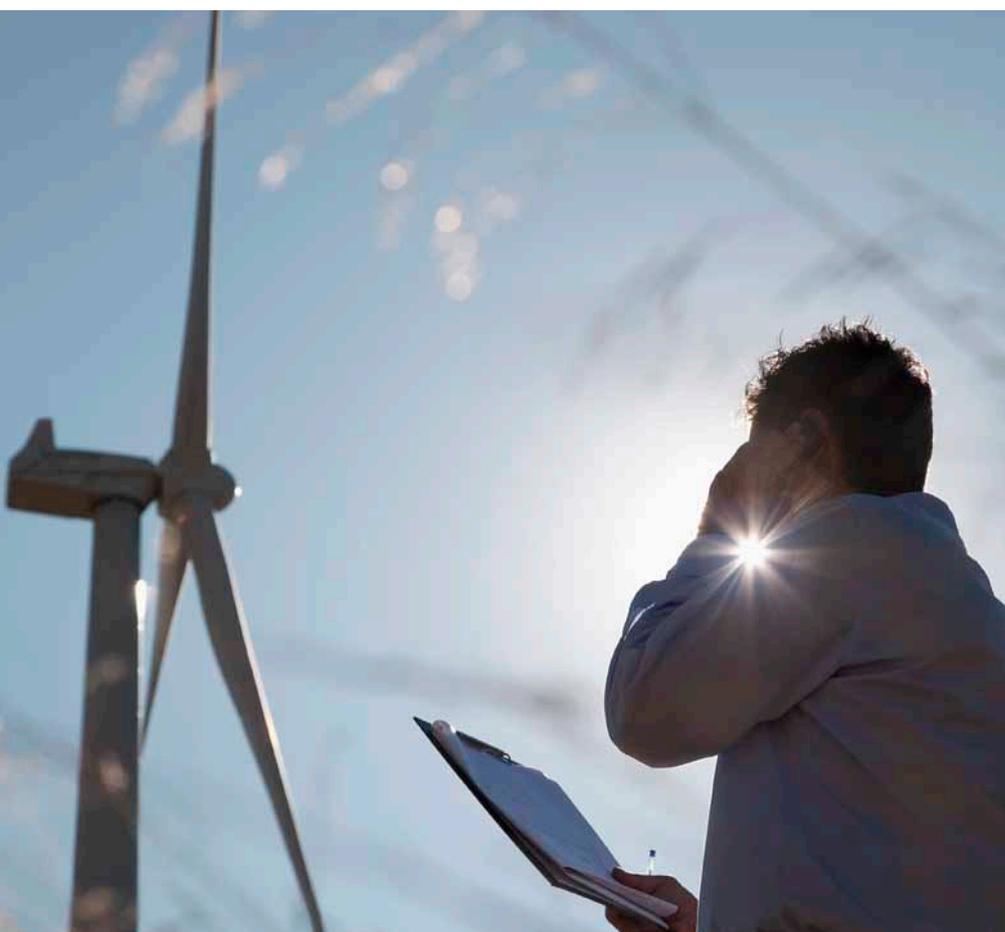




Die Macht der Forschung

Deutsches Forschungszentrum steigert Rechenleistung mit High-Performance-Computing-Cluster um das 10-fache



„Wir können die Kapazität unseres Clusters ohne Störungen erweitern.“

Thomas Schnicke, Leiter der Arbeitsgruppe Wissenschaftliches Rechnen und Wissenschaftliches Datenmanagement, UFZ

Kundenprofil



Unternehmen	Das Helmholtz-Zentrum für Umweltforschung(UFZ)
Branche	Forschung und Lehre
Land	Deutschland
Mitarbeiterzahl	1.000
Website	www.ufz.de

Geschäftsbedarf

Das UFZ benötigte einen starken High-Performance-Computing-Cluster (HPC), um den steigenden Bedarf der Wissenschaftler für komplexe Simulationen zu decken.

Lösung

Das UFZ implementierte einen HPC-Cluster auf Dell™ PowerEdge™-Servern mit Hilfe des "Dell Premier Partner" circular, GmbH.

Vorteile

- IT-Team spart Zeit mit schlüsselfertiger Lösung
- Wissenschaftler schließen Berechnungen mit 1.000 Prozent mehr Rechenleistung schneller ab
- Modular erweiterbares HPC-Cluster zu implementieren, um sowohl den aktuellen wie auch den zukünftigen Bedarf der Wissenschaftler zu decken
- Mit Dell ProSupport™ bewahrt die Organisation die Forschung vor Störungen

Lösungen

- HPC

Im Helmholtz-Zentrum für Umweltforschung (UFZ) erforschen Wissenschaftler die Ursachen und Folgen der weit reichenden Veränderungen der Umwelt.

„Die HPC-Berater von Dell haben zusammen mit den Experten von Circular eine großartige Plattform erarbeitet, auf der sie aufbauen konnten.“

Thomas Schnicke, Leiter der Arbeitsgruppe Wissenschaftliches Rechnen und Wissenschaftliches Datenmanagement, UFZ

Im Helmholtz-Zentrum für Umweltforschung (UFZ) erforschen Wissenschaftler die Ursachen und Folgen der weit reichenden Veränderungen der Umwelt. Sie befassen sich mit Wasserressourcen, biologischer Vielfalt, den Folgen des Klimawandels und Anpassungsmöglichkeiten, Umwelt- und Biotechnologien, Bioenergie, dem Verhalten von Chemikalien in der Umwelt, ihrer Wirkung auf die Gesundheit, Modellierung und sozialwissenschaftlichen Fragestellungen. Ihr Leitmotiv lautet: Unsere Forschung dient der nachhaltigen Nutzung natürlicher Ressourcen und hilft, diese Lebensgrundlagen unter dem Einfluss des globalen Wandels langfristig zu sichern. Das UFZ beschäftigt an den Standorten Leipzig, Halle und Magdeburg über 1.000 Mitarbeiter. Es wird vom Bund sowie von Sachsen und Sachsen-Anhalt finanziert.

Die Helmholtz-Gemeinschaft leistet Beiträge zur Lösung großer und drängender Fragen von Gesellschaft, Wissenschaft und Wirtschaft durch wissenschaftliche Spitzenleistungen in sechs Forschungsbereichen: Energie, Erde und Umwelt, Gesundheit, Schlüsseltechnologien, Struktur der Materie, Verkehr und Weltraum. Die Helmholtz-Gemeinschaft ist mit über 32.000 Mitarbeiterinnen und Mitarbeitern in 18 Forschungszentren und einem Jahresbudget von rund 3,4 Milliarden Euro die größte Wissenschaftsorganisation Deutschlands. Ihre Arbeit steht in der Tradition des Naturforschers Hermann von Helmholtz (1821-1894).

Ausgangssituation

Die Wissenschaftler der Organisation müssen ihr Wissen austauschen, da sich viele der wissenschaftlichen Fachbereiche überschneiden. Alle Forschungsgruppen haben zudem allgemein Bedarf an der Verarbeitung komplexer Berechnungen und großer Datenmengen. Um diesen Anforderungen gerecht zu werden, installierte die UFZ eine Hochleistungsrechnerumgebung (HPC). Das UFZ betreibt bereits ein Linuxcluster (LiClus) mit einer Skalierung bis maximal 256 Cores. Bei der Software-Implementierung dieses Clusters waren bereits Mitarbeiter der circular GmbH erfolgreich beteiligt. Dieses bestehende Cluster sollte erheblich erweitert werden um dem deutlich erhöhten Bedarf an Rechenleistung gerecht zu werden.

So funktioniert es

Hardware

Dell™ PowerEdge™ R710 Server mit Intel® Xeon® X5630-Prozessoren

Dell PowerEdge R410 Server mit Intel® Xeon® X5650-Prozessoren

QLogic InfiniBand Interconnect

Hauptanforderungen

Dem Anforderungsprofil entsprechend hat circular folgende Lösung unterbreitet:

- 4-fache Core-Skalierbarkeit gegenüber LiClus-Cluster (Skalierung bis zu 1032 Cores)
- QDR-InfiniBand Interprozess-Kommunikation (40 Gbit/s non-blocking)
- Nutzung vorhandener Racks (Luftkühlung, maximal 10 kW pro Rack)
- Paralleles Filesystem (GPFS)
- Grid Engine (Job-/Batchmanagement)
- Provisioning mittels xCAT

Wissenschaftliche Anwendungen

Auf diesem neuen Cluster sollen hauptsächlich Parallele Anwendungen aus dem Bereich Umweltingformatik (OpenGeosys) und der Hydrosystemmodellierung (Klimamodell: ECHAM) zum Einsatz kommen.

Thomas Schnicke, Leiter der Arbeitsgruppe Wissenschaftliches Rechnen und Wissenschaftliches Datenmanagement (WRWD) innerhalb der Abteilung WKDV (IT-Abteilung des UFZ), sagt: „Als wissenschaftliches Institut beschäftigen wir uns ständig mit neuen Projekten, um Umweltsysteme besser simulieren und Umweltprobleme lösen zu können. Insbesondere komplexere Simulationen mit höherer Genauigkeit erfordern mehr Rechenleistung. Daher wächst unser Bedarf, Daten zu verarbeiten, sehr schnell.“

Das UFZ beschloss daher, ein modular erweiterbares HPC-Cluster zu implementieren, um sowohl den aktuellen wie auch den zukünftigen Bedarf der Wissenschaftler zu decken. Thomas Schnicke wollte außerdem eine Umgebung installieren, die sich von seinem Team leichter verwalten ließe. Er meint: „Wir wollten eine schlüsselfertige Lösung, die jeden zukünftigen Bedarf skalieren würde.“

Im Rahmen einer EU-weiten Öffentlichen Ausschreibung wurde der Wettbewerb gestartet. Ausgeschrieben wurde die betriebsfertige Lieferung und Installation einer Clustererweiterung mit 1032 Cores. Besonderer Wert wurde dabei auf die schlüsselfertige ("Turn-Key Solution") Implementierung der komplexen Hardware- und Software-Gesamtlösung gelegt. Somit waren nicht nur Qualität und Preis der Hardware-Komponenten ausschlaggebend, sondern auch die Systemintegration.

Nach sechs Monaten, erhielt die circular Informationssysteme GmbH den Zuschlag für ihr auf DELL-Komponenten basierende Gesamtlösung. „Die HPC-Berater von Dell haben zusammen mit den Experten von Circular eine großartige Plattform erarbeitet, auf der sie aufbauen konnten. Der effiziente Ablauf war sehr essentiell, wodurch uns Circular die Umgebung liefern konnte, die wir haben wollten“, führt Thomas Schnicke aus.

Übersicht der Hardware-Komponenten

Das beauftragte Gesamtkonzept der circular GmbH bestand aus folgenden funktionalen Einzelkomponenten:

- 2* Frontend-Nodes für das Pre- und Postprocessing (Cluster-Login, Compiler)
- 86* Compute-Nodes für die eigentlichen Berechnungen
- 3* Storage-Nodes für performantes paralleles Filesystem über QDR-InfiniBand
- 1* FC-Storage mit ca. 40 TB Brutto-Kapazität
- 1* QDR-InfiniBand Infrastruktur mit einem zentralen InfiniBand-Switch
- 1* Administrations-Netzwerk (Gigabit-Ethernet) mit 10GbE-Uplink
- 1* Management-Netzwerk (Gigabit-Ethernet)

Konkrete Realisierung der Hardware

Die eigentliche Rechenlast des Clusters wird durch die Compute-Nodes getragen. Diese 86 Nodes wurden mit Dell PowerEdge R410 Server realisiert. Die Ausstattung mit jeweils 2 HexaCore-Prozessoren des Typs Intel X5650 (2.66 GHz, 6 Cores) ergibt somit 12 Cores je Node / 1032 Cores für alle Compute-Nodes zusammen. Für Berechnungen stehen jedem Compute-Node 64 GB Hauptspeicher zur Verfügung (4 GB pro Core + 4 GB Diskless Boot waren gefordert). Insgesamt verfügt das HPC-Cluster somit über 5.5 TB an Hauptspeicher.

Hauptgründe für die Wahl des Dell PowerEdge R410 Servers waren seine Energieeffizienz (Netzteile mit hohem Wirkungsgrad), die Remote-Managebarkeit und das ausgezeichnete Preis-/Leistungsverhältnis.

Thomas Schnicke entschied sich für die Dell PowerEdge R410 aufgrund des ausgewogenen Verhältnisses zwischen Leistung und Energie-Effizienz. Das gesamte Geschäftskonzept aus circular Expertise und Dell Hardware war für ihn ausschlaggebend. Zur Realisierung der Frontend- sowie Storage-Nodes wurden einheitlich 5 Stück Dell PowerEdge R710 Server verwendet. Dabei verfügen die Frontend-Nodes über die identische CPU der Compute-Nodes (Intel X5650), während die Storage-Nodes aus lizenztechnischen Gründen des parallelen Filesystems lediglich QuadCore-CPU's erhielten (Intel E5630). Alle Dell PowerEdge R710 Server sind dabei redundant über zwei Ports mit dem QDR-InfiniBand Netzwerk verbunden.

Hauptgründe für die Wahl des Dell PowerEdge R710 Servers waren auch hier seine Energieeffizienz (Netzteile mit hohem Wirkungsgrad), die Ausbaubarkeit (I/O-Karten), die hohe Verfügbarkeit (redundante HotSwap-Netzteile) und die Remote-Managebarkeit (Dell iDRAC6 Enterprise).

Außerdem gab es für das UFZ keinen entsprechenden Anstieg im Energieverbrauch, weil die Intel-Prozessoren sehr energieeffizient arbeiten trotz höherer Rechenleistung pro CPU-Core. Mit mehr Rechenleistung können die Wissenschaftler komplexere Simulationen durchführen und genauere Ergebnisse erhalten. Dies ist für die Zukunft des UFZ zwingend erforderlich, da viele der datenintensiven Projekte zur notwendigen Finanzierung des Unternehmens beitragen.

Als zentraler QDR-InfiniBand-Switch kam ein QLogic 12800-120 Director-Switch (non-blocking) zum Einsatz, der ebenfalls von Dell geliefert wurde. Dieser ist durch modulare Leaf-Einschübe bis zu 216 Ports ausbauen (momentan 108 Ports) und erlaubt somit eine einfache zukünftige Erweiterung des Clusters. Die beiden Gigabit-Ethernet Netzwerke basieren auf ein Core-/Leaf-Design mit Cisco Switches der Baureihe C2960 (4 Stück) und C3560X (2 Stück).

Ben Langenberg, Mitarbeiter der Arbeitsgruppe WRWD meint dazu: „Die modulare Realisierung aller drei Netzwerke erlaubt eine einfache Skalierung. Insbesondere die Ausbaureserve des zentralen QDR-InfiniBand-Switches ermöglicht eine Erweiterung um den Faktor 2 ohne zusätzliche Hierarchie-Ebenen.“

Konkrete Realisierung der Software

- RedHat Enterprise Linux Server Release 5.5 (RHEL 5.5) auf Dell PowerEdge R710 Server
- CentOS Release 5.5 auf Dell PowerEdge R410 Compute-Nodes
- xCAT (Extreme Cloud Administration Toolkit) Version 2.5 für das Provisioning und Management der Compute-Nodes (Diskless Clients mit Boot-over-IB)
- GPFS Parallel File System (Version 3.4) mit RDMA over QDR-InfiniBand
- Grid Engine Version 6.2 Update 5 für das Job-/Batchmanagement

Um die Verfügbarkeit des HPC-Clusters zu erhöhen wurden weiterhin zentrale Dienste HA-fähig implementiert. Insbesondere folgende Dienste sind hochverfügbar:

- Linux-HA Implementierung auf beide Frontend-Nodes
- Folgende Services sind HA: DHCP, TFTP, HTTP (Monitoring), xCAT und NFS
- Grid Engine (Master/Shadowmaster Implementierung unter Nutzung des parallelen Filesystems)

Projekterfahrung

Die enge Zusammenarbeit aller Projektbeteiligten (UFZ: Dr. Thomas Wieser, Leiter der IT, UFZ, Thomas Schnicke, Ben Langenberg, Christian Krause / circular: Ralph Hinsche, Michael Mertin, Florian Rätzer) war Garant für die höchst erfolgreiche Projektumsetzung. Der Kunde UFZ lobte das Projektmanagement, die reibungslose Installation, Implementierung und Integration und die unbürokratische und effiziente Beseitigung von nachgeschalteten technischen Herausforderungen.

FAZIT

Thomas Schnicke sagt: „Unsere neuen Systeme basieren auf Dell PowerEdge-Servern mit aktuellen Intel Xeon-Prozessoren, daher haben wir jetzt eine 10-mal höhere Rechenleistung als bei der vorherigen Lösung, aber nur dreimal so viele Kerne.“ Der Linpack-Benchmark ergab dabei eine Gesamt-Rechenleistung von 9,7 TFlops. Die Leistungsaufnahme bei Volllast liegt etwa 3-mal höher als bei dem kleineren Vorgängersystem; daher konnte die Energieeffizienz des neuen Systems um 10% verbessert werden.

Entscheidend ist, dass Thomas Schnickes Team die Rechenleistung und Speicherkapazität erhöhen kann, ohne die Simulationen der Wissenschaftler zu beeinträchtigen. „Wir können die Kapazität unseres Clusters ohne Störungen erweitern. In einer Organisation wie der unseren können wir den Benutzern nicht einfach sagen, dass sie für zwei Wochen, in denen wir an der Infrastruktur arbeiten, keine Berechnungen durchführen können.“



Hier können Sie alle Dell Partner-Fallstudien nachlesen dell.de/partner

