern. Einst optimierte Instanzen bedienen jetzt möglicherweise weniger Nutzer und Datenverkehr. Oder aber sie werden durch steigende Nachfrage überlastet. Durch die Überwachung der Nutzung können Sie langfristige Trends ermitteln, die basierend auf vCPU- und RAM-Auslastung auf eine nötige Vergrößerung oder Verkleinerung Ihres Maschinentyps hindeuten können.

2

**Identifizieren Sie inaktive VMs.** Der einfachste Weg, die Rechnung Ihres Cloud-Anbieters zu reduzieren, besteht darin nicht mehr genutzte Ressourcen zu entfernen, z. B. Konzeptstudien, die inzwischen nicht mehr priorisiert werden, oder Zombie-Instanzen, die niemand gelöscht hat.

3

**Lassen Sie VMs automatisch starten und herunterfahren.** Produktionssysteme sind in der Regel rund um die Uhr in Betrieb. VMs in Entwicklungs-, Test- oder privaten Umgebungen werden jedoch möglicherweise nur in den Geschäftszeiten verwendet und die Deaktivierung kann Geld sparen.

## Nutzen Sie Preisanziehe

Machen Sie sich mit den Anreizen vertraut, die Ihr Cloud-Anbieter für die kostensparende Verwaltung Ihrer VM-Flotte bietet. Zum Beispiel bietet Google Cloud Rabatte für die zugesicherte Nutzung, die ideal bei planbaren Steady-State-Workloads sind. Sie können sich für ein oder drei Jahre verpflichten und im Gegenzug erhebliche Einsparungen bei Ihrer VM-Nutzung erzielen.

## Automatisieren von Kostenoptimierungen

Der beste Weg, um sicherzustellen, dass Ihr Team stets die besten Verfahren zur Kostenoptimierung anwendet, ist die Automatisierung dieser Verfahren. Kennzeichnungen können Ihnen helfen, Instanzen nach Nutzungsanforderungen zu kategorisieren. Zum Beispiel können Sie so Instanzen, die nur von Entwicklern in den Geschäftszeiten verwendet werden, von Produktionsinstanzen trennen. Die automatische Skalierung bietet Ihnen die Möglichkeit, die Kapazität nur bei Bedarf zu erhöhen. Sie können die Skalierung bei steigendem Datenverkehr problemlos vornehmen und dann automatisch wieder herunterfahren, wenn der Bedarf an Instanzen sinkt.



# Google Cloud und AMD: Gemeinsam sind wir stark

VMs mit AMD EPYC™ Prozessoren liefern zuverlässige, Workload-optimierte Performance mit herausragender Flexibilität, fortschrittlichen Sicherheitsmerkmalen und beeindruckender Leistungs- und Kosteneffizienz. Die Kombination aus x86-Architektur und führender Kerndichte schafft ein attraktives Preis-Leistungs-Verhältnis, das auf einem innovativen Preismodell basiert, mit dem Kunden ihre Infrastruktur optimieren können. Die verbesserte Google Cloud VM-Performance gestützt durch AMD EPYC™ Prozessoren ermöglicht es Ihnen, Aufgaben schneller abzuschließen, sodass Ihre VM in kürzerer Zeit zur nächsten Abfrage wechseln kann und die Gesamtverarbeitungszeit reduziert wird.

VMs mit AMD auf Google Cloud ermöglichen Kostenersparnisse bis zu 13 % gegenüber vergleichbaren Alternativen,<sup>4</sup> sodass Sie Ressourcen geschäftskritischen Aufgaben und Innovationen zuweisen können. Verwandeln Sie Ihre bestehenden Workloads mit wenigen Klicks zu AMD EPYC™ VMs und verlassen Sie sich auf ein branchenübliches x86-Ökosystem.



VMs mit AMD EPYC™ Prozessoren  
liefern zuverlässige Workload-optimierte  
Performance mit herausragender Flexibilität.



# Tools für die Kostenoptimierung

## AMD Cloud Cost Advisor:

Erhalten Sie Echtzeiteinblicke in die voraussichtlichen Kostenersparnisse, wenn Sie zu Cloud-Instanzen mit AMD in Google Cloud wechseln.

## Rabattanalysebericht für zugesicherte Nutzung

in der Cloud-Konsole: Hilft Ihnen, die Effektivität der von Ihnen erworbenen Verpflichtungen zu verstehen und zu analysieren und sogar abzuschätzen, wie Ihre Ressourcenbasis auf der Grundlage historischer Daten aussieht.

## Benutzerdefinierte

Dashboards: Erhalten Sie detailliertere Kosteneinblicke in Ihre Umgebung, indem Sie Kosten mithilfe von Kennzeichnungen Abteilungen oder Teams zuweisen. So können Sie eine Ressource auf Grundlage einer vordefinierten Geschäftskennzahl zuordnen und die Ausgaben im Laufe der Zeit verfolgen.



## Google Cloud Deployment

Manager: Automatisieren Sie die Erstellung und Verwaltung von Google Cloud-Ressourcen, einschließlich Compute Engine, um Leitplanken zu erstellen, bevor Sie eine Cloud-Ressource bereitstellen.

## Recommenders:

Google Cloud bietet mehrere Empfehlungstools, mit denen Sie Ressourcen optimieren können, darunter ein Empfehlungs-Tool für inaktive VMs, das inaktive VMs und dauerhaft bestehende Laufwerke auf Grundlage von Nutzungskennzahlen ermittelt. Das Empfehlungs-Tool für die richtige Dimensionierung zeigt Ihnen, wie Sie Ihren Maschinentyp auf Grundlage von Änderungen in der vCPU- und RAM-Nutzung herabdimensionieren.

## Cloud-Abrechnungsberichte:

Erhalten Sie einen Überblick über Ihre Compute-Engine-Kosten.

# Lytics verbessert die Effizienz, um höhere Margen zu ermöglichen

Die Kundendatenplattform Lytics stellt hohe Anforderungen an Rechenleistung und Speicherkapazität, um seine ausgeklügelten Orchestrings-Engines zu betreiben. Lytics ist zu Google Cloud migriert, um die Kosten zu senken. Das Unternehmen konnte sofortige Kostenersparnisse pro Byte und Kern sowie eine verbesserte Performance feststellen. Das Streben nach Effizienz bei Lytics ist damit jedoch noch nicht beendet.

Als Google Cloud AMD EPYC™-CPU-betriebene Instanzen auf den Markt brachte, war es an der Zeit, die Optionen des Unternehmens zu überprüfen. Lytics führte einen Drei-Wege-Test der bestehenden Intel basierten Google Cloud Platform N1-Instanzen im Vergleich zu N2-Instanzen und den AMD basierten N2D-Instanzen aus. Basierend auf den Ergebnissen beschloss Lytics, seine Workloads zu N2D-Instanzen zu migrieren. „Die N2D-Instanzen mit AMD EPYC™ CPUs verbessern unsere Plattformmargen enorm“, sagt Kathryn Gruenefeldt, VP of Engineering bei Lytics. „Wenn wir die Kundensupport-Kosten reduzieren können, können wir diese Ersparnisse als attraktive Preise an unsere Kunden weitergeben.“<sup>5</sup>



„Die N2D-Instanzen mit AMD EPYC CPUs verbessern unsere Plattformmargen enorm.“

- Kathryn Gruenefeldt, VP of Engineering, Lytics

## Ergebnisse

13% Kostenreduktion

14-20%  
Bessere Performance





# FullStory treibt Wachstum mit höherer Geschwindigkeit und Kapazität voran

FullStory bietet Echtzeit-Analysen für tausende Kunden in den Sektoren eCommerce-Einzelhandel, SaaS und in anderen digitalen Bereichen. Dabei müssen hunderte Instanzen auf tausenden Kernen ausgeführt werden. Das Unternehmen wollte den Service, den es seinen Kunden anbietet, durch eine Reduktion der Latenzen auf bestehenden Plattformen verbessern und neue Großaufträge gewinnen. FullStory stand vor einer teuren Erweiterung seiner Cloud-Infrastruktur. Doch dann erkannte das Unternehmen, dass AMD EPYC™ CPU-gestützte Google Cloud-Instanzen die bessere Wahl waren. Nach Auswertung der Optionen migrierte das Unternehmen seine Datenbank-Cluster, die eine Größe von bis zu hunderten Terabyte erreichen können, auf eine einzelne N2D-Instanz. Jaime Yap, Director of Engineering bei FullStory, zufolge war die Migration nahtlos: „Es waren nur wenige Softwareänderungen nötig, da es sich in beiden Fällen um 64-Bit-x86-Architekturen handelt. Die Ergebnisse waren überragend.“

Der N2D-Cluster konnte die Latenzbedenken für die größten Kunden von FullStory vollständig ausräumen und das Unternehmen war in der Lage, mehr Großkunden an Bord zu bringen – und das ohne eine kostspielige Instanzenerweiterung mit einem zusätzlichen neuen Cluster. Doch dies waren nicht die einzigen Ersparnisse. „Im Kostenvergleich waren die N2D-Instanzen etwa 13 % günstiger als N1.“<sup>6</sup>



„Es waren nur wenige Softwareänderungen notwendig, da es sich in beiden Fällen um 64-Bit-x86-Architekturen handelt.“

- Jaime Yap, Director of Engineering, FullStory

## Ergebnisse

# 40%

Latenzreduktion bei Abfragen großer Kunden

# 13%

 Kostenreduktion

Kein weiterer Datenbank-Cluster nötig



# Qubit profitiert von maximaler Stabilität bei reduzierten Kosten

Die KI-gestützte Produktempfehlungs-Engine von Qubit stellt 9,5 Millionen personalisierte Erfahrungen für 364 eCommerce-Shops bereit, darunter einige der größten Luxusmarken im Einzelhandel. Diese Art von Volumen erfordert eine zuverlässige Datenverarbeitungsinfrastruktur, insbesondere während der Datenverkehrsspitzen, wenn die Einzelhändler den größten Teil ihrer Umsätze erzielen. Qubit musste diese Zuverlässigkeit aufrechterhalten und gleichzeitig die Kosten optimieren. Der Google Cloud-Ansprechpartner des Unternehmens empfahl den AMD EPYC™ Prozessor, der eine leistungsäquivalente Alternative zu den N2-highmem-4 Intel-basierten Instanzen bei geringeren Kosten bot.

Aber würde er wirklich die Stabilität bieten können, auf der der Ruf des Unternehmens ruht? Um das herauszufinden, migrierte Stephen Boyle, Infrastructure Team Lead, die Dienste von Qubit während des Betriebs auf die Staging-Server und führte sie dort dann für drei Wochen vor Black Friday aus. „Es gab überhaupt keine Probleme“, sagt Boyle. „Es war supereinfach. Und dann sparten wir mit der Migration Geld!“ Dieser Testlauf gab Qubit das Vertrauen, den gesamten Empfehlungsdienst in der Woche vor dem Black Friday zu migrieren. Dazu gehörten 160 N2D-Instanzen mit vier AMD EPYC™ Prozessorkernen und jeweils 32 GB Speicher, die Google Kubernetes Engine (GKE) ausführen. Die Migration benötigte nur fünf Stunden.<sup>7</sup>



„Mit den Ersparnissen, die wir durch die Migration auf die AMD EPYC gestützten Google Cloud Platform-Instanzen erzielen konnten, können wir neue, innovative Technologie erwerben und diese schnell im Markt einsetzen.“

- Sergio Iacobucci, Director of Marketing and Partnerships, Qubit

## Ergebnisse

- ➔ Felsenfeste Stabilität
- ➔ Schnelle Migration

13% Kostenreduktion



# Optimieren Sie Ihre Cloud-Ausgaben

Die Migration Ihrer geschäftskritischen Workloads auf Google Cloud VMs mit AMD EPYC™ Prozessoren ist der erste Schritt in die digitale Transformation, die Ihrem Unternehmen mehr Agilität, Flexibilität und Effizienz beschert.

Um Ihr Cloud-Computing-Budget optimal zu nutzen, sollten Sie Kostenoptimierungsstrategien in jedem Schritt auf Ihrem Weg in die Cloud integrieren. AMD EPYC™ Prozessoren sind leistungsstark sowie kosteneffektiv und bilden das Fundament für die Cloud-Kostenoptimierung. In Kombination mit der leistungsstarken Technologie und den Fachkenntnissen von DoIT ermöglichen sie beträchtliche Kostenersparnisse und Leistungsverbesserungen. Darüber hinaus bieten Google Cloud und AMD ein ganzes Paket an Kostenoptimierungs-Tools, um Ihr Cloud-Budget zu optimieren.

**Beginnen Sie noch heute.**



## Verweise und Quellen

<sup>1</sup> Blog: [Neue AMD EPYC basierte Compute-Engine-Familie, jetzt in der Betaversion](#)

<sup>2</sup> Blog: [N2D-VMs mit den neuesten AMD EPYC CPUs ermöglichen ein durchschnittlich 30 % besseres Preis-Leistungs-Verhältnis](#)

<sup>3</sup> [Die neuen Tau-VMs bieten ein führendes Preis-Leistungs-Verhältnis für skalierte Workloads](#)

<sup>4</sup> VMs mit AMD in der Google Cloud sorgen für Kostenersparnisse von bis zu 13 % gegenüber vergleichbare Alternativen auf Grundlage von:

- N2D Rome bietet 13 % geringere Kosten als die vergleichbare N-Serie - <https://cloud.google.com/blog/products/compute/announcing-the-n2d-vm-family-based-on-amd>

- N2D Milan hat den gleichen Preis wie N2D Rome - <https://cloud.google.com/blog/products/compute/3rd-gen-amd-epyc-comes-to-compute-engine-n2d-machine-family>

- T2D hat den gleichen Preis wie N2D - Preise verfügbar unter <https://cloud.google.com/compute/vm-instance-pricing> - zum Beispiel beträgt der T2D-On-Demand-Preis in Iowa 0,027502 USD / vCPU-Stunde, während der N2D-On-Demand-Preis in Iowa 0,027502 USD / vCPU-Stunde beträgt.

- C2D bietet bis zu 13 % geringere Kosten als C2 - Preise verfügbar unter <https://cloud.google.com/compute/vm-instance-pricing> - zum Beispiel beträgt der C2-On-Demand-Preis in Iowa 0,03398 USD, während der C2D-On-Demand-Preis in Iowa 0,029563 USD beträgt. Mit einer Prozentdifferenzformel lässt sich ermitteln, dass C2D ca. 13 % günstiger ist als C2.

<sup>5</sup> [FullStory-Fallstudie](#)

<sup>6</sup> [Lytics-Fallstudie](#)

<sup>7</sup> [Qubit-Fallstudie](#)