

Linux Cluster für bioinformatische Anwendungen

Das Ziel dieser Diplomarbeit war Design, Planung und Aufbau eines Beowulf Clusters, der hardwaremäßig aus gewöhnlichen PCs besteht, die über ein Netzwerk verbunden sind und gemeinsam parallele Applikationen unter Linux berechnen können. Dabei wurde der Cluster und die Applikationen auf die speziellen Bedürfnisse der Bioinformatik zugeschnitten.

Aus dieser Diplomarbeit resultiert ein voll funktionstüchtiger Beowulf Cluster, der aus 24 Knoten mit insgesamt 48 Prozessoren und einem Master mit weiteren 2 Prozessoren besteht. Die Knoten untereinander sind mit zwei 100 Mbps FastEthernet-Karten verbunden und greifen damit auf das 240 GB große Plattensystem des zentralen Masters zu, der mit Gigabit-Ethernet integriert ist. Für die Überwachung des Clusters wurde ein plattformunabhängiges Monitoring-tool entwickelt, das laufend aktuelle Statusinformationen des Clusters anzeigt. Weiters wurden gängige Werkzeuge der Bioinformatik wie SRS (Sequence Retrieval System) , BLAST (Basic Local Alignment Search Tool), WU-BLAST und FASTA für den Clusterbetrieb installiert. Dieses Tool stellt eine Kernkomponente eines genomischen Informationsmanagementsystems dar, die für Forschungsarbeiten der modernen Biowissenschaften unabdingbar ist.

Schlüsselworte: Parallelisierung, Bioinformatik, Beowulf Cluster, Linux

Linux Cluster for Bioinformatic Applications

Objective of this master thesis was the design and creation of a Beowulf Cluster consisting of lowcost PC's interconnected by a network and executing parallel tasks running on Linux. The cluster and the applications were optimized for bioinformatic purposes. This master thesis resulted in a fully functional compound of 24 PC's with 48 processors including a master with two additional processors. The nodes are interconnected with two 100 Mbps FastEthernet-Networks and access the 240 GB disksystem of the central master through a Gigabitinterface. Additionally, a platform independent monitoring tool was developed and common bioinformatic applications were installed: SRS (Sequence Retrieval System), BLAST (Basic Local Alignment Search Tool), WU-BLAST and FASTA The developed tool represents a central component of a genomic informationmanagement system which is essential for modern molecular Biosciences.

Keywords: Parallelisation, Bioinformatics, Beowulf Cluster, Linux,