Hewlett Packard Enterprise

Business Whitepaper

Künstliche Intelligenz für ein autonomes Rechenzentrum

Inside HPE InfoSight und die empfohlene Engine

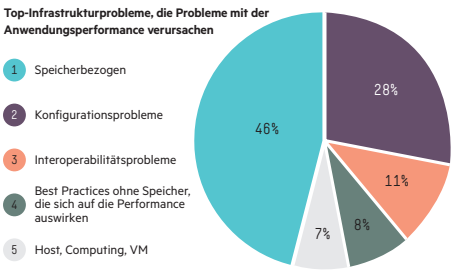
Inhaltsverzeichnis

Einführung

Die Infrastrukturverwaltung ist immer mit Schwierigkeiten, Unsicherheiten und Zeitverschwendung verbunden. Das liegt daran, dass IT-Experten Tage, Nächte und Wochenenden damit verbringen müssen, um Probleme zu lösen, die Anwendungen und Betriebsabläufe unterbrechen, um die Infrastruktur manuell zu optimieren. Außerdem wächst die Herausforderungen mit der Anzahl der Anwendungen und der Abhängigkeit von der Infrastruktur.

Glücklicherweise gibt es eine bessere Lösung. HPE InfoSight ist eine Lösung mit künstlicher Intelligenz (KI), die Probleme in der gesamten Infrastruktur vorhersieht und beseitigt und so für eine optimale Leistung sowie eine effiziente Ressourcennutzung sorgt.

In diesem Paper werden wir untersuchen, wie **HPE InfoSight** mit der Recommendation Engine den Weg für ein autonomes Rechenzentrum ebnet, damit die IT ihre Energie auf den Mehrwert für das Unternehmen konzentrieren kann.



Argumente für ein autonomes Rechenzentrum

Jeder Unternehmer ist sich des digitalen Wandels bewusst. Grundlage dafür ist jedoch die Notwendigkeit, dass die Infrastruktur Daten konsistent und zuverlässig den Anwendungen bereitstellt. Unternehmen können sich einfach keine Unterbrechungen oder Verzögerungen oder den Grad an High-Touch-Ressourcen leisten, der heute benötigt wird.

Speicherplattformen der nächsten Generation wie **Enterprise Flash Array von HPE Storage**, erhöhen die Speicher- und Applikations-Performance-Spezifikationen. Eine schnelle Speicherung allein kann jedoch keinen zuverlässigen, unterbrechungsfreien Zugriff auf die Daten gewährleisten oder die erforderliche manuelle Pflege überflüssig machen. Die Komplexität der Infrastruktur wirkt sich unweigerlich auf die Unternehmen und die Menschen auf, die sie verwalten.

So sehr die IT daran arbeitet, ihr Geschäft voranzubringen, so sehr hält sie die Infrastruktur weiterhin auf. Das Ergebnis ist ein endloser Zyklus von Break-Fix-Tune-Repeat.

Das herkömmliche Monitoring und die Unterstützung reichen nicht mehr aus.
Die IT-Abteilung hat sich schon immer auf Monitoring-Tools zur Fehlerbehebung in ihrer Umgebung verlassen. Leider hat dies dazu geführt, dass die Mitarbeiter Dutzende von Stunden damit verbracht haben, Logdateien zu prüfen und Diagramme zu interpretieren, um etwas Einblick in die Ursache einer Störung zu erhalten, um diese beheben zu können.

Wenn die Fehlersuche zu schwierig wird, wendet sich die IT-Abteilung an Anbieter. Branchenweit führt Support jedoch auch zu zeitraubenden, mehrstufigen Eskalationen.

Da die Infrastruktur immer wichtiger wird, wird dieses Modell nicht mehr ausreichen. Es ist nicht mehr akzeptabel, von einer Störung zu erfahren, nachdem sie aufgetreten ist. Unternehmen benötigen eine Lösung, die die Verwaltung und Unterstützung von Infrastrukturen verändert - eine Lösung, die Probleme vorhersagen kann, bevor sie auftreten.

Infrastrukturarbeiten sind aufwändig und mühsam.
Das fortwährende Sicherstellen der optimalen Leistung für jede Anwendung ist mit mühsamen manuellen Eingriffen verbunden. Für ständig wechselnde Arbeitslasten erfordert die **Feinabstimmung** in der Infrastruktur spezialisierte Ressourcen, die oft mit zeitraubendem Ausprobieren verbunden sind. Eine Überbereitstellung ist ein einfacher Ausweg, der aber mit Kosten für mehr als das, was benötigt wird, verbunden ist. Auch wenn sich die geschäftlichen Anforderungen nicht ändern, kann es **verpasste Gelegenheiten** geben, um die Performance mit vorhandenen Ressourcen zu verbessern. Möglicherweise würde das Verschieben einer Anwendung von einem AFA zu einem Hybrid oder das Ändern der Größe eines Volumens einen Unterschied ausmachen. Dies nicht zu wissen, kann ein enormer Kostenfaktor sein, der sich nicht zu Nutze gemacht wird.

Im Idealfall erhält die IT-Abteilung Empfehlungen dafür, was sie wann tun soll, um die Leistung und die verfügbaren Ressourcen zu optimieren.

KI überwindet Einschränkungen

Als Mensch können wir die Gegenwart sehen und uns ein wenig an die Vergangenheit erinnern, die mit den Tools übereinstimmt, die IT-Administratoren zur Verwaltung ihrer Umgebungen einsetzen. Um die Belastung durch die Verwaltung der Infrastruktur zu verringern, ist es jedoch erforderlich, dass Probleme gut vorhergesagt werden können, bevor sie auftreten, und dass umfassende Informationen über die zugrundeliegenden Workloads und Ressourcen verfügbar sind, um zu wissen, wie die Umgebung optimiert werden kann. Herkömmliche Tools reichen aus diesen Gründen nicht aus:

- **Sie können nicht voneinander lernen:** Analytics, die lediglich über lokale Systemmetriken berichten, sind nur begrenzt aufschlussreich, da das Verhalten von Tausenden von Peer-Systemen nicht für die Erkennung und Diagnose von Entwicklungsproblemen genutzt werden kann. Im Gegensatz dazu kann ein globaler Ansatz zur Datenerhebung und -analyse Beobachtungen aus einer immensen Vielfalt von Workloads zusammenfassen. So können seltene Ereignisse, die an einem Standort erkannt wurden, an einem anderen präventiv vermieden und häufiger auftretende Ereignisse früher und genauer erkannt werden.

- **Analytics beschränkt sich auf Infrastruktursilos:** Probleme, durch die Anwendungen unterbrochen werden, können irgendwo in der Infrastruktur ihre Ursache haben. Tools bieten einen Systemstatus pro Gerät an, der nur einen Teil des Gesamtbildes darstellt. Hingegen können Cross-Stack-Analytics, die mit mehreren Ebenen korrelieren einschließlich Anwendungen, Berechnung, Virtualisierung, Datenbanken, Netzwerke und Speicher, ein Gesamtbild schaffen.
- **Fehlende Domänenexpertise:** Vorhersagemodelle erfordern weitreichende Domänenerfahrungen—ein Verständnis aller Betriebsabläufe, Umgebungs- und Telemetrieparameter in jedem System des Infrastruktur-Stacks. Allgemeine Analytics-Verfahren können nur eine bestimmte Tiefe erreichen. Die Zusammenarbeit von Domänenexperten und KI kann Algorithmen für maschinelles Lernen erschaffen, mit denen die Ursache von historischen Ereignissen komplexeste und bedrohliche Probleme vorhersagen kann.
- **Handlungsunfähigkeit:** Im Idealzustand läuft der autonome Betrieb ohne menschliches Eingreifen. Dazu muss nicht nur bekannt sein, welche Änderungen vorgenommen werden müssen, um ein Problem zu vermeiden oder die Umgebung zu verbessern, sondern auch, sie müssen auch im Auftrag des Administrators durchführen werden können. Um diesen Automatisierungsgrad zu erreichen, bedarf es einer bewährten Historie von automatisierten Empfehlungen, die die notwendige Vertrauensebene schaffen können.

Die von künstlicher Intelligenz angetriebene Infrastruktur kann diese Einschränkungen durch den folgenden Rahmen überwinden:

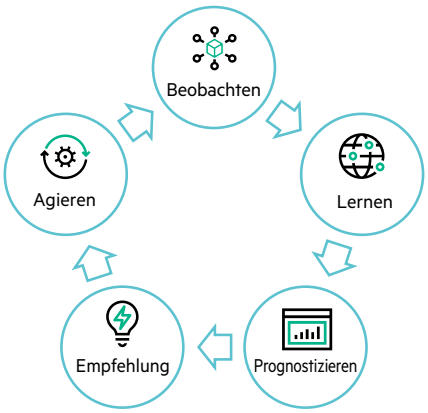


Abbildung 1. Künstliche Intelligenz für Infrastruktur-Framework

1. **Beobachten:** Durch die gleichzeitige Überwachung aller Systeme in einer installierten Basis entwickelt KI ein stationäres Verständnis der idealen Betriebsumgebung für jedes Workload und jede Anwendung. Dann kann abweichendes Verhalten durch Erkennung der unterstrichenen E/A-Muster und Konfigurationen jeder Umgebung erkannt werden.
2. **Lernen:** Tiefe Systemtelemetrie in Verbindung mit globaler Konnektivität schafft eine Datengrundlage, die die Erfahrungen jedes angeschlossenen Systems nutzt. Maschinelles Lernen in der Cloud beschleunigt Wissen und allgemeines Lernen der KI.
3. **Vorhersagen:** Für jedes neu entdeckte Problem kann die KI lernen, das Problem vorherzusagen und Pattern-Matching-Algorithmen zu verwenden, um festzustellen, ob ein anderes System in der installierten Basis anfällig ist. Darüber hinaus kann die Anwendungsperformance modelliert und auf neue Infrastrukturen abgestimmt werden, die auf historischen Konfigurationen und Workload-Mustern basieren.
4. **Empfehlen:** Auf Basis von Predictive Analytics ermittelt die KI die geeignete Empfehlung zur Verbesserung und Sicherstellung der idealen Umgebung. Empfehlungen sind operative Systementscheidungen, die die IT entlasten und das Rätselraten bei der Verwaltung ihrer Infrastruktur beenden.
5. **Umsetzen:** Durch das gegenseitige Vertrauen zwischen Infrastruktur und KI können Empfehlungen automatisch im Auftrag der IT-Administratoren umgesetzt werden. Wenn keine Automatisierung verfügbar ist, können spezifische Empfehlungen durch die Automatisierung von Supportfällen gegeben werden.

KI kann Ihre Infrastruktur überwachen, kontinuierlich von einer globalen installierten Basis lernen und das Gelernte anwenden, um Probleme vorherzusagen und zu vermeiden und das Rätselraten bei der Verwaltung der Infrastruktur zu beenden. KI kann die Infrastruktur intelligenter und zuverlässiger machen.

Vorteile von HPE InfoSight:
86% der Probleme werden automatisch vorhergesagt und gelöst[1]

99,9999 % nachgewiesene Verfügbarkeit[2]

79% Reduzierung des IT-Speichers OPEX[3]

85% weniger Zeitaufwand für Storage-Probleme[4]

Globales Lernen

KI und maschinelles Lernen erfordern enorme Datenmengen, die über die begrenzten Protokolle und Metriken herkömmlicher Hardwareplattformen hinausgehen. Die HPE-Speicherplattformen, die **Intel® Xeon® Prozessoren und SSDs** nutzen, wurden mit tiefenddiagnostischen Sensoren entwickelt. Da HPE InfoSight diese Daten seit 2010 sammelt, ergibt sich aus der Breite der Telemetrie ein architektonischer Vorteil.

HPE InfoSight: KI im Rechenzentrum

HPE InfoSight wurde aus der Überzeugung heraus gegründet, dass Infrastrukturmanagement und Support weiterentwickelt werden müssen. Anstatt sich mit unerwarteten Problemen und reaktivem Anbieter-Support zu befassen, sollte die KI die Infrastruktur so intelligent machen, dass sie Probleme frühzeitig erkennen und ohne menschliches Zutun lösen kann. Nur in diesem Self-Healing-Modell können Unternehmen ihre Ressourcen am effizientesten nutzen, um ihr Geschäft voranzutreiben.

HPE InfoSight ist eine KI-Plattform, die das Rechenzentrum autonom macht. Aufbauend auf einem einmaligen Datenerfassungs- und -analyseansatz sammelt und analysiert HPE InfoSight in jeder Sekunde Millionen von Sensordatenpunkten aus unserer weltweit vernetzt installierten Basis. Diese Sensordaten bieten umfassende Messungen des Betriebs und des Zustands jedes Systems, Subsystems und der umgebenden IT-Infrastruktur. Aus diesen Daten werden **Predictive Analytics** und **Recommendation Engines** erlernt, was unsere Kunden in erheblichem Maße betreffen kann.

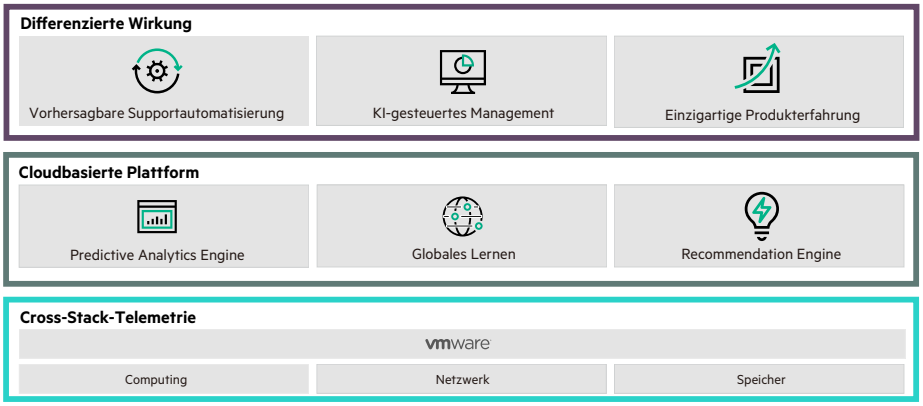


Abbildung 2. HPE-InfoSight-Plattform

Predictive Analytics Engine

Vorausschauend, um Störungen zu beseitigen und die IT voranzubringen.

HPE InfoSight bietet Predictive Analytics, die sich über den gesamten Lebenszyklus der Infrastruktur erstrecken - von der Planung bis zur Erweiterung.

- **Für die Planung:** Die neue Infrastruktur wird durch die Vorwegnahme der Leistung und der benötigten Ressourcen auf der Grundlage verschiedener Anwendungen, die in unserer installierten Basis zu sehen sind, angepasst. Durch Telemetrie aus den eingesetzten Systemen verfeinert HPE InfoSight kontinuierlich seine maschinell erlernten Modelle für eine bessere Größengenauigkeit.
- **Sobald Arrays bereitgestellt werden:** Predictive Analytics transformiert das Produkt- und Supporterlebnis. HPE InfoSight sucht ständig nach Frühindikatoren für Probleme und löst diese automatisch, bevor der Kunde überhaupt bemerkt, dass es ein Problem gibt. Wenn HPE InfoSight ein neues Problem erkennt, lernt es, das Problem vorherzusagen und andere Systeme in der Installation vor demselben Problem zu bewahren.
- **Den Lebenszyklus abschließen:** HPE InfoSight prognostiziert den zukünftigen Kapazitäts-, Leistungs- und Bandbreitenbedarf auf Basis historischer Nutzung und autoregressiver sowie Monte-Carlo-Simulationen.

[1] "**Ein neuer Maßstab für die Systemverfügbarkeit**," 2017

[2] "**HP Get 6-Nines-Garantie**," 2017

[3, 4] "**Die finanziellen Auswirkungen von HPE InfoSight Predictive Analytics**," 2017

Predictive Analytics geht über die reine Speicherung hinaus
Die Prognosefähigkeiten von HPE InfoSight gehen über die Speicherung hinaus.

Beispielsweise hat HPE InfoSight eine katastrophale All-Path-Down-Situation für HPE Nimble Storage-Kunden aufgrund eines möglichen Problems mit einer Netzwerk-VIC-Karte im Host vorhergesagt und verhindert. Unter Einsatz von HPE InfoSight, HPE Nimble Storage erkannten die Supportmitarbeiter von Nimble, dass der Fibre Channel-Wiederherstellungsmechanismus aufgrund eines Double-Abort-Problems in der Karte ausfallen könnte. HPE InfoSight verwendete einen Signaturmuster-Algorithmus, um 100 Kunden zu identifizieren, die für dieses Problem anfällig sind, und wendete einen Workaround an, der das Problem verhinderte.

Wie bei **HPE Nimble Storage** gezeigt, prognostiziert und löst HPE InfoSight automatisch 86 % der Probleme. Dies bedeutet eine Reduzierung der IT-Betriebskosten um 79 %, 85 % weniger Zeitaufwand für Storage-Probleme und über 99,9999 % der nachgewiesenen Verfügbarkeit der gesamten HPE Nimble Storage-Installationsbasis.

Das Aufkommen der selbstfahrenden Autos.
Empfehlungsmaschinen werden branchenübergreifend eingesetzt, um alles zu automatisieren und zu optimieren, von Online-Shopping-Carts bis hin zu Geschäftsprozessen. Ein Bereich, in dem sie einen sehr großen Einfluss haben, ist die Entwicklung von selbstfahrenden Autos. Empfehlungen teilen einem fahrerlosen Auto mit, wie schnell es fahren kann, wann es bremsen muss und wie Kollisionen vermieden werden. Die richtige Empfehlung zur richtigen Zeit kann einen Unfall vermeiden oder geschehen lassen.

Recommendation Engine

Müheloses Management der Infrastruktur

Für eine autonome Infrastruktur benötigt HPE InfoSight die Fähigkeit, nicht nur zu sehen, was vor uns liegt, um Probleme vorherzusagen, sondern auch um dynamisch intelligente Empfehlungen und Entscheidungen zu treffen, die jede Umgebung proaktiv verbessern und optimieren. Es muss anwendungsbewusst sein, um die richtige Empfehlung zur richtigen Zeit zu bedienen, ohne andere Anwendungen zu beeinträchtigen.

Durch eine Recommendation Engine baut HPE InfoSight die Prognosefähigkeiten aus, um der IT-Abteilung automatisch mitzuteilen, wie sie Probleme vermeiden, die Leistung proaktiv verbessern und Ressourcen optimieren kann. Die Engine berät quasi auf der Grundlage der Erfahrungen aus seiner Wissensbasis.

Da viele der schwierigeren Aufgaben im Infrastrukturmanagement mit der Systemleistung zusammenhängen, werden wir uns genauer ansehen, was die Recommendation Engine für das Performance Management leistet.

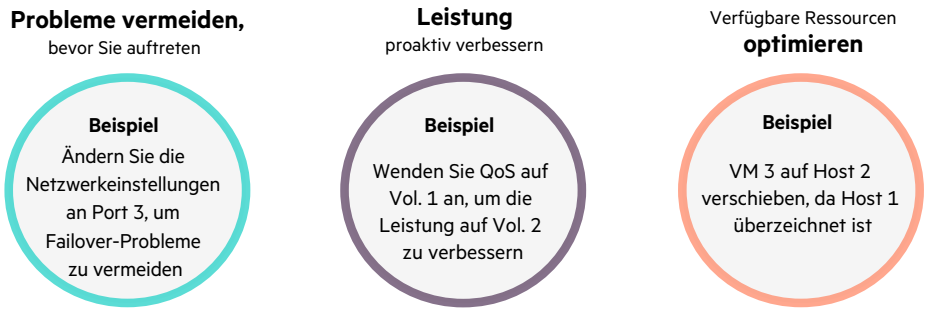


Abbildung 3. Vorteile der HPE InfoSight Recommendation Engine

KI-Leistungsempfehlungen
Die Realität heute ist, dass die Sicherstellung optimaler Leistung eine zeitraubende und kostspielige Angelegenheit ist. Sie ist zum einen reaktiv, da Probleme mit der Anwendung unerwartet auftreten. Zum anderen sind die Analysen anderer Tools nicht intelligent genug, um zu verstehen, warum ein Problem aufgetreten ist und wie man es lösen kann. Drittens ist es problematisch, da es zu viel manuelles Tuning und Rätselraten gibt.

Durch fortgeschrittenes maschinelles Lernen identifiziert die Recommendation Engine in HPE InfoSight Möglichkeiten zur Leistungssteigerung basierend auf E/A-Workload-Mustern, ermittelt genau die Variablen mit der größten Wirkung und liefert proaktiv die richtige Empfehlung zur Leistungssteigerung. Die Recommendation Engine nimmt Ihnen das Rätselraten ab und optimiert Leistung und Ressourcen.

Entwicklung der Recommendation Engine

In diesem Abschnitt werfen wir einen genaueren Blick auf die Recommendation Engine. Wir stellen das Design und die Architektur vor.

Abbildung 4 zeigt den Problem Space von möglichen Infrastrukturproblemen auf einem Balkendiagramm mit der Art der Probleme und der Häufigkeit, die jeweils beschriftet sind. Die Probleme lassen sich in der Regel in zwei Kategorien einteilen, nämlich **einfach und verbreitet**, die grau markiert sind und **komplex und einmalig**, die blau markiert sind—und eine Pareto-Verteilung bilden. Es ist wichtig, die Belastungskurve zu beachten, die mit der Art der Probleme korreliert.

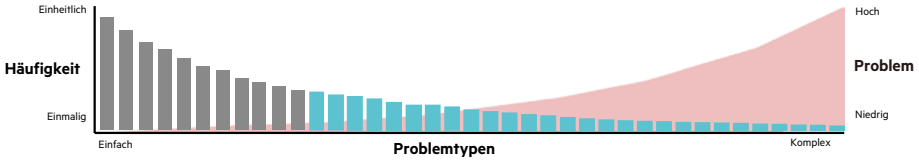


Abbildung 4. Problemraum korreliert mit dem relativen Belastungsindex

Einfache und verbreitete Probleme, wie z.B. ausgefallene Laufwerke, sind häufiger, machen aber nur einen kleinen Prozentsatz der Belastung für IT-Administratoren aus. Die Häufigkeit dieser Probleme macht es einfacher, sie vorherzusagen und mit einer automatisierten Lösung zu beheben. Die Realität in IT-Umgebungen ist jedoch, dass Probleme sehr unterschiedlich sein können, und es sind die Probleme, die komplex und einzigartig sind - diejenigen, die unerwartet auftauchen und zahlreiche Personen und Ressourcen erfordern, um sie zu lösen -, die für die größten Belastungen verantwortlich sind.

Unternehmen müssen das gesamte Spektrum an Problemen kennen, von den grundlegendsten bis hin zu den komplexesten und vielseitigsten, den vorhergesagten und automatisch gelösten. Einfache Probleme lassen sich erkennen, indem man sich nur wenige Daten qualitativ mit fest programmierten Regeln zur Auslösung von Ereignissen und Alarmmelden ansieht. Während andere Anbieter behaupten, Empfehlungen anzubieten, beschränkt sich ein Großteil ihrer Kenntnisse in der Regel auf die Lösung von Problemen auf der einfachen und verbreiteten Möglichkeiten.

Bei komplexen und einzigartigen Problemen steigen die Anzahl der Variablen und die für eine diagnostische Bestimmung erforderliche quantitative Präzision nahezu exponentiell an. Je komplexer die Probleme werden, desto fehleranfälliger und ineffizienter werden hartcodierte Regeln mit zahlreichen quantitativen Variablen. Selbst die talentiertesten Expertinnen und Experten kämpfen mit Herausforderungen, die über das einfache Schwellenverhalten für quantitative Probleme hinausgehen (z.B. sollte dieses Problem ausgelöst werden, wenn Sensor X über die Schwelle Y steigt). Und allzu oft werden auch diese Lösungen eher aus anekdotischen Erfahrungen als aus einer gründlichen Analyse abgeleitet.

Die HPE InfoSight Recommendation Engine geht über die einfachen und häufigen Probleme hinaus, um die komplexen und einzigartigen Probleme zu erkennen und zu vermeiden. Mit KI und maschinellem Lernen können wir langfristige Prozesse aufnehmen und Empfehlungen zur Vermeidung von Geschäftsunterbrechungen geben.

Entwurfsmethodik für die KI-Leistungsempfehlungen

Der Aufbau einer stabilen Recommendation Engine für die Performance erfordert die Beantwortung einiger wichtiger Fragen.

Sensoren alleine haben nicht genug Kontext
Analog zu einem biometrischen Screening gibt es keinen abschließenden Weg zu wissen, ob eine systolische Blutdruckmessung von 133 mmHg problematischer ist als 121 mmHg, ohne dass viel mehr Kontext und Daten über den menschlichen Körper vorhanden sind. Daher ist die Diagnose und Empfehlung für Werte zwischen 120 und 139 mmHg gleich und basiert auf Annahmen und nicht auf wissenschaftlichen Überlegungen. Folglich ist es falsch anzunehmen, dass eine durchschnittliche Leselatenz von 10 ms einen größeren Einfluss auf die Performance hat als 5 ms, da dieser Metrik der vollständige Kontext fehlt.

Frage 1: Sind Performance-Metriken tatsächlich ein genauer Indikator für ein nicht optimiertes System oder ein potenzielles Problem?

Sensoren sammeln Echtzeitmessungen ihrer Umgebung mit dem Ziel, Ereignisse oder Veränderungen zu erkennen. Normalerweise verlassen sich IT-Administratoren auf den Wert dieser Sensoren (d. h. Leselatenz, Schreiblatenz, IOPS, Durchsatz und andere), um festzustellen, ob das entsprechende Verhalten problematisch ist. Dieser Ansatz ist jedoch fehlerhaft, da Sensoren allein nicht über vollen Kontext verfügen, um festzustellen, ob ihre Werte wirklich einen Einfluss auf die Anwendung und das Endkundenerlebnis haben.

Unterschiedliche Workloads und Anwendungen haben unterschiedliche Leistungsmerkmale und Auswirkungen für das Endkundenerlebnis. Beispielsweise sind große Blockoperationen wie Backup-Jobs natürlich latenter, aber weniger reaktionssensitiv als transaktionale Workloads. Die Annahme einer höheren Latenz bedeutet, dass es Probleme gibt, die zu Fehlalarmen und Zeitverlusten bei der Verfolgung falscher Ereignisse führen - ein grundlegendes Problem für das Event-Management.

Design-Ansatz

Wie hoch die Latenz wirklich ist, hängt von der Empfindlichkeit der zugrundeliegenden Anwendung ab. Mit Hilfe der globalen Systemtelemetrie in HPE InfoSight haben wir maschinell erlernte Modelle mit typischer Leistung entwickelt, um Ereignisse, die für den Benutzer tatsächlich von Bedeutung sind, genauer zu identifizieren. Wir haben diese Modelle mit Hilfe von Kundenfalldaten validiert, die die **potenzielle Wirkung**, auch **Latenzschweregrad** genannt, wiedergeben, die sich negativ auf die Leistung auswirken kann.

Ergebnis

Wie in Abbildung 5 dargestellt, versteht HPE InfoSight den wahren Einfluss der Latenz und liefert einen Schweregradindex innerhalb eines definierten Zeitrahmens als orangefarbene und zugehörige Zahlenwerte (1 bis 10). Dunklere Orangetöne deuten auf eine höhere potenzielle Latenz hin.

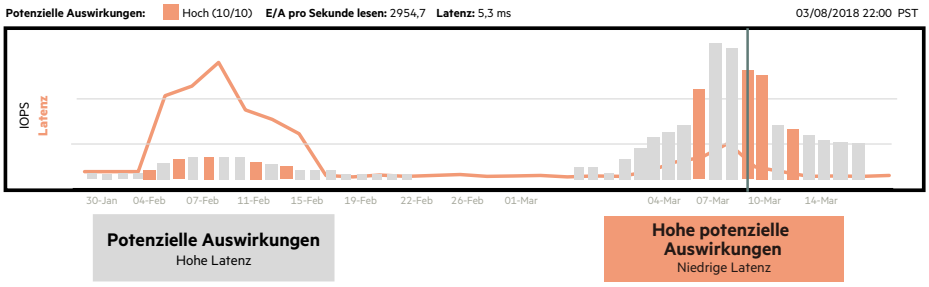


Abbildung 5. Historische IOPS-Aktivität mit potenziellen Auswirkungen in orangefarbener Farbe und einem numerischen Latenzschweregrad.

Diese Visualisierung filtert das Rauschen und ermöglicht es IT-Administratoren, sich nur auf die wichtigen Ereignisse zu konzentrieren. Das Ergebnis ist die Eliminierung von Fehlalarmmeldungen sowie die Fähigkeit zu erkennen, wann die Leistung verbessert werden kann.

Frage 2: Welche Faktoren können die Leistung der Anwendung aufgrund der auf dem System ausgeführten Workloads beeinflussen und in welchem Umfang?

Jetzt, da wir wissen, ob und wann Sensormessungen ein nicht optimiertes System anzeigen, wird im nächsten Schritt die Ursache dafür bestimmt.

Traditionell durchlaufen IT-Administratoren eine Reihe von Test- und Fehlertests, um ein Performance-Problem zu lösen, in der Hoffnung, dass etwas funktioniert und das Problem fernhält. Dieses Rätselraten ist jedoch zeitaufwendig und löst das Problem oft nicht auf unbestimmte Zeit, wenn überhaupt.

Design-Ansatz

Um sicherzustellen, dass unser System Probleme im gesamten Problembereich mit hoher Präzision erkennen kann, integrieren wir die Analyse aus den in Frage 1 beschriebenen Modellen mit zwei Arten von maschinellen Lernmodellen: **fachkundig geschult** und **weltweit geschult**. Die von Experten geschulten Modelle werden anhand spezifischer Beispiele von seltenen Ereignissen, die von unseren Support-Technikern benannt wurden, geschult und validiert. Weltweit geschulte Modelle werden mit unserer installierten Telemetrie geschult und bewertet, um ungewöhnliche Probleme zu erkennen, indem nach erwarteten Korrelationen mit der Latenz gesucht wird oder wenn ein System im Vergleich zu den Erwartungen hinter den Erwartungen zurückbleibt.

Dieser hybride Ansatz stellt sicher, dass HPE InfoSight in der Lage ist, komplexe und einzigartige Probleme zu lösen.

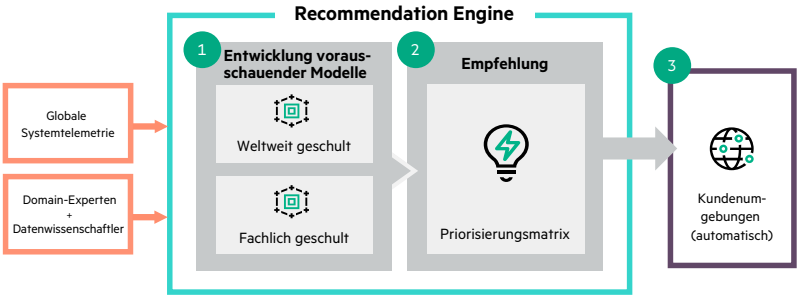


Abbildung 6. Architektur für die empfohlene Engine

Von Experten geschulte Modelle

Unsere von Experten geschulten Modelle sind Klassifikatoren, die von Menschen gekennzeichnete Fälle eines Problems verwenden, die in der Praxis beobachtet wurden. Durch den Unterstützungsprozess trainieren unsere Datenwissenschaftler Klassifikatoren, um neue Fälle dieser Ereignisse im Feld ohne menschlichen Eingriff und mit hoher Genauigkeit zu erkennen. Die Erweiterung der Telemetrie mit menschlichen Kennzeichnungen stellt sicher, dass das System auch bei ungewöhnlichen Ereignissen eine korrekte Diagnose und Empfehlung ausgeben kann.

Weltweit geschulte Modelle

Von Experten geschulten Modelle eignen sich gut, um Bedingungen zu identifizieren, die einzeln wahr oder falsch sind, aber nicht gut für Probleme geeignet sind, die mehrere Ursachen haben können, und in unterschiedlichem Maße zusammen auftreten. Wenn auf einem bestimmten System mehrere unterschiedliche Latenzfaktoren erkannt werden, ist es wichtig, mit eine konsistenten Methode zu bestimmen, welche davon am stärksten für das erkannte Problem verantwortlich sind. In dieser Situation wird ein menschlicher Experte wahrscheinlich keine Schulungsbeispiele in ausreichender Menge erstellen. Stattdessen schulen wir Modelle mit unserer globalen installationsbasierten Telemetrie, um quantifizieren zu können, wie verschiedene Latenzquellen (oft nichtlinear) zu einer beobachteten Latenz beitragen. Mit diesem Modell können wir ermitteln, welche Probleme zuerst gelöst werden müssen. Die Breite und der Umfang unserer Telemetrie erlaubt es uns, sehr umfassende Diagnosemodelle zu erstellen, die sonst unmöglich zu schulen wären.

Ergebnis

Unser hybrider Maschinenlernansatz verbessert kontinuierlich die Genauigkeit des Fehlerüberwachungssystems und dessen Reichweite und minimiert so unbekannte Probleme. Das Ergebnis ist eine genaue Ursachendiagnose für jedes System in unserer installierten Basis.

Gründe, die für maschinelles Lernen sprechen

Das maschinelle Lernen eignet sich ideal für Probleme, die die Untersuchung mehrerer quantitativer Variablen gleichzeitig erfordern und Signaturen erfordern, die keine prägnante qualitative Beschreibung enthalten. Von Menschen erstellte Regeln sind nicht so gut geeignet, diese Probleme auf die gleiche Weise zu lösen, so wie es für einen Menschen unwahrscheinlich wäre, einen Code zu schreiben, der bestimmt, ob eine Matrix von Pixeln mit dem Gesicht einer bestimmten Person übereinstimmt.

Multivariate Analyse

Expertengeschulte Klassifikatoren haben sich z. B. bei der Identifizierung von Instanzen mit **SSD Bandbreitensättigung als nützlich erwiesen:** ein ungewöhnliches Ereignis, bei dem ein hoher E/A-Durchsatz auf SSDs geleitet wird. Dieses Szenario ist interessant, weil wir festgestellt haben, dass die Betrachtung einer SSD-Metrik (z.B. Latenz, Warteschlangentiefe, IOPS, MB/s, Anteil der letzten aktiven Millisekunden usw.) völlig unzureichend ist, um genau zu bestimmen, ob die SSD Upstream-Performance-Probleme verursacht. Stattdessen müssen mehrere Proben dieser Metriken gleichzeitig untersucht werden. Betrachtet man nur eine Metrik, so ergibt sich eine Heuristik, die entweder eine große Anzahl von False Positives erzeugt (geringe Genauigkeit) oder einen großen Teil der problematischen Ereignisse nicht identifiziert (geringer Rückruf). Um ein Modell herzustellen, das gleichzeitig eine hohe Präzision und einen hohen Wiedererkennungswert erreichen konnte, war ein multivariates, maschinengelerntes Modell erforderlich. Aufgrund der quantitativen Komplexität des Problems war unser maschinengelernter Klassifikator in der Lage, dieses Problem viel effektiver zu erkennen als jede der vorhergehenden menschlichen Heuristiken.

Frage 3: Was ist die richtige Empfehlung für eine Leistungssteigerung?

Aus den Ergebnissen der Fragen 1 und 2 kann HPE InfoSight ermitteln, ob es Möglichkeiten gibt, das Kundenumfeld zu verbessern. Der Ansatz zu Frage 3 führt dazu, dass HPE InfoSight IT-Administratoren automatisch mitteilt, was sie tun sollen, um die Situation zu verbessern.

Design-Ansatz
Die einfachste, aber ineffizienteste Empfehlung ist, die Notwendigkeit eines Hardware-Upgrades zu empfehlen und den Kunden einfach zu mitzuteilen, dass die Ressourcen über die physischen Grenzen hinausgehen und daher größere Hardware benötigt wird. Im Gegensatz dazu bietet HPE InfoSight eine viel umfangreichere Reihe von Empfehlungen, einschließlich, aber nicht beschränkt auf QoS-Limits, Software-Updates, Workload-Änderungen, Konfigurationsänderungen und Hardware-Upgrades. HPE InfoSight beinhaltet Anwendungen, Ressourcen und Präferenzen (z.B. Tageszeit und Wochentage, Latenzempfindlichkeit) für jedes System. Und es nutzt dieses Verständnis, um die Empfehlungen zu priorisieren.

Details, die mit der Empfehlung geliefert werden, informieren diejenigen Workload-Teile der Benutzer, die eine gesättigte Ressource verbrauchen (z.B. die Einheiten, die die meiste Speicher-CPU verwenden, wenn das Array CPU-gebunden ist). Diese Details sind von entscheidender Bedeutung, da sie es dem Benutzer ermöglichen, zwischen einer workloadbasierten Problembehebung (d.h. Drosselung der Volumenaktivität oder anderweitige Dämpfung der Volumenanforderungen) und einer Hardwarekorrektur (d.h. Hinzufügen von Hardware zum System, um die Fähigkeiten des Systems zu erweitern und den Ressourcenengpass zu verringern) zu entscheiden.

Ergebnis
Vor HPE InfoSight mussten sich IT-Administratoren mit dem Problem der Verwaltung der Speicherleistung auseinandersetzen. Dies war ein reaktiver Prozess, der viel Zeit in Anspruch nahm, um Diagramme und Protokolle zu interpretieren und die Infrastruktur manuell zu optimieren.

Mit der Recommendation Engine müssen sich die Kunden einfach keine Sorgen mehr um die Performance machen. HPE InfoSight informiert die IT-Abteilung, wenn es eine Möglichkeit gibt, die Leistung zu verbessern und sagt ihnen, was sie tun sollen. Sie können ihre Speichersysteme anspruchsvoll betreiben, mehrere Anwendungen konsolidieren und brauchen sich keine Sorgen darüber machen, ob sich die Anwendungen aufgrund der Infrastruktur verlangsamen. Sie wissen, dass ihr System in einem optimalen Zustand ist.

Zusammenfassend lauten die von HPE InfoSight generierten Empfehlungen:

- **Automatisch:** Jederzeit für jeden Kunden weltweit verfügbar
- **Präventiv:** Sieht Engpässe voraus, bevor sie sich auf Unternehmen auswirken können.
- **Umfassend:** Mit maschinellem Lernen komplexe und einzigartige Probleme vorhersagen
- **Bindend:** Jenseits von Hardware-Upgrades und spezifischen Betriebsänderungen

Der Weg zu einem autonomen Rechenzentrum

Unternehmen müssen heute den ununterbrochenen Zugriff auf Daten für alle ihre wachsenden Anwendungen sicherstellen. Dies wird jedoch mit der Komplexität der Infrastruktur und den Anforderungen an die begrenzten Ressourcen immer schwieriger. CIOs können es sich nicht mehr leisten, durch ihre Infrastruktur aufgehalten zu werden.

Unsere Vision ist nichts anderes als ein autonomes Rechenzentrum, das keine ständige Aufmerksamkeit, kein manuelles Tuning und keine reaktive Fehlersuche mehr benötigt. Dies ist ein Rechenzentrum, in dem sich die Infrastruktur selbst verwaltet, selbst wiederherstellt und optimiert. Dies mag zwar weit von der Realität entfernt erscheinen, aber Unternehmen mit einer Infrastruktur von **HPE InfoSight** können diese Vision eher früher als später verwirklichen. Und der Schlüssel ist künstliche Intelligenz.

Als die fortgeschrittenste künstliche Intelligenz der Branche hat HPE InfoSight die Verwaltung und Unterstützung der Infrastruktur grundlegend verändert. Durch Cloud-basiertes maschinelles Lernen werden Probleme vorhergesagt und verhindert, während gleichzeitig die optimale Leistung und Verfügbarkeit der unterstützten Infrastruktur bereitgestellt wird. Und mit fast einem Jahrzehnt Erfahrung und Lernen ist HPE InfoSight immer noch anspruchsvoller und kompetenter.

Aufbauend auf diesen prädiktiven Fähigkeiten bringt uns die Recommendation Engine in HPE InfoSight noch näher an ein autonomes Rechenzentrum heran. Anstatt auf Probleme zu reagieren oder herauszufinden, wie man Ressourcen am besten verwaltet, sieht HPE InfoSight Entwicklungen voraus und sagt den Kunden genau, was zu tun ist, um Probleme zu vermeiden und die Umgebung zu verbessern. Diese Empfehlungen treffen heute intelligente Entscheidungen, die in Zukunft automatisch für unsere Kunden umgesetzt werden können.

Weitere Informationen unter
hpe.com/us/en/storage/infosight.html

Melden Sie sich noch heute an.

a00044051DEE, Mai 2018, Rev. 1