

Warum AIOps? Anwendungs-Performance



Inhalt

- **Warum Anwendungsperformance?**
 - Die Ära moderner Anwendungen
 - Der moderne Anwendungshostingdienst
 - Der wachsende Druck auf die Anwendungsperformance
 - Eine effektive gemeinsame Nutzung von Infrastrukturrressourcen erfordert eine Priorisierung
 - Eine wirksame öffentliche Cloud-Ressourcenbeschaffung erfordert Wissen um die Anwendung

- **Warum Turbonomic Application Resource Management für IBM Cloud Paks?**
 - Anwendungs-Ressourcenmanagement (ARM) sorgt für Performance und Einsparungen
 - Die neuen Anforderungen zur Sicherstellung der Performance moderner Anwendungen

- **Schlussfolgerung**

- **Über Turbonomic, ein Unternehmen von IBM**

Warum Anwendungsperformance?

Die Ära moderner Anwendungen

In der Zeit moderner Anwendungen, in der mittlerweile mehr als zwei Drittel des gesamten Bruttoinlandsprodukts digital entstehen, ¹ hat einzig die Anwendungsperformance für CIOs höchste Priorität, denn die Anwendung *ist* das Geschäft. Die Bereitstellung von Anwendungen ist der Hauptgrund für das Bestehen von IT. Der CIO mit IT-Führungsaufgaben hat keine andere Wahl, als Anwendungsperformance zu liefern, um sicherzustellen, dass das Unternehmen niemals durch die IT eingeschränkt wird. Tatsächlich wird der CIO als Versager betrachtet, wenn er die Anwendungsleistung nicht erbringt; ironischerweise ist es für CIOs sehr vernünftig, ihre geplanten Budgets zu überschreiten. Kurz gesagt, es schädigt das Unternehmen, wenn die Anwendungsperformance nicht sichergestellt wird. Die Herausforderung besteht darin, dass die Anwendungen mit den größten Hindernissen auch die Anwendungen mit den größten Bedarfsschwankungen und somit oft die wertvollsten Apps sind.

Der moderne Anwendungshostingdienst

Investitionen in die Anwendungsentwicklung übertreffen bei weitem den Aufwand des Anwendungshosting. Mit dieser relativen Investition stellen Unternehmen wissentlich und bereitwillig Überkapazitäten ihrer Infrastruktur und Cloudumgebungen bereit, um das Risiko der Anwendungsperformance zu minimieren. Infrastrukturrressourcen sowohl in Rechenzentren als auch in der Cloud werden schnell zu austauschbaren Rohstoffen und damit zu einer weniger zuverlässigen Rechtfertigung für die Kapitalrendite (ROI). Eine schlechte Anwendungsleistung schafft darüber hinaus Misstrauen zwischen Line of Business-Anwendungseigentümern und den ITOps- und CloudOps-Teams, die den Anwendungshostingsservice bereitstellen. Bedenken Sie – die LOB investiert Millionen Dollar in Hunderte Menschen, die neue Nutzer-Erfahrungen entwickeln, setzen ihren Ruf aufs Spiel – nur damit die Ops-Teams den Endbenutzern das langsame „Laderad“ liefern. All diese Investitionen in die Verbesserung der Endbenutzerfunktionalität sind verloren.

Der wachsende Druck auf die Anwendungsperformance

Kapazitätsengpässe bei Anwendungen treten auf, wenn die Infrastruktur – lokal oder in der Cloud – den Bedarf der Anwendung und ihrer Endnutzer nicht decken kann. Der Mangel an Infrastrukturrressourcen ist die häufigste Ursache für Leistungseinbußen bei Anwendungen. Im Gegensatz dazu ist die Architektur des Anwendungscodes, die mit APM-

Tools (Application Performance Management) überwacht wird, in Produktionsumgebungen selten weniger als 10 % der Ursache für eine Verschlechterung der Anwendungsleistung. Da sich die Anwendungsentwicklung auf die Auslieferung von qualitativ hochwertigem Code durch verbesserte Prozesse wie kontinuierliche Integration, kontinuierliche Bereitstellung (CICD), Qualitätssicherung (QA), Vorproduktion und Staging konzentriert, wirkt sich die Codequalität immer seltener auf die Leistung aus.

Eine effektive gemeinsame Nutzung von Infrastrukturrressourcen erfordert eine Priorisierung

In den lokalen Rechenzentren gehen Anwendungen ein, wenn der Bedarf an gemeinsamen Infrastrukturrressourcen nicht priorisiert wird. Angesichts dieser Begrenzung werden Anwendungsbelastungen üblicherweise mit Überkapazitäten versehen und ohne ausreichendes Verständnis und Zusammenhang für die verfügbaren Ressourcen verteilt. Bei der gemeinsamen Nutzung von Ressourcen bedienen sich alle Anwendungen aus einem einheitlichen Ressourcenpool, unabhängig von der Nutzung. Eine überdimensionierte und falsch dimensionierte Ressourcenzuordnung führt zu dauerhafter Überlastung, was Verletzungen der Service-Level-Vereinbarungen (SLA), ineffizienter manueller Fehlersuche, ständiger Ressourcenanpassung und, wie bereits erwähnt, zu ungenutzten Investitionen in die Anwendungsentwicklung führt.

Die heutigen reaktiven und auf eine einzelne Ressource ausgerichteten Monitoring-Tools sind nicht in der Lage, die Beziehung zwischen Anwendungen und Infrastruktur zu verstehen, und sind daher auf manuelle Interpretationen und Eingriffe angewiesen, um Überlastungen der Ressourcen aufzulösen. Je mehr Zeit für die Auflösung benötigt wird, desto größer die Wahrscheinlichkeit, dass eine Entscheidung darüber, was nun zu tun ist, hinfällig wird.

Eine Alternative ist die kontinuierliche Nutzung der dynamischen Anwendungsnachfrage, um die Zuweisung des gesamten Stapels gemeinsam genutzter Infrastrukturrressourcen zu priorisieren. Wenn die Nachfragekurve einer Anwendung steigt, erhöht sich ihre relative Priorität für den Erhalt und die Beibehaltung von Ressourcen. Wenn die Nachfragekurve sinkt, verringert sich ihre relative Priorität für den Erhalt und die Beibehaltung von Ressourcen. Diese kontinuierliche Neupriorisierung des vollständigen Stapels gemeinsamer Ressourcen minimiert Kapazitätsengpässe, ermöglicht eine flexible Ressourcenverteilung über alle Anwendungen hinweg und sichert die Anwendungsperformance.

Eine wirksame öffentliche Cloud-Ressourcenbeschaffung erfordert Wissen um die Anwendung

In der öffentlichen Cloud gehen Anwendungen ein, wenn Instanzen – Ressourcen – nicht ausreichen, um die Nachfrage der Anwendung zu bedienen. Um dieses Risiko anzugehen, stellen Cloud-Administratoren oft Überkapazitäten an Instanzen bereit. In beiden Fällen ergibt sich eine falsche Instanzen-Zuteilung aus dem beschränkten Wissen des Cloud-Administrators um die Ressourcenanforderungen des Anwendungsbedarfs. Außerdem sind Entwickler, die sich auf die Konstruktion der Anwendung konzentrieren, normalerweise wenig interessiert an Profilermittlung und Vorhersage ihrer Ressourcenanforderungen. Daraus folgt, dass grob geratene Schätzungen der Ressourcen vorgenommen werden, während der Cloud-Anbieter das Risiko, die Last und die Konsequenzen dieser groben Schätzung auf den Kunden abwälzt.

Als weitere Komplikation kommt hinzu, dass die richtige Entscheidung bezüglich der Instanz zu einer immer größeren Herausforderung wird, da eine einzelne Amazon Elastic Compute Cloud (Amazon EC2) Instanz-Bereitstellung in Bezug auf Ressourcentyp, zugrundeliegende Hardware, Größenverhältnis, Geografien, Preisgestaltung, Reservierungen, Sparpläne und so weiter Millionen verschiedener Konfigurationsoptionen bietet.

Weiterhin benötigen dynamische Änderungen beim Anwendungsbedarf eine kontinuierliche Neubewertung dieser Schätzung, die für die Auswahl der Cloud-Instanzen verwendet wird. Und schließlich erfordert die Elastizität von öffentlichen Cloud-Anwendungs-Instanzen eine kontinuierliche Kalibrierung in Echtzeit – insbesondere mit der Akzeptanz von kurzlebigen, Container-basierten Anwendungen. Als Ergebnis werden die meisten öffentlich zugänglich Cloudinstanz-Zuweisungen mit Überkapazitäten und manuell und reaktiv als Absicherung von Performance-Risiken bei völlig fehlendem Verständnis des Anwendungsressourcenbedarfs verwaltet. Darüberhinaus fehlt den Monitoring-basierten Cloud-Anbietern und Cloud-Management-Plattformtools das Verständnis für den Anwendungsbedarf zur Sicherstellung der Anwendungsperformance in der Cloud. Sie sind eingeschränkt auf Kostentransparenz, historische Abrechnungen und eine Kostenzuordnung nach Abteilungen – was nichts zu tun hat mit der Performance von Anwendungen in einer öffentlichen Cloud.

Warum Turbonomic Application Resource Management für IBM Cloud Paks?

Turbonomic, ein IBM Unternehmen, bietet die Grundlage für den modernen Anwendungs-Hosting-Service, indem es ein einzigartiges Verständnis der Anwendungsnachfrage nutzt, um kontinuierliche Anwendungs-Resourcing-Maßnahmen und Performance-Analysen bereitzustellen, um die Anwendungsleistung in Echtzeit, über einen längeren Zeitraum und rund um die Uhr zu gewährleisten.

Turbonomic Application Resource Management für IBM Cloud® Paks verwendet ein gemeinsames Datenmodell, so dass Kunden ihre heutigen und zukünftigen Anwendungen zuverlässig mit Ressourcen ausstatten können, unabhängig davon, ob sie vor Ort, in öffentlichen Clouds oder in der Edge laufen.

Anwendungs-Ressourcenmanagement (ARM) sorgt für Performance und Einsparungen

Turbonomic's ARM nutzt KI-gestützte Analysen und hilft somit, die Anwendungsperformance durch einen kontinuierlichen Abgleich der Echtzeit-Anwendungsbedarfsanforderungen mit Ressourcenarten und -größen sicherzustellen. Resourcing-Aktionen schließen das Starten und Stoppen, anfängliche und kontinuierliche Verteilung, Skalierung und Ändern der Größe ein.

Im Gegensatz dazu geben Hunderte von Tools – und sogar Tabellenkalkulationen – an, den Kunden Geld zu sparen, aber deren „Empfehlungen“ basieren häufig auf einfachen Grenzwert-Warnungen, die Performanceprobleme mit sich bringen können. In einer kürzlich durchgeführten Umfrage gaben 39 % der FinOps-Experten an, dass es für sie die größte Herausforderung sei, Ingenieure zum Handeln zu bewegen.² Die Ursache? Mangelndes Vertrauen in die vorhandenen Tools, die zur Generierung dieser „Handlungen“ verwendet werden.

Die neuen Anforderungen zur Sicherstellung der Performance moderner Anwendungen

Die Turbonomic AIOps-Plattform verwendet ein umfassendes einheitliches Datenmodell, das alle gemeinsamen Ressourcen, von denen die Anwendungsperformance abhängt, aufnimmt, standardisiert und verwaltet. Am wichtigsten ist, dass es die Topologie der Lieferkettenbeziehungen – „Stitching“ – zwischen jeder Ressourcenbeschaffungsbhängigkeit von der Anwendung bis zur Infrastruktur erstellt.

Im Vergleich dazu können manuelle Tools und Toolfragmente die Performance nicht sicherstellen, da sie die Ressourcen isoliert überwachen und, wenn überhaupt, nur einen begrenzten Überblick über die Anwendung haben. Anwendungsperformance erfordert das Verständnis und die Verwaltung aller Ressourcenabhängigkeiten in genau der richtigen Menge, Reihenfolge und im richtigen Zeitrahmen.

Schlussfolgerung

In einer Zeit moderner Anwendungen besteht das Unternehmen in seiner Anwendung. Da Anwendungen komplexer werden und es mehr Abhängigkeiten sowie vielfältigere und verteilte Umgebungen gibt, wächst das Risiko für Anwendungsperformance und Benutzererfahrung exponentiell. Der einzige Weg, diesen Herausforderungen zu begegnen, ist es, einen anwendungsbezogenen Ansatz zu implementieren, der den Anwendungsbedarf und das Angebot verfügbarer Ressourcen kontinuierlich auswertet und verlässlich Empfehlungen ausgibt, denen sowohl Geschäftswie Anwendungseigentümer vertrauen können. Mit der Zeit und mit wachsendem Vertrauen ist dann die nächste Stufe eine Automatisierung dieser Aktionen. Das Ergebnis sind dann sehr leistungsfähige Anwendungen, massive Einsparungen bei den IT-Ausgaben und die Chance, geschäftliche Innovation zu entfesseln. Das ist die Magie von AIOps.

Mehr als zwei Drittel des weltweiten BIP sind mittlerweile digital.¹

- Die Bereitstellung von Anwendungen ist der Hauptgrund für das Bestehen von IT.
- CIOs müssen sicherstellen, dass das Vorankommen ihres Unternehmens niemals durch die IT eingeschränkt wird.
- Die Anwendungsentwicklung beträgt das dreifache der Kosten für das Anwendungshosting.
- Unternehmen stellen zu viele Ressourcen zur Verfügung, um das Risiko einer Verschlechterung der Anwendungsleistung zu mindern.
- Der Mangel an Infrastrukturressourcen ist die häufigste Ursache für Leistungseinbußen bei Anwendungen.
- Anwendungen gehen ein, wenn der Bedarf an gemeinsamen Infrastrukturressourcen nicht priorisiert wird.
- Die heutigen ressourcenorientierten Überwachungstools können ohne ein Verständnis für den Anwendungsbedarf lediglich auf negative Situationen reagieren.
- Eine kontinuierliche Neupriorisierung des gesamten Stapels gemeinsam genutzter Ressourcen auf der Grundlage des Anwendungsbedarfs ist die einzige Möglichkeit, die Anwendungsleistung zu gewährleisten.

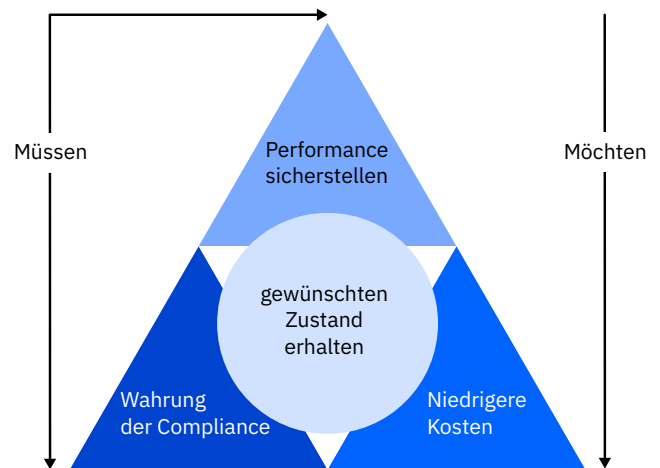


Abbildung 1. Anwendungen befinden sich im gewünschten Zustand, bei dem die Performance sichergestellt ist, während die Compliance zu günstigsten Kosten erhalten wird.

- Cloud-Administratoren haben nur begrenzte Kenntnis von Anwendungsressourcenanforderungen.
- Anwendungsentwickler sind fokussiert auf Geschäftslogik und geben Entscheidungen zur Ressourcenverteilung an das IT-Personal ab.
- Die Auswahl der richtigen öffentlichen Cloud-Ressourcen ist komplex mit Millionen von Konfigurationsmöglichkeiten.
- Die meisten Cloud-Instanz-Optionen sind überproportioniert, um das Risiko von Performanceeinbußen zu minimieren.
- Die richtige Dimensionierung öffentlicher Cloud-Instanzen erfordert das Wissen um Anwendungsbedarfe.

Über Turbonomic, ein Unternehmen von IBM

Turbonomic, ein IBM Unternehmen, bietet Software für das Management von Anwendungsressourcen (Application Resource Management, ARM) zur Sicherstellung der Anwendungsperformance und -steuerung durch die dynamische Bereitstellung von Ressourcen für Anwendungen in Hybrid- und Multi-Cloud-Umgebungen an. Turbonomic Network Performance Management (NPM, Management der Netzleistung) bietet moderne Monitoring- und Analyselösungen, die Unternehmen, Netzbetreibern und Anbietern von Managed Services dabei helfen, eine kontinuierliche Netzwerkleistung in Netzwerken verschiedener Anbieter zu gewährleisten.

Für weitere Informationen siehe ibm.com/Cloud/turbonomic.

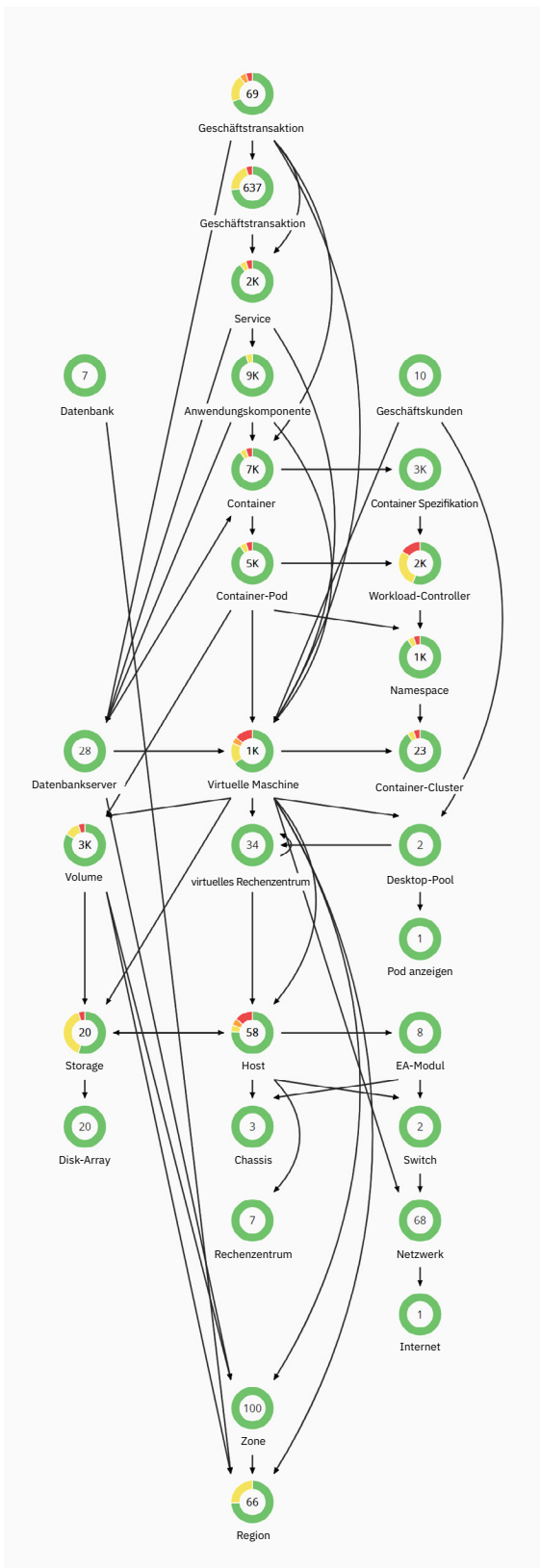


Abb. 2. Turbonomic erstellt automatisch die Topologie der Lieferkettenbeziehungen.

© Copyright IBM Corporation 2021

IBM Deutschland GmbH
IBM-Allee 1
71139 Ehningen

IBM Österreich
Obere Donaustraße 95
1020 Wien

IBM Schweiz
Vulkanstrasse 106
8010 Zürich

Hergestellt in den USA
November 2021

IBM, das IBM-Logo und IBM Cloud sind Marken oder eingetragene Marken der International Business Machines Corporation, in den USA bzw. anderen Ländern. Andere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM-Marken finden Sie unter ibm.com/trademark.

Turbonomic ist eine eingetragene Marke von Turbonomic Inc. einem Unternehmen von IBM.

Dieses Dokument ist zum Zeitpunkt der Erstveröffentlichung aktuell und kann von IBM jederzeit geändert werden. Nicht alle IBM Angebote sind in jedem Land, in welchem IBM tätig ist, verfügbar.

Vertragsbedingungen und Preise erhalten Sie bei den IBM Geschäftsstellen und/oder den IBM Business Partnern. Die Produktinformationen geben den derzeitigen Stand wieder. Für IBM Produkte gelten die Gewährleistungen, die in den Vereinbarungen vorgesehen sind, unter denen sie erworben werden.

¹ [IDC gibt weltweite Vorhersagen zur digitalen Transformation für 2021 bekannt](#), IDC, 29. Oktober 2020.

² [Zustand der FinOps Bericht 2021](#), The FinOps-Foundation, 2021.