



Link: <https://www.cio.de/a/wie-sie-public-cloud-kosten-mit-ki-automatisch-optimieren,3607653>

Workloads in die richtige Cloud schieben

Wie Sie Public-Cloud-Kosten mit KI automatisch optimieren

Datum: 13.11.2019

Unternehmen, die in der Public Cloud außer reservierten Instanzen auch Low-Priority-Kapazität nutzen, erzielen enorme Kostenvorteile. Allerdings besteht bei solchen Instanzen das Risiko, dass der Provider den Zugriff jederzeit beenden kann. Erfahren Sie hier, wie ein Cloud-Service, der mit künstlicher Intelligenz (KI) arbeitet, vollautomatisch die preisgünstigsten verfügbaren Low-Priority-Instanzen für den sicheren Betrieb Ihrer Workloads ermittelt.

Immer mehr Unternehmen nutzen zusätzlich Kapazität aus der Public Cloud, um ihre Daten zu verarbeiten, zu verwalten und zu analysieren. Das bestätigen aktuelle Umfragen wie der **Cloud-Monitor 2019**¹, den Bitkom Research für die Wirtschaftsprüfungsgesellschaft KPMG erstellt hat.

Die Hälfte der Firmen, die eine Enterprise Cloud (Public und Private) betreiben, gibt dafür pro Jahr durchschnittlich mehr als 1,1 Millionen Euro aus. Und 2019 sollen sich die Cloud-Ausgaben noch um ein Viertel erhöhen, so der "Rightscale 2019 State of the Cloud Report", eine Umfrage von Flexera unter knapp 800 Technologieexperten.

Kostenfalle Public Cloud

Daher hat für rund zwei Drittel der Befragten die oberste Priorität, die Kosten zu optimieren. Das erweist sich in der Praxis als komplexe Aufgabe, weil es mehrere Möglichkeiten gibt, Rechenkapazität aus einer Public Cloud zu beziehen, die sich preislich gewaltig unterscheiden.

- Am teuersten sind On-Demand- bzw. Pay-as-you-go-Instanzen, die kurzfristig in Anspruch genommen werden können und einen einfachen und flexiblen Zugriff auf Computing-Ressourcen bieten, ganz ohne Laufzeitbindung.
- Deutlich günstiger sind reservierte Instanzen. Hier wird voraussichtlich benötigte Rechenleistung - in der Regel für einen Zeitraum von ein bis drei Jahren - gegen Rabatt im Voraus bezahlt. Allerdings ist der Markt sehr volatil. Diese Preise können ein Jahr später schon höher sein als für On-Demand-Instanzen, da Provider permanent die Preise justieren.
- Am günstigsten sind Low-Priority-Instanzen, die AWS zum Beispiel als Spot-Instanzen bezeichnet. Das sind überschüssige Kapazitäten der Public-Cloud-Provider, welche diese günstig vertreiben. Gegenüber On-Demand-Instanzen kann man bis zu 80 Prozent sparen, wie sich anhand des **Amazon Instance Advisor**² nachvollziehen lässt. Das geht jedoch zulasten der Verfügbarkeit, denn der Provider kann den Zugriff auf Low-Priority-Kapazitäten ohne Vorwarnung jederzeit mitten im Betrieb beenden und zu reservierten Instanzen umwidmen.

Darüber hinaus spielen bei den Preisen für Rechenkapazität auch die Region, in der ein Public-Cloud-Provider sie zur Verfügung stellt, die Anzahl der Betriebsstunden und andere Faktoren eine Rolle. Das erschwert es, stets die günstigste Rechenleistung zu ermitteln, vor allem wenn ein Unternehmen viele Anwendungen und Services - nicht selten sind es Hundert und mehr - in der Public Cloud betreibt.

Auf dem Markt verfügbare IT-Werkzeuge geben zwar Empfehlungen zur Optimierung der Kosten. Das funktioniert aber nur zu einem gewissen Grad. Viel Einsparpotenzial wird allein dadurch verschenkt, dass etliche der damit verbundenen Aufgaben manuell durchgeführt werden müssen.

Einsparpotenziale optimal ausschöpfen

Abhilfe schaffen hier Cloud-Services, die mit künstlicher Intelligenz (KI) und maschinellem Lernen (ML) arbeiten und dafür sorgen, dass Instanzen optimal genutzt werden. Benötigt ein Unternehmen zusätzlich Low-Priority- oder On-Demand-Instanzen, zum Beispiel bei dringenden Computing-Anforderungen oder für Anwendungen mit flexiblen Start- und Endzeiten, stellen solche Services die günstigsten verfügbaren Low-Priority-Instanzen dafür bereit und starten sie automatisch.

Zudem können solche KI-Lösungen automatisch dafür sorgen, dass ungenutzte Instanzen heruntergefahren werden. Häufig vergessen Anwender, ungenutzte Rechenkapazitäten wieder freizugeben.

Nach diesem Konzept arbeitet die Compute-Management-Lösung "Cloud Cost Optimization" von wenovate. Der Cloud-Service überwacht kontinuierlich die Preise am Low-Priority-Markt und identifiziert automatisch die aktuell preiswertesten Instanzen. Das umfasst gegenwärtig das Angebot von Amazon Web Services (AWS), der Google Cloud Platform und von Microsoft Azure.

Die KI-Algorithmen von wenovate ermitteln anhand historischer und statistischer Daten, wann der Provider die aktuell genutzten Low-Priority-Instanzen beendet und ersetzt diese durch andere Low-Priority-Instanzen, die zu diesem Zeitpunkt am günstigsten sind.

Gibt es keine preiswerten freien Instanzen am Low-Priority-Markt, startet wenovate Cloud Cost Optimization dedizierte On-Demand-Instanzen und stellt so sicher, dass Anwendungen und Workloads zu jedem Zeitpunkt reibungslos ausgeführt werden können. Auf der anderen Seite kehrt der Service selbständig zu freier Rechenkapazität zurück, sobald Low-Priority-Ressourcen verfügbar werden.

Mit diesem Konzept stellt wenovate Cloud Cost Optimization sicher, dass ein Cluster immer kostengünstig ist. Es priorisiert reservierte Instanzen (reserved instances) und stellt sicher, dass diese vollständig genutzt werden. Und nur wenn keine reservierten Instanzen verfügbar sind, startet es zusätzliche Berechnungsinstanzen.

Cloud-Kosten um bis zu 80 Prozent senken

Auf diese Weise kann ein Unternehmen zu jedem Zeitpunkt preisgünstige Computing-Ressourcen aus der Public Cloud nutzen und im Vergleich zu On-Demand- bzw. Pay-as-you-go-Kapazität bis zu 80 Prozent sparen. Das ist ein echter Mehrwert, zumal für wenovate Cloud Cost Optimization keine Lizenzgebühren anfallen. Der Kunde entrichtet als Gebühr lediglich einen gewissen Prozentsatz der eingesparten Kosten für den Cloud-Account, mit dem die wenovate-Lösung verknüpft ist.

Alle aktuellen Kostenersparnisse und Analysen zum Sparpotenzial führt der Cloud-Service in einem Dashboard zusammen, wo sie optisch ansprechend visualisiert werden. Damit behalten Nutzer jederzeit den Überblick über alle relevanten Informationen.

Das und die kurze Einführungszeit - je nach Umfang und Komplexität der genutzten Computing-Kapazitäten liegt sie zwischen wenigen Minuten und zwei Tagen - sind ein weiteres großes Plus der Lösung.

Erfahren Sie hier³, wie mit wenvate Cloud Cost Optimization das Einsparpotenzial, das in Public-Cloud-Accounts schlummert, bestmöglich ausgeschöpft werden kann.

Links im Artikel:

¹ <https://w.idg.de/2xudCe1>

² <https://w.idg.de/2okdOMc>

³ <https://w.idg.de/35maeSo>

IDG Business Media GmbH

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium in Teilen oder als Ganzes bedarf der schriftlichen Zustimmung der IDG Business Media GmbH. dpa-Texte und Bilder sind urheberrechtlich geschützt und dürfen weder reproduziert noch wiederverwendet oder für gewerbliche Zwecke verwendet werden. Für den Fall, dass auf dieser Webseite unzutreffende Informationen veröffentlicht oder in Programmen oder Datenbanken Fehler enthalten sein sollten, kommt eine Haftung nur bei grober Fahrlässigkeit des Verlages oder seiner Mitarbeiter in Betracht. Die Redaktion übernimmt keine Haftung für unverlangt eingesandte Manuskripte, Fotos und Illustrationen. Für Inhalte externer Seiten, auf die von dieser Webseite aus gelinkt wird, übernimmt die IDG Business Media GmbH keine Verantwortung.