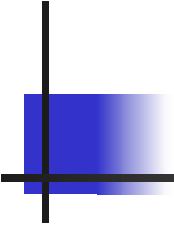


Parallele und verteilte Systeme



**Institut für Informatik
Universität Heidelberg**

Prof. Thomas Ludwig

pvs.informatik.uni-heidelberg.de



Agenda

- Arbeitsgruppen am Instituts
- Bachelorstudiengang
- Lehre
- Forschungsarbeiten
- Kooperationen
- Sonstiges



Arbeitsgruppen am Institut

Prof. **Ambos-Spies**: Theoretische Informatik und mathematische Logik

Prof. Peter **Bastian**: Praktische Informatik und wissenschaftliches Rechnen (IWR)

Prof. Georg **Bock**: Simulation und Optimierung (IWR)

Prof. Donald **Kossmann**: Datenbanksysteme

Prof. Volker **Lindenstruth**: Technische Informatik (Physik)

Prof. Thomas **Ludwig**: Parallele und verteilte Systeme

Prof. Barbara **Paech**: Software-Systeme

Prof. Gerhard **Reinelt**: Diskrete Optimierung

Prof. Gabriel **Wittum**: Technische Simulation (IWR)



Bachelorstudiengang

- Begonnen WS 2001/02
- 3-jährige Ausbildung, starker Matheanteil
- 60 Plätze pro Jahrgang
- Belegung: 10 / 25 / 20
- Ab WS 2004/05 Masterstudiengang

- Grundstudium
 - Betriebssysteme und Netzwerke
14+13 Vorlesungsstunden
 - Softwarepraktikum „Parallele Programmierung“
- Hauptstudium
 - Cluster-Computing
 - Softwarepraktikum „Aufbau und Betrieb von Rechnerclustern“
 - Seminare: „Betriebssystemaspekte des Cluster-Computing“



Forschungsgebiet

Parallele Ein-/Ausgabe für Hochleistungsrechner

Parallele Ein-/Ausgabe

- Auf der Programmierenebene
 - Viele parallele Prozesse schreiben in dieselbe Datei
 - Zugriff zu verschiedenen/identischen Positionen
 - Disjunkte/nichtdisjunkte Sicht auf die Dateien
- Auf der Systemebene
 - Datei ist physikalisch über viele Platten verteilt
- Typisches Abbildungsproblem; Lokalität wichtig



Fragestellungen

- Benutzerebene
 - Welche Zugriffssemantik eignet sich für welche Anwendung? (gemeinsamer/individueller Dateizeiger, gemeinsame/individuelle Datenbereiche)
 - Wie kann ich meine E/A charakterisieren? (sequenziell, zufällig, kleine/große Datenmengen)
 - Wie kann ich die E/A-Leistung messen?
 - Wie kann ich meinen Datendurchsatz optimieren?



Fragestellungen

■ Systemebene

- Wie werden die Daten einer Datei über die Platten hinweg verteilt?
- Wie mißt man hier die Lasten im System?
- Automatische Lastoptimierung durch Verlagerungen von Dateiteilen?
- Wie optimiert man Zugriffe auf die Metadatenserver? (Zentrale Komponente als Engpaß im System)
- Was passiert beim Absturz des Programms und beim Ausfall von Platten?
- Wie übertrage ich Riesendateien auf andere Systeme? (Datensicherung)
- Wie verwende ich Riesendateien in anderen Programmen? (Datenvisualisierung)



Laufende Arbeiten

Drei Dissertationen zum Thema

- Hipolito Vasquez (AG PVS)
 - Dynamischer Lastausgleich in parallelen Dateisystemen
- Stefan Friedel (IWR)
 - Einsatz paralleler E/A im wissenschaftlichen Rechnen
- Christoph Biardzki (LRZ, München)
 - Optimierung der Leistung der Metadatenserver in parallelen Dateisystemen



Laufende Arbeiten...

- Bachelor-Arbeiten

- Performance Analysis Framework

- Einbau von Instrumentierungen in PVFS2

- Anschluß von trace-basierten und online-Werkzeugen

- Vergleich verschiedener Verteilungen

- Anschluß an die distribution-Schnittstelle in PVFS2

- Evtl. dynamisch änderbare Verteilungen



Anwendungsfelder

- Numerische Anwendungen mit massivem Datenaufkommen
 - 1-10 GByte/Rechnerknoten
 - Physik (IWR und andere)
 - Steigendes Interesse Dritter
 - Aufnahme in Ausbuantrag zu Helics
 - Bioinformatik
 - Noch Partner gesucht



Kooperationen

- Leibniz-Rechenzentrum München
Dissertation Biardzki
- Argonne National Labs: Rob Ross
Erweiterungen zu PVFS
- ParTec München
Erweiterungen der Para-Station
Verwaltungssoftware für Cluster
- Lehrstuhl Prof. Bode München
Bioinformatik (1 weitere Dissertation)

Kooperation mit PVFS-Gruppe

- Früher Zugriff auf Quellen von pvfs2
- Deutliche Überarbeitung ggü. pfvs (=pvfs1)
- Mehr interne Schnittstellen
 - Zur Vernetzung: BMI
 - Zur eigentlichen Datenspeicherung: Trove
 - Zur Verteilungssteuerung: Distribution
- Kernel 2.4 für paralleles FS
 - 2.6 für Unix-Zugriffe

Test- und Produktions-Cluster

- Eigenes Cluster $(4+1)*2$
Doppelprozessorboards XEON 2.4GHz mit Gigabit-Ethernet
- Helics-Cluster: 256 Doppelprozessorboards
AMD Athlon, 2GByte/Knoten, Myrinet/2000,
Clos-Netzwerk, MPICH/GM, PBSpro/maui