

Hochleistungsrechencluster zur Unterstützung der biomedizinischen Forschung

GMDS – Mannheim, 6. September 2010

B. Löhnhardt, M. Quade, D. Skrowny, M. Sohns,

Prof. Dr. H. Bickeböller, Prof. Dr. U. Sax



Vortragsinhalte

- Herausforderungen in der aktuellen Forschung
- Praxisbeispiel „Genetische Epidemiologie“
- Hochleistungsrechencluster in virtuellen Forschungsumgebungen
- Evaluation
 - Vergleichsaufbau
 - Vergleichsergebnisse
- Diskussion und Ausblick



<http://www.wissgrid.de>



<http://services.medigrid.de>

Herausforderung in der aktuellen Forschung

- Hohe Bedarfe an Rechenleistung existieren in der med. Forschung.
- Oftmals ist keine ausreichende IT-Infrastruktur innerhalb der Forschungsinstitute verfügbar.
- Die gegenwärtige Förderpolitik sieht eine getrennte Förderung von Forschung und der dafür benötigten IT-Infrastruktur vor.
- Die Beschaffung von dedizierten institutsinternen Rechenclustern ist oft nicht sinnvoll:
 - Hoher Investitions- und Personalaufwand besteht.
 - Oftmals ist keine regelmäßige Nutzung erforderlich, dafür können Peaks vorhanden sein.

3

Praxisbeispiel

- Nationales Genomforschungsnetz Herzinsuffizienz:
 - Eine genomweite Studie zur diastolischen Herzinsuffizienz wird durchgeführt.
 - Die statistische Auswertung der umfangreichen Primärdaten erfolgt durch die Genetische Epidemiologie der Universitätsmedizin Göttingen.
- Imputationsanwendung:
 - Nicht genotypisierte Marker werden in den Studiendaten geschätzt.
 - Verwendete Software: Markov Chain based Haplotyper (MACH)
- Konkrete Probleme:
 - Computerraum oder eigener Arbeitsplatzrechner werden genutzt.
 - Hoher Installations- und Wartungsaufwand vorhanden.

4

Lösung „Hochleistungsrechencluster“

- Visualisierungscluster der Universitätsmedizin Göttingen:
 - Besteht aus 9 Rechenknoten mit 96 Prozessoren.
 - Rechenjobs im Batchbetrieb können verarbeitet werden.
 - Umfangreiche lokale sowie Remote-Visualisierung möglich.
- Verbund von Ressourcen (Grid Computing):
 - Deutsche Grid Initiative (D-Grid)
 - MediGRID



Grid Computing

(Quelle: <http://heprc.phys.uvic.ca>)



Visualisierungscluster Göttingen

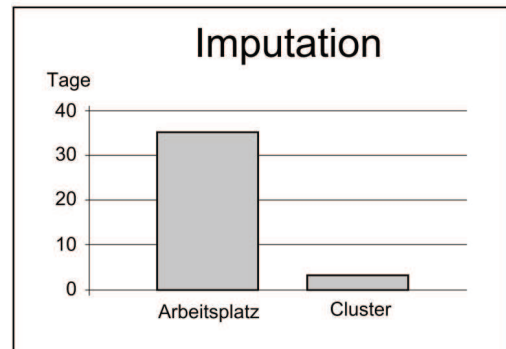
5

Evaluation: Aufbau

- Imputation mit MACH:
 - Die Imputation jedes Chromosoms erfolgt in 2 voneinander abhängigen Phasen.
 - Getrennte Imputation für Sample genotypisiert auf 500K Affymetrix-Chips und Sample genotypisiert auf 6.0 Affymetrix-Chips.
 - Als Referenz dient HapMap2.
 - Insgesamt sind 88 Einzelschritte (bei 22 Chromosomen) notwendig.
- Durchführung in zwei Arbeitsumgebungen:
 - Arbeitsplatzrechner in der Genetischen Epidemiologie
 - Visualisierungscluster der Universitätsmedizin Göttingen

Evaluation: Ergebnisse

- **Gesamtlaufzeit:**
 - Arbeitsplatzrechner: ~830 Stunden
 - Visualisierungscluster: ~82 Stunden
- **Zeiteinsparung:**
 - **über 90 %**
 - **mehr als 1 Monat Ersparnis**



Vergleich der Berechnungsdauer in Tagen

- **Weitere Vorteile der Nutzung des Clusters für den Forscher:**
 - Installations- und Wartungsaufgaben werden übernommen.
 - Routinen zur Datensicherung, dem Datenmanagement und der Langzeitarchivierung werden zur Verfügung gestellt.
 - Es stehen dem Wissenschaftler mehr Zeit für die eigentlichen Schwerpunkte seiner Forschung zur Verfügung.

7

Diskussion und Ausblick

- Erhebliche Einsparungen durch die Verwendung des Clusters.
- Als Bestandteil einer virtuellen Forschungsumgebung wird dem Forscher der Zugriff auf verteilte Ressourcen ermöglicht.
- **Ausblick:**
 - Rechenoperationen können durch CUDA beschleunigt werden.
 - Eine dynamische, bedarfsgerechte Einbindung weiterer Grid-Ressourcen wird im weiteren Projektverlauf angestrebt.
 - Aufwendige Visualisierungen der Ergebnisse sind möglich.

8

Kontaktmöglichkeiten



UNIVERSITÄTSMEDIZIN : UMG
GÖTTINGEN

Universitätsmedizin Göttingen

Medizinische Informatik

<http://www.mi.med.uni-goettingen.de/>

Benjamin Löhnhardt

Computational Medicine und Grid-Computing

Benjamin.Loehnhardt@med.uni-goettingen.de

Tel.: (0551) 39 - 22842



<http://www.wissgrid.de>



<http://services.medigrid.de>