



SUCCESS STORIES

NETAPP GX STORAGE FÜR HIGH PERFORMANCE COMPUTING AM LEIBNIZ-RECHENZENTRUM



„Mit Data ONTAP GX ist NetApp eine Architektur gelungen, deren Performance, Konsistenz und Skalierbarkeit sich von vergleichbaren Systemen abhebt. Wir freuen uns, die weltweit erste GX-Installation in einer Supercomputing-Umgebung zu betreiben.“

Dr. Herbert Huber, Leiter der Gruppe Compute-Server, Leibniz-Rechenzentrum

AUF EINEN BLICK

Ort Garching bei München, Deutschland

Branche Forschung und Lehre

Kundenprofil

Das Leibniz-Rechenzentrum (LRZ) ist eines von drei nationalen Supercomputing-Zentren in Deutschland und darüber hinaus IT-Dienstleister für Forschung und Lehre an allen Hochschulen in München. www.lrz.de

Herausforderung

Projektdateien im Terabytebereich sollten für einen neuen Höchstleistungsrechner schnell, zuverlässig und konsistent bereitgestellt werden.

Lösung

Ein NetApp FAS3050 Cluster mit sechs Knoten und Data ONTAP GX Software liefert den Datenservice für die Projekt-Shares. Snapshots und Datenreplikation sorgen für Backup und Disaster Recovery.

Vorteile

- Zehnfache Metadaten-Performance
- Global Namespace vereinfacht Management
- Bewährte hohe Verfügbarkeit und Datenintegrität
- Flexible, für Benutzer transparente Datenverwaltung
- Snapshot- und Replikationstechnologie für hohe Datensicherheit
- Klare Skalierbarkeit für Ausbauphase 2 mit doppelter Leistung und Kapazität

DER KUNDE

Das Leibniz-Rechenzentrum

Die Ursprünge des Leibniz-Rechenzentrums (LRZ) gehen bis 1962 zurück, als Hans Piloty und Robert Sauer an der Bayerischen Akademie der Wissenschaften den Vorläufer der heutigen Kommission für Informatik gründeten. Unter diesem Dach wurde das LRZ mit Mitteln des Freistaats Bayern als gemeinsames Rechenzentrum für Forschung und Lehre für alle Münchner Hochschulen geschaffen. Es ist nach dem großen Philosophen und Universalgelehrten Gottfried Wilhelm Leibniz (1646-1716) benannt, der das binäre Zahlensystem entwickelte, die Grundlagen der Differentialrechnung schuf und eine der ersten mechanischen Rechenmaschinen für die vier Grundrechenarten entwarf.

Heute erbringt das LRZ allgemeine Dienste der Informationsverarbeitung für die Hochschulen in München und für die Bayerische Akademie der Wissenschaften. Es stellt mit dem Münchner Wissenschaftsnetz eine leistungsfähige Kommunikationsinfrastruktur bereit, ist Kompetenzzentrum für Datenkommunikationsnetze und betreibt umfangreiche Platten- und automatisierte Bandspeicher für Backup- und Archivierung. Der Datenbestand am LRZ verdoppelt sich jährlich und beträgt aktuell 1,7 Petabyte. Schließlich ist das LRZ ein Supercomputing-Zentrum für technisch-wissenschaftliches Hochleistungsrechnen und steht Benutzern aus ganz Deutschland zur Verfügung.

DIE HERAUSFORDERUNG

Hochleistungsspeicher für optimale Projektbedingungen

Seit Juli 2006 betreibt das LRZ eine Linux-basierte SGI Altix 4700 mit 4.096 Intel Itanium-2-CPU's, mehr als 17 Terabyte Hauptspeicher und einer Rechenleistung von 26 TFlop/s. Der Supercomputer mit Namen HLRB II – Höchstleistungsrechner in Bayern – ist einer von drei nationalen Supercomputern in Deutschland. Er rangiert derzeit auf Platz 18 der internationalen „Top500“-Liste der Supercomputer und steht bundesweit für Top-Projekte der deutschen Wissenschaftsgemeinde zur Verfügung. Die Plattform wird für Simulationen auf den Gebieten Strömungsdynamik, Festkörperphysik, Materialforschung, Astrophysik, Geowissenschaften, Chemie, Biowissenschaften und Medizin genutzt. HLRB I, der Vorgänger des aktuellen Systems, war eine Hitachi SR 8000 mit 1.368 Prozessoren und lokalem Plattenspeicher.

Das Plus an Performance auf der Compute-Seite sollte auch die Storage-Seite reflektieren. Für die Systemverantwortlichen am LRZ stand bereits im Vorfeld der Ausschreibung für den Supercomputer fest, dass die neue Umgebung zwei verschiedene Datenspeicher und Filesysteme erfordern würde, um optimale Projektbedingungen zu schaffen.

„Wir haben es am LRZ mit zwei Datentypen zu tun. Ein Projekt umfasst im Wesentlichen die eigentlichen Projektdaten der Wissenschaftler mit Eingangsdaten, Programmcodes und Endergeb-

„Die Synthese aus Global Namespace, hoher Performance und der NetApp eigenen Art von Verfügbarkeit, Skalierbarkeit und Management erwies sich für unsere Supercomputing-Umgebung als ideal. Storage mit hoher Transaktionsrate und Zuverlässigkeit ist hier ein absolutes Muss. GX hat unsere Benchmark-Bedingungen im ersten Anlauf erfüllt und bietet ein Performance-Niveau, das für den Benutzer spürbar ist. NetApp trägt dazu bei, dass wir den unterschiedlichsten Wissenschaftsprojekten optimale Arbeitsbedingungen bieten können.“

Christoph Biardzki, Storage-Betreuer, Leibniz-Rechenzentrum

nissen, die auch archiviert werden, sowie temporäre Zwischenergebnisse aus den laufenden Simulationen. Während der Umfang einer temporären Datei mehrere Terabyte erreichen kann, bewegen wir uns bei den Projektdateien im unteren Gigabyte-Bereich“, erklärt Dr. Bernd Reiner, Storage-Betreuer am LRZ. „Auch die Rahmenbedingungen sind völlig andere. Der Anspruch an Datenkonsistenz und Datenverfügbarkeit ist bei den temporären Dateien nicht so hoch wie bei den Projektdateien. Hier müssen Storage und Filesystem vom Anfang bis zum Ende eines Projektes große Mengen an vergleichsweise kleinen Dateien schnell und absolut zuverlässig handhaben.“

Das LRZ würde parallel mehr als hundert wechselnde Projekte simultan zu verwalten haben, so dass es auch auf eine möglichst einfache Art des Storage-Managements und zuverlässige Methoden für Backup und Disaster Recovery ankam.

DIE LÖSUNG

NetApp FAS3050 mit Data ONTAP GX für Fileservice und Datenablage

Mit Data ONTAP GX hat NetApp Mitte 2006 ein neues Betriebs- und Filesystem für High Performance Computing herausgebracht. GX steht ganz in der Entwicklungstradition von NetApp und kombiniert die für Supercomputing-Umgebungen nötige Performance mit einfachem Management und hoher Datenintegrität.

NetApp Lösungen sind am LRZ seit einigen Jahren für Fileservices, Datensicherung und Archivierung im Einsatz, so dass die Management-Vorteile und hohe Zuverlässigkeit der NetApp Architektur bekannt waren.

Inwieweit sich Data ONTAP GX für die Bedingungen am LRZ eignen würde, sollten Tests zeigen. Das LRZ hatte auf reale Projektbedingungen zugeschnittene Benchmarks entwickelt. Die Metadaten-Performance war ebenso Testobjekt wie der sequenzielle Datendurchsatz. Dieser Wert gibt Aufschluss über die Systemleistung beim Einspielen einer Ergebnisdatei in ein Projekt-Share. Die bereits mit Entwicklungsversionen der Software erzielten Ergebnisse waren laut Christoph Biardzki, einem der Storage-Experten am LRZ, äußerst vielversprechend und überzeugten auch SGI, Generalunternehmer für das HLRB-II-Projekt.

Der neue Supercomputer besteht aus 16 Partitionen mit jeweils 256 Prozessoren und verfügt über eine 4 GBit Fibre-Channel-Umgebung mit aktuell 300 Terabyte an schnellem Plattenspeicher für die Zwischenergebnisse. Hier kann CXFS, das verteilte Cluster-Filesystem von SGI, seine Möglichkeiten der Parallelverarbeitung im großen Stil ausspielen. Der NetApp Storage mit 40 Terabyte Nettokapazität auf Fibre Channel-Festplatten wird von allen Partitionen über NFS angebunden und dient als Speicher für sämtliche projektrelevanten Daten und Software. Anhand des

Global Namespace von Data ONTAP GX können mehrere, voneinander unabhängige Datenspeichersysteme als große Festplatte mit allen Dateien und Dateisystemen präsentiert und verwaltet werden.

„Die Implementierung verlief schneller und einfacher als wir erwartet hatten. Neben den lokalen NetApp Technikern hatten wir für das Fine-Tuning auch einen Mitarbeiter des Special Deployment Teams aus Israel im Haus“, erinnert sich Biardzki. „Es war natürlich ein Risiko anhand der Betaversion der Software eine derartige Investitionsentscheidung zu treffen. Neben den Benchmarks sprachen aber auch die guten Erfahrungen mit dem NetApp-Support in der Vergangenheit dafür. Die Ergebnisse haben uns Recht geben und wir haben die erste GX Installation weltweit in Betrieb genommen.“

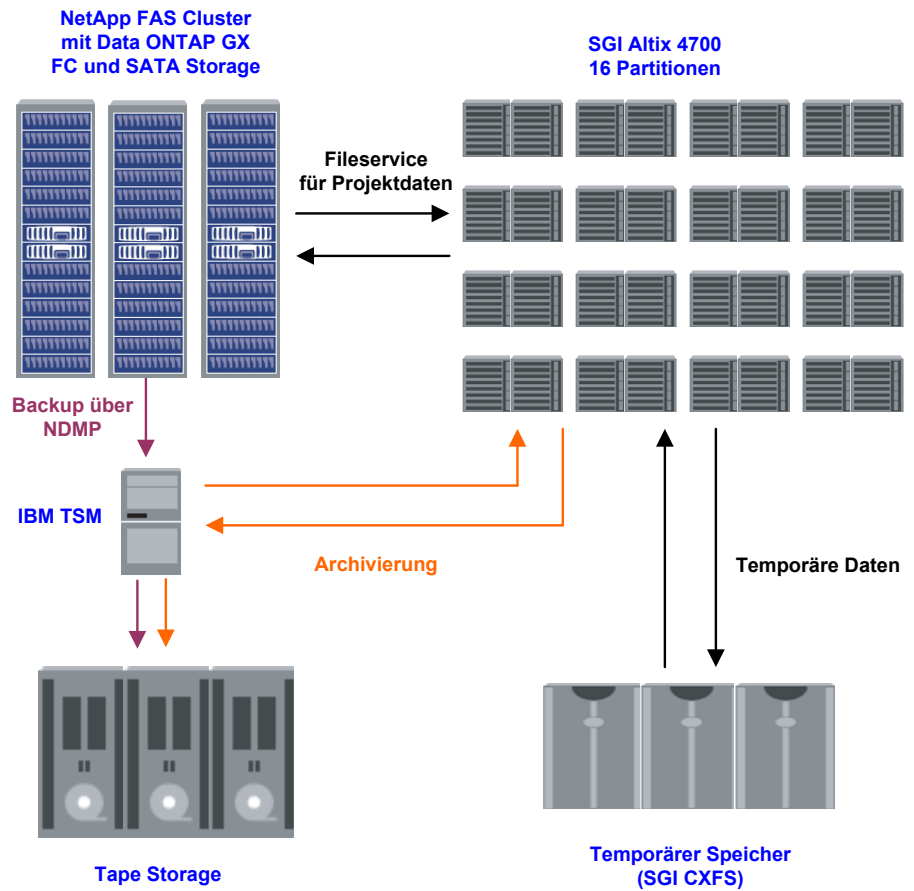
Aufgrund ihrer Wichtigkeit werden die Daten auf dem FAS Storage online mit der Snapshot-Technologie von Data ONTAP GX gesichert. Zusätzlich wird die in GX integrierte Möglichkeit zur Replikation genutzt, um die Daten auf zusätzliche SATA-Festplatten zu spiegeln. Schließlich wird über NDMP ein Backup auf STK T10000 Bandlaufwerke geschrieben. Für die Langzeitarchivierung können Benutzer ein von IBM TSM gesteuertes Archivsystem einsetzen, das auf eine Bänderumgebung mit aktuell 2.300 Terabyte Kapazität zugreift.

DIE VORTEILE

Global Namespace vereinfacht Management und sorgt für reibungslosen Projektbetrieb auf jeder Ausbaustufe

Jedes Forschungsprojekt mit Rechenerlaubnis auf dem HLRB II erhält auf dem NetApp Storage dediziert Speicherkapazität zugewiesen. Die Zulassung eines Projekts für den Rechner und die Zuteilung von Rechenzeitkontingenten geschieht durch ein nationales Gremium aus hochrangigen Wissenschaftlern. Sie setzt Erfahrung im Umgang mit Höchstleistungsrechnern und Parallelverarbeitung bei den Bewerbern voraus und richtet sich auch danach, ob das Projekt zur Rechenarchitektur des HLRB II passt. Ist ein Projekt genehmigt, richtet das LRZ Projekt-Shares ein und validiert das Projekt auf dem HLRB II.

Der gemeinsame Namensraum (Global Namespace), den Data ONTAP GX bietet, gehört zu den großen Vorteilen von GX. „Würde man beispielsweise einfache Fileserver als Dateisystem einsetzen, kommt es früher oder später zu Leistungs- oder Kapazitätsgrenzen, die dazu zwingen, weitere Dateisysteme aufzumachen. Da die Systeme voneinander unabhängig sind, muss man sie separat verwalten, und Benutzer verbringen viel Zeit damit einzelne Dateien zu lokalisieren. Anders beim Global Namespace von Data ONTAP GX, das in dieser Hinsicht dem am LRZ bewährten Andrew File System ähnelt. Hier werden sämtliche Dateisysteme und Dateien unter einem Dach zusammengefasst, obwohl sie auf unterschiedlichen Speichersystemen liegen“, führt Biardzki aus. „Jedes Projekt hat seinen eigenen Speicherbereich, der in den Global Namespace integriert wird und sich nach Belieben zwischen den sechs Maschinen des Clusters verschieben lässt. Diese administrative Funktionalität war uns sehr wichtig, damit Kapazitätsanpassungen und Rekonfigurationen im laufenden Betrieb erfolgen können. Der Prozess aus Dateneingabe, Simulationsrechnung und Ergebnisausgabe bleibt so vom Storage-Management unbeeinträchtigt.“



Die Storage-Struktur für den HLRB II am LRZ. Die Kette aus File Service, Datenverarbeitung und Ergebnisablage ergänzt ein Datensicherungskonzept aus online Snapshots sowie Backup und Archivierung auf schnelle Bandroboter von StorageTek, zum Teil unterstützt durch Platten-Caching. IBM Tivoli Storage Manager (TSM) steuert das Archiv, auf das die Benutzer direkten Zugriff haben.

„NetApp bietet uns Storage der Hochgeschwindigkeits- und Hochverfügbarkeitsklasse mit dem Komfort des Network-Attached Storage. Wir würden diese Entscheidung jederzeit wieder treffen. NetApp Storage unter Data ONTAP GX ist ein klarer Fall für Applikationen, die mit zahlreichen Dateien operieren, eine hohe Verarbeitungsleistung brauchen und absolute Datenintegrität erfordern.“

Dr. Bernd Reiner, Storage-Betreuer, Leibniz-Rechenzentrum

Bei mittlerweile mehr als 120 Projekten gleichzeitig sind Global Namespace und flexible Volumeverwaltung eine große Entlastung für das Storage-Team. Die Projekte sind sehr unterschiedlich, was die Größe der Daten-Sets, die Anzahl der involvierten Mitarbeiter, deren Bedarf an gemeinsamem Datenzugriff und die Verweildauer des Projekts auf dem NetApp System betrifft. Für die Benutzer ist der Global Namespace völlig transparent. Das heißt, sie „sehen“ nur ihren jeweiligen Projektbereich – ohne direkten Bezug zum physischen Speicher.

Der Skalierungspfad ist bereits klar vorgezeichnet: Die aktuelle Konfiguration des Rechners repräsentiert die erste Ausbauphase. In der zweiten Phase – ab dem zweiten Quartal 2007 – wird das LRZ Rechenleistung und Speicherkapazität des Rechners verdoppeln. Für den NetApp Speicher heißt das zusätzliche Plattenkapazität bei mehr sequenziellem Durchsatz und gleich bleibend hoher Metadaten-Performance. Aufgrund des Global Namespace wird die logische Integration mit ein paar Mausklicks erledigt sein.

Um Faktor 10 höhere Performance

Die Performance des Systems hat sich im täglichen Einsatz bewährt und liefert ein Niveau, das für die Benutzer spürbar ist. Neben der eher physikalisch bedingten

Durchsatzperformance spielt auch die Metadaten-Leistung eine große Rolle. Metadaten liefern Informationen wie etwa Speicherort, Größe, Urheber etc. einer Datei. Sind sie getrennt von den Daten abgelegt, wirkt sich dies negativ auf die Performance aus, sobald viele Dateien, beispielsweise zum Kompilieren, vom Speicher abgerufen werden. Die Handhabung der Metadaten war beim Vorgängerrechner – dem HLRB I – ein großes Problem, das oft von Benutzern bemängelt wurde, die interaktiv mit dem System gearbeitet haben.

Bei NetApp liegen Daten und Metadaten im selben Volume. Kurze Wege begünstigen die Performance beim Umgang mit hunderten kleiner Dateien. Im Vergleich zum temporären CXFS-Filesystem des neuen Supercomputers, das auf massive Parallelverarbeitung aber wenige Dateien in Terabytegröße ausgelegt ist, wird daher eine um Faktor 10 höhere Metadaten-Performance erreicht. „Diese Architektur macht NetApp Storage zur Idealbesetzung für unsere Projekt-Shares“, sagt Biardzki.

Konsistenz und Datensicherheit

Die dedizierte Einrichtung von zwei verschiedenen Speicherplattformen folgte auch der Notwendigkeit der Konsistenz und Datensicherheit. Beides wichtige Kriterien, um die Ergebnisqualität sicher-

zustellen. Für die Aussagekraft und Richtigkeit der Simulationsberechnungen ist eine konsistente Datenhaltung auf dem Fileserver unerlässlich. „Bei NetApp können wir sicher sein, dass Daten korrekt abgelegt und genauso wieder abrufbar sind. Die Anforderungen an den temporären Speicher sind dagegen weniger streng – die Daten könnten schließlich neu berechnet werden. Wegen der extremen Parallelverarbeitung und der schiereren Anzahl der Komponenten können sich dort Fehler schneller ausbreiten“, erklärt Biardzki. „Für die Sicherheit der Daten in den Projekt-Shares nutzen wir die integrierten Snapshots und Replikation. Diese äußerst komfortable Art für online Backup, Spiegelung und Disaster Recovery konnten uns andere Filesysteme nicht bieten, und wenn doch, dann nur unter hohen Performance-Einbußen. NetApp und Data ONTAP GX bieten das optimale Feature-Set für die Storage-Anforderungen der Projekt-Shares.“

PRODUKTE

Hardware: 6 FAS3050 mit 50TB Nettospeicher

Software: Data ONTAP GX, Snapshots, SnapRestore, SnapMirror auf SATA-Platten, NDMP Backup

Protokolle: NFS



www.netapp.de

Network Appliance GmbH
Bretonischer Ring 6
D-85630 Grasbrunn
Tel.: +49 (0)89 900594-0
Fax: +49 (0)89 900594-99
E-Mail: info_vertriebCE@netapp.com

© 2007 Network Appliance, Inc. Alle Rechte vorbehalten. Angaben können ohne Vorankündigung geändert werden. NetApp und das Network Appliance Logo sind eingetragene Warenzeichen und Network Appliance und Snapshot sind ein Warenzeichen von Network Appliance, Inc. in den USA und anderen Ländern. Alle anderen Marken oder Produkte sind Warenzeichen oder eingetragene Warenzeichen der entsprechenden Rechtsinhaber. 01/2007