

SERVICE & SUPERCOMPUTING

Durch guten Support zur Höchstleistung

BENUTZER, BETREUER UND ADMINISTRATOREN ARBEITEN ENG ZUSAMMEN, UM DAS OPTIMUM AUS DEM HÖCHSTLEISTUNGSRECHNER AM LEIBNIZ-RECHENZENTRUM HERAUSZUHOLEN



GUDRUN KAISER

Löwenanteil der Berechnungsarbeit ausmachen. Die Leistung von 2 TeraFlop/s (= 2 Millionen MFlop/s) der am LRZ installierten SR8000 lässt sich folgendermaßen beschreiben: Wenn ein Mensch in zehn Sekunden zwei zwölfstellige Zahlen multipliziert, dann müssten neuntausend mal so viele Menschen, wie heute auf der Erde leben, Multiplikationen durchführen, um die Leistung der Maschine zu erreichen.

Große Datenmengen aus dem Hauptspeicher

Damit große Rechnungen schnell (mit hohen MFlop/s-Raten) ausgeführt werden können, ist es wichtig, dass der Prozessor schnell genug mit den Daten, mit denen er rechnet, versorgt werden kann. Und da sieht es bei den PCs von der Stange gar nicht mehr so gut aus, wenn es sich um große Datenmengen handelt. Schon ein einziger SR8000-Prozessor überflügelt dabei einen modernen PC mit mindestens der sechsfachen MFlop/s-Leistung. Der Grund für diese „Umkehr“ der Verhältnisse im realen Betrieb ist die sogenannte Speicherbandbreite, also die Zahl der Bytes, die pro Sekunde aus dem Hauptspeicher in den Prozessor geladen werden können. Um nicht nur auf ein paar wenigen Daten aus dem Cache, das ist der kleine schnelle Speicher auf dem Prozessorchip selbst, sondern auf allen Daten im Hauptspeicher flott rechnen zu können, müssen Supercomputer-Hersteller spezielle Hardware einsetzen. Denn ein Eng-

VON RALF EBNER UND
HELMUT HELLER

Vordreieinhalb Jahren wurde am Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften der Supercomputer Hitachi SR8000-F1 als „Höchstleistungsrechner in Bayern“ (kurz HLRB) in Betrieb genommen. Damit etablierte sich das Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (LRZ) neben dem Höchstleistungsrechenzentrum Stuttgart (HLRS) und dem John-von-Neumann-Institut (NIC) des Forschungszentrums Jülich als drittes deutsches Höchstleistungsrechenzentrum.

Die Hitachi SR8000-F1 am Leibniz-Rechenzentrum

Höchstleistungsrechner im Vergleich zum PC

Dreieinhalb Jahre sind im IT-Bereich schon eine lange Zeit. Mittlerweile bekommt man für den Privatgebrauch Standard-PCs mit Prozessoren, die mit beinahe der zehnfachen Taktfrequenz der SR8000-Prozessoren (375 MHz) arbeiten. Allerdings ist für die Rechenleistung nicht nur die Taktfrequenz ausschlaggebend, sondern die tatsächlich erreichten MFlop/s, d.h. Millionen Floating-Point-Operationen pro Sekunde, die in naturwissenschaftlichen und technischen Simulationen den

pass bei der Speicherbandbreite lässt die Rechenleistung des Prozessors bei technisch-wissenschaftlichen Simulationen ohne weiteres um den Faktor 100 schrumpfen.

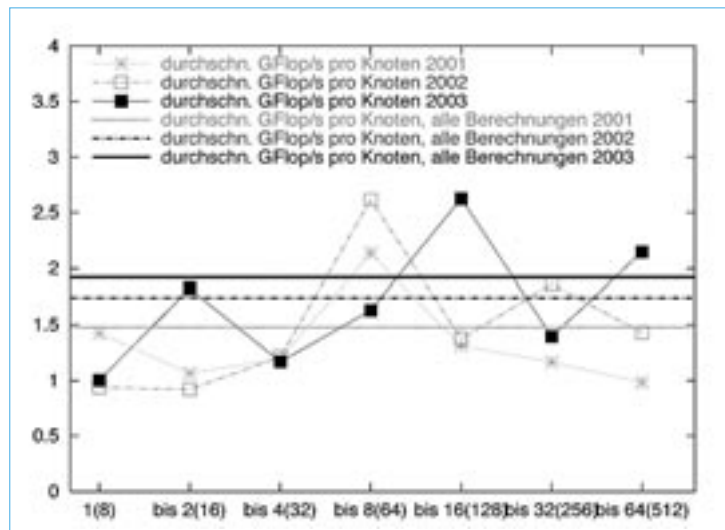
Leistungsfähiges Verbindungsnetzwerk

Ein einzelner Prozessor, der schnell auf große Datenmengen im Hauptspeicher zugreifen kann, macht noch keinen Supercomputer aus. Die zu simulierenden Aufgabenstellungen sind meist so groß, dass zu ihrer Lösung viele solcher Prozessoren zusammenarbeiten – der HLRB hat 1512 Prozessoren –, wobei die Daten auf die den Prozessoren zugeordneten Hauptspeichermodule aufgeteilt werden müssen. Während der Berechnung stehen die Prozessoren dann in regem (Daten-)Austausch miteinander. Um also auch auf vielen Prozessoren eine flüssige Berechnung zu erreichen, ist ein sehr leistungsfähiges Verbindungsnetzwerk notwendig, ein sog. Crossbar-Switch, der bis zu 950 MByte/s zwischen zwei Prozessoren und insgesamt 580.000 MByte/s mit extrem kurzen Verzögerungen (Latenzzeiten) von 17 Mikrosekunden übertragen kann.

Qualität von Hardware, Software und Administration

Damit eine so große Maschine überhaupt benutzbar ist, ist eine hohe Qualität der Einzelteile (sowohl Hard- als auch Software) nötig, wie folgende kleine Überschlagsrechnung zeigt: Geht man etwa davon aus, dass bei einem einzelnen „Standard“-Computer mit einer Wahrscheinlichkeit von 1% innerhalb einer Woche eine Störung (Absturz, Hardware-Schaden etc.) auftritt bzw. eben mit 99% Wahrscheinlichkeit keine Störung, so müsste bei der SR8000 mit ihren 1512 Prozessoren mindestens mit folgender Wahrscheinlichkeit eine Störung in den nächsten sieben Tagen auftreten:

$$1 - 0,99^{1512} > 99,99 \% !$$



Durchschnittliche GFlop/s-Rate (=1000 MFlop/s) aller HLRB-Rechenaufträge, gruppiert nach der Anzahl der verwendeten Prozessoren. Hier erkennt man eine deutliche durchschnittliche Leistungssteigerung über die drei Betriebsjahre hinweg, speziell bei den großen Rechenaufträgen mit bis zu 64 Knoten, die mittlerweile ein Drittel der HLRB-Rechenzeit verbrauchen.

Wenn die SR8000 also die gleiche „Qualität“ wie ein Standard-PC hätte, wäre ein Ausfall innerhalb der nächsten sieben Tage fast sicher! Und hier sind noch nicht einmal die zusätzlichen komplexen Spezialkomponenten für den Crossbar-Switch berücksichtigt, bei denen ebenfalls Störungen auftreten können.

Tatsächlich sind aber ununterbrochene Betriebsdauern von mindestens zwei Wochen die Regel; teilweise sind schon vier Wochen erreicht worden. Insgesamt liegt die Verfügbarkeit der Maschine bei über 99% der Zeit. Dies ist in dieser Größenordnung nicht mehr nur durch Hardware-Qualität zu erreichen, sondern es erfordert auch Administratoren, die sich mit der Vielzahl und den Besonderheiten der Hard- und Software-Komponenten auskennen. Sie überwachen den Betrieb und versuchen zu verhindern, dass kleinere Pannen zu größeren Systemstörungen oder zum Absturz der ganzen Maschine

führen. Außerdem werden vorsorgliche Wartungen durchgeführt, bei denen verdächtige Komponenten schon vor ihrem Lebensende ausgetauscht werden. Schließlich stehen die fleißigen japanischen Hard- und Software-Ingenieure vor Ort und in Fernost rund um die Uhr bereit, um auftretende Probleme schnell zu beheben.

Wofür Forscher den Hochleistungsrechner verwenden

Wissenschaftler der deutschen Hochschulen und der Max-Planck-Institute können den HLRB im Rahmen wissenschaftlicher Projekte nutzen mit dem Ziel, kostspielige oder manchmal gar nicht realisierbare technische oder naturwissenschaftliche Experimente durch Computersimulationen zu ersetzen oder zumindest zu unterstützen. Kennzeichnend für all diese sogenannten Grand-Challenge-Projekte ist der enorme Rechenzeit- und Speicherbedarf. Fast jedes dieser Projekte allein könnte wohl den HLRB auslasten. Ein wichtiges Kriterium für die Zulassung der Projekte ist, dass ihre Rechnungen auf keinem kleineren Computer durchführbar sind. Ein Überblick über die Aufgaben von HLRB-Projekten wurde bereits im Artikel über den Münchner Superrechner in Akademie Aktuell 1/2002 gegeben. Auf der Internet-Seite <http://www.lrz-muenchen.de/projekte/hlrb-projects> finden sich Informationen zu den hier nicht

weiter betrachteten Projekten aus der Hochenergie-, Festkörper-, Astro- und Biophysik, den Geowissenschaften, der Chemie, der Angewandten Mathematik und der Informatik.

Neuere Projektvorhaben sollen im Folgenden kurz erwähnt werden. Nach wie vor stammt gut ein Viertel der Höchstleistungsrechner-Projekte aus den Ingenieurwissenschaften, speziell der Strömungsmechanik (engl. Computational Fluid Dynamics, CFD). Die Einsatzfelder sind naturgemäß sehr vielfältig. Klassische Aerodynamik-Berechnungen von Fahrzeug- und Flugzeugteilen sind hier ebenso vertreten wie die Simulation von Gas- und Flüssigkeits-Strömungen in Rohren und Kanälen, Luftströmungen in Büroräumen oder von Grundwasserströmungen im Erdboden. Für die Industrie bedeutsam sind Strömungssimulationen in komplizierten dreidimensionalen Anordnungen wie etwa Turbomaschinen, Rührwerken oder die Schmelzen, in denen die für die Computerchip-Herstellung benötigten Siliziumkristalle gezüchtet werden. Neuerdings befasst sich ein Projekt auch mit der Vorausberechnung des Lärms, der von Strömungen erzeugt wird.

In der Nano-Technologie werden hochaktuelle Fragestellungen zur Erzeugung kleinster Strukturen aus einzelnen Molekülen angegangen. Solche Strukturen sollen als Mini-Sensoren, -Halbleiter, -Computer, -Maschinen oder -Kraftwerke dienen. Hier liefern Computer-Simulationen wertvolle Informationen, welche Molekülstrukturen dazu geeignet sind und mit welchen Verfahren man sie produzieren könnte. In der Biochemie werden die elektronischen Eigenschaften der Desoxyribonukleinsäure (DNA) und ihre Eignung als molekulares Kabel für Bio-Chips untersucht. Zwei Projekte wollen zum genaue-

ren Verständnis beim biologischen Vorgang des Sehens beitragen: Zum einen sollen die in der Netzhaut ablaufenden Reaktionen beim Sehen simuliert werden, zum anderen werden die entstehenden Verknüpfungen in der Großhirnrinde untersucht. Ein aufstrebender Bereich in den so genannten Biowissenschaften sind Anwendungen für die Gen-Analyse. Hierzu müssen mit schwierigen mathematischen und informationstechnischen Verfahren Übereinstimmungen in riesigen Mengen von Gen-Sequenzen gefunden werden.

Neue Projektanträge können jederzeit gestellt oder verlängert werden. Dies ist durchaus nicht bei allen Rechenzentren üblich. Durch den zu jedem Zeitpunkt möglichen Starttermin kann jedes Forschungsteam möglichst rasch die notwendigen Berechnungen durchführen, unnötige Verzögerungen der wissenschaftlichen Arbeit werden vermieden. Doch mit der Genehmigung eines HLRB-Projekts beginnt erst die eigentliche Arbeit für die Forscher, denn es gilt, sich mit den speziellen Gegebenheiten (Architektur, Benutzer- und Programmierschnittstellen) der Maschine vertraut zu machen, und innerhalb der zugestandenen Rechenzeit die nötigen Simulationen sinnvoll zu planen.

Das LRZ als moderner Dienstleister

Mit einem großen Fundus an Schriften im Internet, die die Herstellerdokumentation zusammenfassen, Erfahrungen dokumentieren und die spezielle Umgebung des HLRB beschreiben, versucht das LRZ, den Benutzern den Start auf dem HLRB zu erleichtern.

(<http://www.lrz-muenchen.de/services/compute/hlrb>).

Speziell auf die SR8000 zugeschnittene Kurse zu Rechnerarchitektur, Programmierung, Optimierung und Software-Werkzeugen sind ein wei-

terer wichtiger Baustein im LRZ-Service. Diese Kursangebote werden auch von vielen Benutzern von außerhalb Münchens trotz des Reiseaufwands gern angenommen, ersparen sie sich doch die selbständige mühevoll durchgeführte Dokumentation, denn trotz aller Bemühungen fallen die Anleitungen sowohl des LRZ als auch die der Hersteller unweigerlich sehr technisch aus.

Statusseminare und Workshops

Mithilfe eines Höchstleistungsrechenzentrums sollen auch die Benutzer untereinander wissenschaftliche und technische Erfahrungen austauschen. Bei den beiden großen Statusseminaren „High Performance Computing in Science and Engineering“, die 2002 in Zusammenarbeit mit dem Kompetenznetzwerk für wissenschaftliches Hoch- und Höchstleistungsrechnen in Bayern (KONWIHR) an der TU München in Garching und 2003 zusammen mit dem Höchstleistungsrechenzentrum Stuttgart durchgeführt wurden, konnten die Teilnehmer wertvolle Einblicke in die Arbeit und Techniken ähnlich gelagerter Projekte gewinnen.

Ein besonderer Workshop, ebenfalls im Rahmen von KONWIHR, ist der jährlich stattfindende „Period Performance Workshop“. Hier werden gezielt Anwenderprogramme ausgewählt, die offensichtlich noch nicht so performant auf der SR8000 laufen, und dann beim Workshop zusammen mit den Benutzern intensiv analysiert und verbessert. Erfreulich ist, dass die so behandelten Programme dann bald ganz oben auf der Rangliste der Rechenaufträge mit den höchsten MFlop/s-Raten (also mit der besten Leistung) erscheinen.

Die Zeitschrift inSiDE (innovatives Supercomputing in Deutschland, <http://www.hlr.de/news-events/inside>) wurde vor kurzem als ge-

meinsames Organ der drei deutschen Höchstleistungsrechenzentren ins Leben gerufen. HLRB-Projektmitarbeiter finden hier im Überblick interessante Veranstaltungshinweise und Artikel aus dem Bereich des Höchstleistungsrechnens.

Benutzerbetreuung für HLRB-Projekte

Der Wissenschaftsrat hat zusätzlich zu den Systemadministratoren noch weitere Stellen für Wissenschaftler empfohlen, die sich nicht mit dem Maschinenbetrieb direkt, sondern mit der Benutzer- und Anwendungsseite befassen. Der Freistaat Bayern ist dieser Empfehlung gefolgt und hat mit den Geldmitteln, die für den Rechner und seinen Betrieb je zur Hälfte vom Land und vom Bund zur Verfügung gestellt wurden, auch zusätzliche Stellen für HLRB-Benutzerbetreuer genehmigt. Um von Synergien zu profitieren, sind diese Betreuer für alle Computer-Server am LRZ zuständig. Der Betreuungsaufwand für HLRB-Projekte ist jedoch besonders hoch, da die Projekte allesamt an der vordersten Front der Forschung arbeiten, und da es nur wenige vorgefertigte Programme gibt. Am ehesten können noch Chemiker von dem halben Dutzend „Standard“-Programmen profitieren, die wie gewohnt auch auf dem HLRB nutzbar sind. Aber das Gros der Projekte ist auf Mitarbeiter angewiesen, die sich sowohl in ihrem Fachgebiet als auch in der Programmierung so gut auskennen, dass sie Programme „nach Maß“ für ihre Aufgaben schreiben oder zumindest anpassen können.

Leistungssteigerung von Benutzerprogrammen

Der HLRB ist zwar leistungsfähig, aber nur, wenn die Programme sich an diese leistungsfähige Architektur anpassen. Nicht selten ist man nach langen Anpassungs- und Übersetzungszeiten zunächst ein-

mal enttäuscht, dass das Programm auf dem HLRB langsamer als auf dem eigenen Arbeitsplatzrechner läuft. Dann muss in sogenannten Profilen nach den Programmteilen gesucht werden, die die Engpässe beim Fortgang der Berechnungen darstellen. Diese Programmteile sind so umzuschreiben, dass sie durch gleichmäßig fortlaufende Speicherzugriffe und ausgewogenes Kommunikationsverhalten die Architektur des Supercomputers optimal ausnutzen. Trotz der bereitgestellten Dokumentationen und Anleitungen sind die Wissenschaftler der Institute dabei immer wieder auf den Rat und die tatkräftige Unterstützung des Rechenzentrums angewiesen.

Zur Rentabilität der Benutzerbetreuer

Wenn es einem Benutzerbetreuer gelingt, bei Programmen, die pro Jahr ca. 3% der Rechenzeit auf der Maschine benötigen, eine Beschleunigung um den Faktor 4 zu erreichen (realistisch sind oft noch höhere Beschleunigungen), dann können in diesen 3% der Betriebszeit viermal so viele Rechnungen durchgeführt werden; die Nachfrage nach Rechenzeit ist – wie oben geschildert – ja vorhanden. Wollte man stattdessen die größere Anzahl an Rechnungen durch zusätzliche Hardware ermöglichen, müsste man bei den auf ein Jahr umgelegten Anschaffungs- und Betriebskosten des HLRB mit ca. 9 Mio. EUR folgende Summe ausgeben:

$$(4 - 1) \times 3\% \times 9 \text{ Mio. EUR / Jahr} = 810.000 \text{ EUR / Jahr.}$$

Um die gleichen wissenschaftlichen Erkenntnisse durch die vierfache Zahl an Simulationen zu erzielen, können durch Einstellung eines Benutzerbetreuers (großzügig veranschlagte Kosten von ca. 70.000 Euro pro Jahr) also 740.000 EUR/Jahr, d.h. über 90% gespart werden!

Identifizierung problematischer Rechenaufträge

Bisweilen ist den Benutzern nicht bewusst, dass ihre Programme nicht optimal laufen. Deshalb wird kontinuierlich eine große Menge von Leistungsdaten direkt aus der Hardware ausgelesen und versucht, die relevanten, d.h. die auf mögliche Probleme hindeutenden Teile herauszufiltern und den betreffenden Benutzern zuzuordnen. Es ist unabdingbar, dass eine solche ständige Überwachung sowohl der Maschine selbst als auch der rechenenden Prozesse erfolgt, damit nicht unnötig Rechenzeit vergeudet wird. In Umfragen unter den Benutzern des HLRB wird der Support durch das LRZ immer wieder gelobt. Eine gute Beratung ist natürlich auch im Interesse des Leibniz-Rechenzentrums der Bayerischen Akademie der Wissenschaften. Nur mit optimierten Programmen ist ein gutes Verhältnis der sustained performance (also der Leistung, die über einen längeren Zeitraum durch Anwendungsprogramme erreicht wird) zur peak performance (also der Leistung, die bei optimal angepasstem Programm theoretisch möglich ist) erreichbar. Und nur so ist eine Nutzung garantiert, die die hohen Kosten der Maschine rechtfertigt.

Das Betriebskonzept des HLRB hat auch den „Nationalen Koordinierungsausschuss von Höchstleistungsrechnern“ im Wissenschaftsrat überzeugt und dazu bewogen, die Aufstellung eines Nachfolgerechners für den HLRB am Leibniz-Rechenzentrum zu empfehlen. Dieser Rechner soll Ende 2005 in einem Neubau in Garching mit einer Leistung von 40 TeraFlop/s installiert werden. Neben den bisher stärksten Nutzern von Höchstleistungsrechnern in den Disziplinen Physik und Ingenieurwissenschaften ist in Zukunft mit einer weiteren Zunahme in den Bereichen Chemie und Biowissenschaften zu rechnen.

