



www.microstaxx.de



## Anwenderbericht

Client/Server  
Computing

High Performance  
Computing

Netzwerke

Open Source

Security

Storage

Service

Topical Compute Cluster an der LMU München

# High Performance Computing für geophysikalische Grundlagenforschung

In der weltweit eng vernetzten Geophysik-Community ist der Lehrstuhl für Geophysik der Ludwig-Maximilians-Universität (LMU) München eines der führenden als auch größten Institute des an sich kleinen aber feinen Fachgebiets. Durch intensive Grundlagenforschung im Bereich Plattentektonik haben die Münchner Geophysiker weltweit die Nase vorn in der Voraussage der bei kontinentalen Plattenbewegungen wirkenden Kräfte. Um von den immer knapper zur Verfügung stehenden Rechenzeiten der großen Rechenzentren unabhängiger zu werden, entschloss sich der Lehrstuhl für Geophysik des Department für Geo- und Umweltwissenschaften, einen eigenen Cluster zur Durchführung von umfangreichen, wissenschaftlichen Simulationen aufzubauen.

## DER KUNDE

Der Lehrstuhl Geophysik im Department für Geo- und Umweltwissenschaften an der Ludwig-Maximilians-Universität München ([www.geophysik.uni-muenchen.de](http://www.geophysik.uni-muenchen.de)) beschäftigt sich mit dem Aufbau und der Dynamik des Erdkörpers, dem Vergleich mit erdähnlichen Planeten, der Struktur von Geomaterialien sowie einem brandaktuellen Thema: der Gefährdung durch Erderschütterungen. Typische Berechnungen in der Geophysik sind große, strömungsmechanische Gleichungssysteme nach Navier-Stokes (z.B. zur Beschreibung von Bewegungen im Erdmantel) und seismologische Modellrechnungen, die einen Rechencluster über mehrere Tage oder sogar Wochen hinweg beanspruchen und dabei nicht nur einmal durchgeführt werden müssen.

## DIE HERAUSFORDERUNG

Diese Aufgaben gehören zu den anspruchvollsten physikalischen Berechnungen die heutzutage durchgeführt werden können. Sie übersteigen die Grenzen der größten HPC (high performance computing) Systeme um einen Faktor von 10 bis 100, sowohl bei den Anforderungen an Kapazität als auch an die Leistungsfähigkeit. Dafür sollte ein mittleres Segment der Datenverarbeitung eingeführt werden, um einerseits das universitätsübergreifende Rechenzentrum zu entlasten und andererseits für die Geophysiker die Möglichkeit zu schaffen, eigene, auch langwierige Berechnungen ohne große Wartezeiten und umfangreiches Antragswesen durchzuführen.

## DIE LÖSUNG

„Topical computing“ (Rechnereinheit zu einem speziellen Thema) war die aufzubauende Lösung. Diese erlaubt die Optimierung der Schlüsselkomponenten (CPU, Netzwerkverbindung, Arbeitsspeicher) für die relevanten Applikationen.



**„Nur durch das fundierte Linux Know-How des MICROSTAXX Teams konnte die anspruchsvolle Cluster-Lösung auf Open Source Basis zeitnah realisiert werden. Das neue System erfüllt all unsere Anforderungen an Performance, Sicherheit und einfache Administration.“**

**Jens Oeser,  
IT-Verantwortlicher in  
der Sektion Geophysik**





www.microstaxx.de



## Anwenderbericht

Client/Server  
Computing

High Performance  
Computing

Netzwerke

Open Source

Security

Storage

Service

So wurde ein eigener Cluster (Tethys – Tectonic high performance simulator) speziell für die Unterstützung der drei wichtigsten Anwendungen aus den geophysikalischen Forschungsbereichen entworfen: Konvektion des Erdmantels, Ausbreitung seismischer Wellenbewegungen und geodynamische Modellierung.

„Serienmäßig gefertigte Hochleistungs-Linux-Cluster sind geeignet, die Grenzen der größten HPC Systeme durch die Verwendung von preisgünstiger PC Hardware von der Stange zu entschärfen“, so Jens Oeser, IT-Verantwortlicher am Lehrstuhl, „so haben wir uns zu dieser Vorgehensweise entschlossen.“

Mit dem Design des Clusters, Aufbau der Hardware sowie der Beratung bei der Installation der Softwareverwaltung und dem Cluster-Management wurde das auf Forschung und Lehre spezialisierte Systemhaus MICROSTAXX ([www.microstaxx.de](http://www.microstaxx.de)) in München beauftragt, dass sich durch besonderes Linux Know-How im high-end Umfeld qualifiziert hat. Nach einer Projektlaufzeit von ca. 3 Monaten wurde der Cluster im Juli 2006 in Betrieb genommen.

Erste Benchmarkergebnisse zeigen eine parallele Effizienz der geophysikalischen Simulationen auf dem Cluster von über 80% und eine kumulierte Systemleistung von mehr als 200 GFLOPs. Durch Optimierungen an den selbst geschriebenen Programmen ist eine weitere Verbesserung vorstellbar. „Wir kamen zu dem Schluss, das kosteneffiziente Beowulf Cluster eine zunehmend größere Rolle in der Ausführung von groß angelegten kapazitätsorientierten, geowissenschaftlichen Simulationen spielen werden,“ so Oeser abschließend.

### TECHNISCHE GEGEBENHEITEN

128 Prozessor AMD Opteron Gigabit Ethernet Linux Cluster

#### **Hardware Spezifikationen der 64 Rechenknoten:**

Anzahl Prozessoren:	2 pro Knoten
Prozessortyp:	AMD Opteron 250 (64 bit, single core)
Prozessorgeschwindigkeit:	2.4 GHz
L1 cache:	64/64KB (data/instruction)
L2 cache:	1MB (data + instruction)
Arbeitsspeicher:	2 GB RAM (DDR1)
Festplattenkapazität:	160 GB
Netzwerkschnittstellen:	1000T Ethernet (2 Ports)

#### **Hardware Spezifikationen des Cluster Head:**

Prozessor:	1 INTEL XEON CPU
Prozessorgeschwindigkeit:	3.0 GHz, Single Core
Arbeitsspeicher:	2 GB RAM (DDR1)
Festplattenkapazität:	2 TB Storage Subsystem
Netzwerkschnittstellen:	mehrere 10 GBit und 1 GBit Ethernet Ports
Sonstiges:	redundante Stromversorgung, Festplatten und Netzwerkschnittstellen



www.microstaxx.de



## Anwenderbericht

Client/Server  
Computing

High Performance  
Computing

Netzwerke

Open Source

Security

Storage

Service

### **Netzwerk:**

1 HP ProCurve 6400cl (6 mal 10 GBit ports) – Cluster Core Switch  
4 HP ProCurve 3400cl (24 mal 1 GBit ports) – Cluster Node Switches

Die Topologie wurde unter Verwendung der HP Mesh Technologie so entworfen, dass der Ausfall des Cluster Core Switch keinen Ausfall des Clusters nach sich zieht.

### **Betriebssystem:**

Die Wahl fiel auf das freie Debian GNU/Linux 3.1 Sarge AMD64 (auf Knoten und Cluster Head), das sich am Lehrstuhl schon vielfach bewährt hat.

### **Software Verteilung:**

Ein PXE-Server dient als Basis für die ferngesteuerte und vollständig automatisierte Installation der Knoten. Installation und Aktualisierung der Knoten übernimmt ein angepasstes FAI (Fully Automatic Installation; entwickelt an der Uni Köln, siehe <http://www.informatik.uni-koeln.de/fai/>).

Ebenfalls über PXE können mittels FreeDOS-basierten Bootdiskettenimages zentral in einem Rebootzyklus BIOS-Updates eingespielt und CMOS-Einstellungen aller Knoten auf einmal gesetzt werden.

Die Kapazität der lokalen Festplatten in jedem Knoten wird durch die Verwendung eines parallelen Filesystems (PVFS2, siehe <http://www.pvfs.org>) als temporärer Zwischenspeicherplatz im gesamten Rechencluster für Berechnungsdaten zur Verfügung gestellt und erreicht eine Kapazität von 8 TB.

### **Management:**

Bei der Überwachung sowie dem Cluster Management wurde bewusst auf teure kommerzielle Pakete verzichtet, statt dessen auf freie Software und eigenentwickelte Skripte gesetzt. Einen Überblick über Auslastung und Aktivität der Knoten, Switches und weiterer Kennzahlen bietet Cacti ([www.cacti.net](http://www.cacti.net)). Die Überwachung und Steuerung der Knoten (Hardwarestatus, Einschalten, Ausschalten, ...) erfolgt über IPMI und selbst geschriebene Shellskripte. Zusätzliche Sensoren von APC behalten den Serverraum im Auge. Geplant ist außerdem der Einsatz der Sun Grid Engine für die Verwaltung der Rechenaufträge.

Ihr Ansprechpartner bei MICROSTAXX:

**Herr Martin Reil**

**MICROSTAXX<sup>®</sup> GmbH**

Wilhelm-Kuhnert-Str. 26  
D-81543 München

Fon: +49-89-48 90 75-0  
Fax: +49-89-48 90 75-99

[www.microstaxx.de](http://www.microstaxx.de)

[vertrieb@microstaxx.de](mailto:vertrieb@microstaxx.de)