

Zentrum für Mikrobiom-Innovation (CMI) überwindet Daten-Engpässe und beschleunigt Entdeckung neuer Methoden zur Verbesserung der menschlichen Gesundheit und der Umwelt

CMI nutzt Panasas ActiveStor zur Entwicklung neuartiger Werkzeuge und Methoden für die Analyse und Manipulation von Mikrobiomen

ZUSAMMENFASSUNG

KUNDE

Zentrum für Mikrobiom-Innovation der Universität Kalifornien (CMI)

GRÖSSE

Mehr als 150 Labors auf dem gesamten Campus

STANDORT

San Diego, Kalifornien, USA

INDUSTRIE

Life Sciences

HERAUSFORDERUNG

- Unterstützung datenintensiver Forschungsarbeiten mit der Notwendigkeit einer Analyse von wachsenden Datenmengen
- Ausführen einer ständig wachsenden Anzahl E/A-intensiver Aufgaben ohne Beeinträchtigung der Systemleistung
- Nutzung modernster wissenschaftlicher Technologien zur Unterstützung innovativer Forschungsmethoden

LÖSUNG

- Unterstützung einer schnellen Analyse, Speicherung und Abfrage großer Datenmengen bei gleichbleibend hoher Leistung
- Bessere Kontrolle der Speicherressourcen bei flexibler Bewältigung unterschiedlichster Arbeitslasten
- Einfachere Wartung und Optimierung des Speichers durch benutzerfreundliche Verwaltung

ERGEBNISSE

- Bessere Benutzerfreundlichkeit, keine Speicherengpässe und Systemverzögerungen mehr für Wissenschaftler.
- Schnellere Untersuchung und Ermittlung von Daten für wissenschaftliche Innovationen
- Bessere Kontrolle und Benutzerfreundlichkeit der Computerumgebung gewährleisten, dass Ressourcen verfügbar sind, wenn sie benötigt werden
- Einfachere Speicherverwaltung für Administratoren



“Die Panasas-Technologie unterstützt die Mission unseres Zentrums, weil sie unsere Untersuchungen niemals beeinträchtigt”

Yoshiki Vázquez-Baeza

Associate Direktor des Bioinformatik-Integrationszentrums der Universität von Kalifornien San Diego Zentrum für Mikrobiom-Innovation

Mikrobiome sind die speziellen, vielfältigen Gemeinschaften aus Bakterien, Viren und anderen Mikroorganismen, die in, auf und um uns herum leben. Sie helfen uns bei der Verdauung und Verarbeitung von Nährstoffen, interagieren mit unserem Immunsystem und übernehmen eine erstaunliche Rolle in unserem Leben und unserer Umwelt – viele dieser Funktionen sind noch nicht erforscht.

Das Center for Microbiome Innovation (CMI) gehört zur Universität von Kalifornien in San Diego und arbeitet daran, die Forschung und das Verständnis von Mikrobiomen zu vertiefen. Das CMI verfügt über modernste Technologien und Fachkenntnisse in den Bereichen klinische Medizin, Bioingenieurwesen, Informatik sowie biologische und physikalische Wissenschaften. Es betreibt auch zentrale Einrichtungen für Datenerzeugung, Bioinformatik und Datenanalyse.

Das Zentrum nutzt hochmoderne Technologien wie Genomik und Metagenomik und verarbeitet jedes Jahr hunderttausende Materialproben von Menschen, Tieren, Lebensmitteln und der bebauten Umwelt. Dank dieser modernen Technologien und mit Hilfe von wissenschaftlicher Instrumente wie DNA-Sequenzierern und Massenspektrometern können CMI-Forscher, Doktoranden und Postdoktoranden in verschiedenen Forschungslabors neue Instrumente und Methoden zur Verbesserung der menschlichen Gesundheit und zum Nutzen der Umwelt entwickeln.

Die genomische Sequenzierung und Analyse von Mikrobiomen erfordert riesige Datenmengen. Dazu erhält jeder der fast 300 Benutzer im CMI lokalen Speicherplatz über zwei Cluster. Allerdings reichen die vorhandenen Ressourcen heute nicht mehr für die Bewältigung umfangreicher E/A-Aufgaben wie die Verarbeitung von DNA-Sequenzen oder die Analyse anderer Multi-Terabyte-Datensätze aus. Versuche, die aktuellen Datenmengen mit herkömmlichen Speichertechnologien zu verarbeiten, führten zu Leistungseinbußen bei den Rechenabläufen und zu Beeinträchtigungen der Auswertungsgeschwindigkeit.

Vor drei Jahren begann eine der wichtigsten Forschungsgruppen des Zentrums mit der Durchführung groß angelegter Metagenomik-Sequenzierungen, einer fortschrittlichen Technologie für mikrobi-

elle Untersuchungen. "Das Ausmaß dieser Art von Berechnungen kann alle anderen Aktivitäten blockieren" erklärt Yoshiki Vázquez-Baeza, Associate Direktor für bioinformatische Integration am CMI. "Das war der Auslöser – wir wussten, dass wir etwas ändern mussten."

„Oft müssen unsere Wissenschaftler für eine Analyse mehrere hundert Dateien zu einer großen Datei verknüpfen“, erklärt Jeff DeReus, Systemadministrator am CMI. "Der Versuch, so viele Lesevorgänge gleichzeitig durchzuführen, würde andere Speichersysteme in die Knie zwingen. Die Leseleistung der Festplatten reichte nicht aus, um diese Art von Arbeitslast neben den anderen täglich im Cluster auszuführenden Aufgaben zu bewältigen".

Datenintensive Unterstützung für Innovation

Nach Prüfung anderer Speichertechnologien und der Evaluierung von Anbietern entschied sich das CMI für eine Panasas ActiveStor® Hochleistungs-Speicherlösung. Systemadministrator DeReus hatte schon früher an Hochschulinstituten mit Panasas-Speichertechnologie gearbeitet und war mit ihrer überlegenen Funktionalität, Verwaltbarkeit und Skalierbarkeit vertraut. "Ich hatte keinerlei Bedenken bezüglich der Wahl von Panasas", sagt er.

Mit der neuen Speicherlösung konnte das CMI eine Basis für kontinuierliche Spitzenleistungen in der Forschung schaffen. "Unsere Teams vergleichen oft mehrere Methoden oder überarbeiten die Benchmarks anhand neuer Datensätze", sagt Vázquez-Baeza. "Wenn wir die Benchmarking-Methoden aufgrund von Speicherproblemen einschränken müssten, könnten wir nicht die gesamte Bandbreite wissenschaftlicher Optionen ausloten", so Vázquez-Baeza. "Die Panasas-Technologie unterstützt den Auftrag unseres Zentrums, da sie unsere Forschungsarbeit niemals einschränkt".

Hohe Arbeitsgeschwindigkeit

Mit Panasas können die CMI-Forscher bisher unerreicht große Datenmengen schnell speichern, abrufen und analysieren. Das schnelle, effiziente und parallele PanFS®-Dateisystem beschleunigt die Arbeit in allen Phasen des rechnergestützten Forschungsprozesses.

"Durch die Bereitstellung der Daten in einem parallelen Hochgeschwindigkeits-Dateisystem wie PanFS können wir die

eigentliche Arbeitslast im Speichercluster verringern", sagt DeReus. "Mitarbeiter können ihre Zwischendateien generieren und die erforderlichen Analysen durchführen. Dies vermindert die Beeinträchtigung anderer Systembenutzer durch solche größeren Datensätze".

Der Auftragsdurchsatz hat sich eindeutig erhöht. "Unsere Wissenschaftler können die Daten schneller verarbeiten", sagt DeReus.

Mit gleichbleibend hoher Leistung unabhängig von der jeweiligen Arbeitslast vermeidet das System Engpässe und hilft CMI-Forschern, ihre Arbeit pünktlich zu erledigen. Dies ist vor allem dann wichtig, wenn Forscher die Einreichungsfristen für wissenschaftliche Arbeiten einhalten, sich um die Präsentation ihrer Ergebnisse auf Industrie- und Hochschulkonferenzen bewerben oder Zuschussanträge einreichen müssen.

Flexibilität und Kontrolle

Oft benötigen die Forscher beim CMI je nach Projekt oder dem aktuellen Stand ihrer Forschungsarbeiten unterschiedlich viel Speicher. Mit der Panasas ActiveStor-Lösung kann sich das CMI ohne arbeits- oder qualifikationsintensive Abstimmungs- und Verwaltungsbemühungen an wechselnde Workload-Anforderungen anpassen.

"Immer mehr Leute bei uns wollen ihre Arbeit erledigen und dabei Datensätze verwenden, die um Größenordnungen umfangreicher sind als alles, was wir bisher gesehen haben", sagt DeReus. "Eine Technologie, die trägt Dateisystemen oder längeren Laufzeiten ihren Schrecken nimmt, ist dabei außerordentlich hilfreich".

"Der Großteil unserer Arbeit läuft in Teams ab", sagt Vázquez-Baeza. "Wenn wir die Daten nicht zwischen Nutzern oder mit anderen Partnern austauschen könnten, wäre das ein Hindernis. Eine zuverlässige Speicherressource wie Panasas ActiveStor erleichtert uns einen Großteil unserer kreativen Arbeit. Dank Panasas-Technologie kann das CMI seine Ziele leichter erreichen".

Vereinfachte Speicherung

Wer an der nächsten großen wissenschaftlichen Entdeckung arbeitet, denkt normalerweise nicht an die Speicherstruktur, die im Hintergrund zur Unterstützung der Arbeit agiert – es sei denn, sie behindert den Arbeitsfortschritt. Das

Fehlen von Beschwerden ist laut DeReus ein zuverlässiger Indikator dafür, dass die ActiveStor-Lösung die Anforderungen der Anwender erfüllt.

Frühere Speicherlösungen erwiesen sich für CMI-Wissenschaftler beim Abschluss ihrer Forschungen als Hindernis. Zu terminkritischen Zeiten erhielten manche Benutzer nicht einmal einen Login-Prompt für ihre Rechner. Dies verärgerte die Wissenschaftler, verlangsamte die Forschungsarbeit und führte zu vermehrten Beschwerden an die IT-Abteilung.

Die Panasas-Lösung beseitigte diese Probleme und vereinfachte den Speicherbetrieb beim CMI. "Heute bemerke ich keine Verlangsamung mehr. Da muss ich keine Jobs mehr verfolgen, die womöglich den Dateisystem-Durchsatz verlangsamen könnten", kommentiert DeReus. "Ich muss nicht mehr versuchen, eine Last aus dem System zu entfernen, um alle anderen wieder zum Laufen zu bringen".

Systemverwaltungsaufgaben zählen im Wesentlichen nicht mehr zu den täglichen Arbeiten. "Die grafische Oberfläche liefert Informationen über alle Performance-Aspekte, sodass wir für die Problemdiagnose nicht mehr in die Details der unteren Systemebenen einsteigen müssen", so DeReus.

Schnellere Einblicke in das Mikrobiom

Um Mikrobiome besser zu verstehen, plant das CMI auch weiterhin den Einsatz wissenschaftlicher Technologien der neuesten Generation. Das CMI konzentriert sich darauf, die Erkenntnisse der Forschung in Lösungen umzusetzen. Daher schätzt man technische Lösungen, die Forscher nicht von ihren wissenschaftlichen Problemen ablenken. Je weniger sich Wissenschaftler Gedanken über die Speicherung machen müssen, umso mehr kann das CMI den Einsatz modernster wissenschaftlicher Technologien vorantreiben. "Dabei darf Technologie niemals ein Hindernis für produktive wissenschaftliche Arbeit sein", erklärt Vázquez-Baeza. "Eine umfassende Kontrolle der Computerumgebung ist für uns von entscheidender Bedeutung. Dank Panasas können wir dies flexibel und ohne Beeinträchtigung unserer Benutzer erreichen".