2.1 Teilprojekt B7

Management von Cluster–Systemen

2.1.1 Antragsteller

Prof. Dr. Ing. Wolfgang Rehm
27.09.1950
Professur Rechnerarchitektur
und Mikroprogrammierung
Fakultät für Informatik
Technische Universität Chemnitz
09107 Chemnitz
Tel.: (0371) 531-1420
Fax: (0371) 531-1806
rehm@informatik.tu-chemnitz.de

2.1.2 Projektbearbeiter

Dipl.-Inf.(FH) C.Dinkelmann
Technische Universität Chemnitz, Fakultät für Informatik

2.2 Ausgangsfragestellung/Einleitung

Cluster-Computing wird allgemein als Lösungsansatz für preiswerte Parallelverarbeitung angesehen. Aus diesem Grund wurden an der Professur Rechnerarchitektur drei Cluster-Prototypen realisiert und teilweise im Rahmen des SFB genutzt. Mit zunehmender Knotenanzahl rückt hierbei das Problem der Administration des Clusters in den Vordergrund. Zu lösen sind insbesondere folgende Aufgaben:

- *Systemverwaltung*. Notwendig ist eine Lösung zur weitgehend automatischen Inbetriebnahme, Konfiguration und Wartung von Knoten des Clusters. Dies schließt ein Recovery im Fehlerfall ein.

- *Software-Konfigurationsmanagement*. Im Gegensatz zu verschiedenen Parallelrechnern erlauben Cluster in der Regel den direkten Nutzerzugriff auf einzelne Knoten des Systems. Dies erfordert die Bereitstellung einer einheitlichen Softwareumgebung auf allen Knoten sowie die Organisation von deren Wartung.

- *Nutzerverwaltung*. Bei der Zuweisung von Resourcen an Nutzer (Menschen und Programme) müssen neben sequentiellen auch kooperierende Prozesse unterstützt werden.

Aus der im Gegensatz zu Rechnernetzen unterschiedlichen Knotenausstattung sowie unterschiedlichen Zugangsmechanismen ergeben sich spezifische Anforderungen, aber auch neue Möglichkeiten, beispielsweise beim Vorhandensein eines lokalen Kommunikationsnetzes.

Aufgrund der Tatsache, dass für Teilprojekt B6 für den Antragszeitraum von 1999–2001 eine halbe Stelle weniger bewilligt wurde als beantragt, und dass es einen unvorhergesehenen Weggang eines Mitarbeiters aus der Grundausstattung für Teilprojekt B6 gab, konnte Teilprojekt B7 nicht im vollen Maße wie ursprünglich vorgesehen durchgeführt werden. Aufgrund dessen wurde die für Teilprojekt B7 beantragte und bewilligte halbe Stelle nach Ablauf der halben Projektzeit mehr auf Teilprojekt B6 konzentriert.

2.3 Forschungsaufgaben/Methoden

Mit zunehmender Akzeptanz des Cluster–Computings für wissenschaftliches Rechnen und zunehmender Größe solcher Systeme spielt deren Administration eine wachsende Rolle. Dies schlägt sich sowohl in der Komplexität der zu lösenden Aufgaben als auch im hierzu nötigenden Zeitaufwand nieder.

Voraussetzung für die Verringerung des administrativen Aufwandes durch nachnutzbare Konfigurationen und Automatisierung ist zunächst die Systematisierung der Management–Aktivitäten in Zusammenhang mit Cluster–Systemen. Diese umfassen:

- *Systemverwaltung.* Der Administrator benötigt Methoden und Werkzeuge zur weitgehend automatischen Inbetriebnahme, Konfiguration und Wartung von Knoten.

- *Software–Konfigurationsmanagement.* Auf Knoten von Cluster–Systemen wird im Gegensatz zu verschiedenen Parallelrechnern in der Regel eine vollwertige Softwareumgebung bereitgestellt. Hier sind Werkzeuge erforderlich, die die Organisation und Wartung von nahezu identischen Softwareumgebungen auf allen Knoten effizient ermöglichen.

- *Nutzerverwaltung.* Notwendig ist die Zuweisung von Resourcen wie Prozessorzeit, Partition und Hauptspeicher an Anwender oder von ihnen gestartete parallele oder sequentielle Programme. Die Lastverteilung muß berücksichtigen, daß sich die Lastanforderungen paralleler Anwendungen zur Laufzeit dynamisch ändern können.

Dabei sollten folgende Teilaufgaben gelöst werden:

1. Analyse der beim Betrieb des Clusters anfallenden Aufgaben mit dem Ziel der Formalisierung des Aufgabenspektrums. Das Ergebnis wird anschließend hinsichtlich von Kriterien wie Häufigkeit des Auftretens, Arbeitsaufwand, Automatisierbarkeit usw. klassifiziert.

2. Analyse bereits verfügbarer Managementlösungen und Konzepte und Beurteilung von deren Einsetzbarkeit für das Management von Clustersystemen anhand des erarbeiteten Aufgabenkatalogs.

3. Auswahl eines Ansatzes und Erstellung eines Managementkonzeptes unter Nutzung der Ergebnisse aus (1) und (2).

4. Prototypische Realisierung der Managementlösung und deren Bewertung anhand des am Lehrstuhl vorhandenen und im Rahmen des SFB genutzten Clustersystems.

Wie Eingangs beschrieben, konnten diese Aufgaben nicht wie geplant in dem Maße durchgeführt werden.

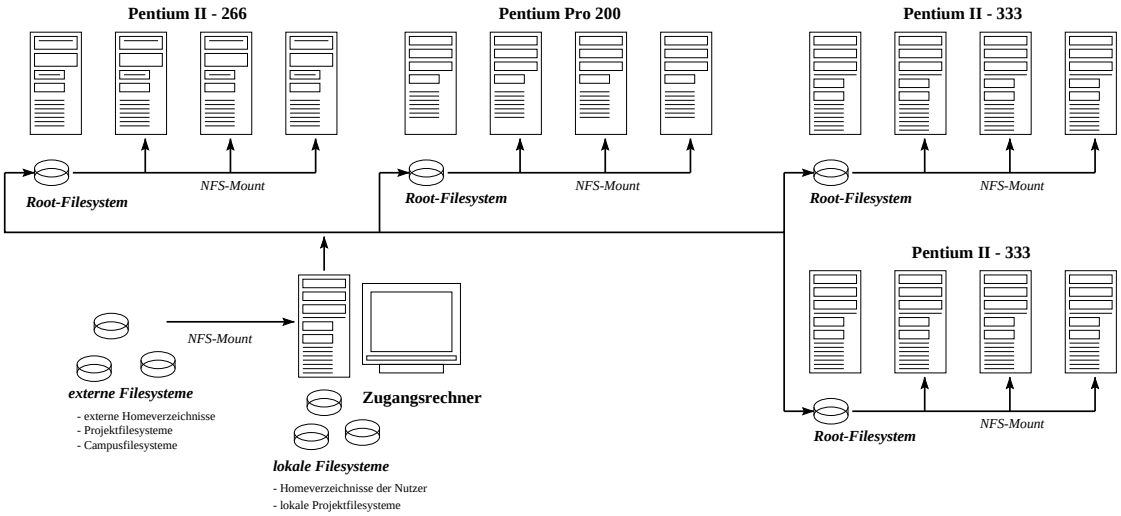
2.4 Ergebnisse

2.4.1 Cluster Infrastrukturen

Um das Management von mittelgroßen Clustern (um die 16 Knoten) zu erleichtern, wurde ein sog. *Submaster*–Konzept eingeführt und erprobt [DRM99]. Die Motivation für dieses Konzepte war dabei zum einen die Reduzierung des Wartungsaufwandes eines Clusters und zum anderen die Erhöhung der Zuverlässigkeit durch Reduzierung von Festplatten, welche sich oft als Schwachpunkt erweisen (mechanische Bauelemente, hohe Beanspruchung, ...).

Abbildung 2.1 zeigt die prinzipielle Cluster–Infrastruktur, wie sie Anfang 1999 in der Ausbaustufe "OSCAR–IV"realisiert wurde.

Abbildung 2.1: Infrastruktur des OSCAR–IV Clusters



Das Submaster–Konzept stellt sich hierbei dadurch dar, dass lediglich ein kleiner Teil der Rechenknoten mit Festplattem ausgerüstet ist (hier jeder vierte). Konkret bedeutet das in diesem Fall, dass sich mehrere Rechner eine Festplatte teilen. D.h.

das Root–Filesystem von mehreren Rechenknoten befindet sich physisch auf einem der Rechenknoten.

Bis dahin existierende Lösungen sahen entweder für keinen der Rechner oder für alle Knoten eine Festplatte vor. Vor– und Nachteile dieser Lösungen liegen dabei auf der Hand.

Während ein vollkommen plattenloser Cluster, bei dem alle Rechenknoten ihr Root–Filesystem von einem einzigen Rechner (in der Regel dem Zugangsrechner) beziehen, den Installations– und Wartungsaufwand auf ein Minimum reduziert, konzentriert sich die NFS Last (Netzwerk Filesystem) auf einen Rechner, wodurch sich die Skalierbarkeit des Systems stark verschlechtert.

Im Gegensatz dazu ist die Skalierbarkeit bzgl. dieses Problems sehr gut, wenn alle Rechenknoten ihre eigene Festplatte besitzen. Jedoch steigt damit auch der Administrationsaufwand und die Ausfallwahrscheinlichkeit des Clusters an.

Das Submaster–Konzept sollte dabei eine Zwischenlösung darstellen, welche einen Kompromiss aus beiden konventionellen Lösungen darstellt.

In der Praxis hat sich dieser Kompromiss zumindest bei dem Forschunscluster "OSCAR" bewährt.

Im Jahr 2000 wurde der Forschungscluster wie geplant im Rahmen des SFB 393 erweitert. Dabei wurden 7 auf Alpha–Prozessoren basierende Knoten hinzugefügt. Diese Knoten wurden in der Art in den Cluster engebettet, dass sie ihr Root–Filesystem direkt vom Zugangsrechner beziehen.

Für die Zukunft ist hier angedacht, alle Rechenknoten des Clusters ohne Festplatten zu betreiben, jedoch eine bestimmte Anzahl dedizierter Root–Filesystem–Server einzuführen, deren einzige Aufgabe die Bereitstellung von Root–Filesystemen für eine bestimmte Anzahl von Rechenknoten ist.

2.4.2 Cluster Management durch Nutzer

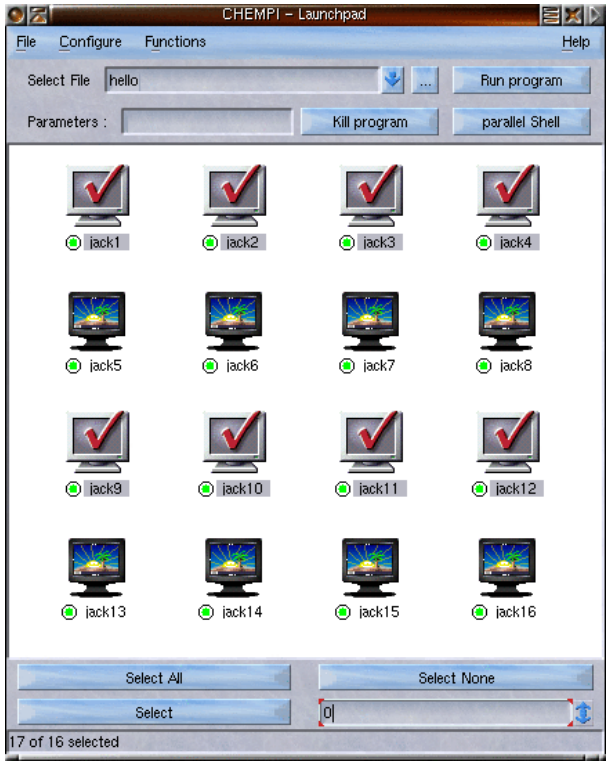
In engem Zusammenhang mit Teilprojekt B6 wurde mit der Entwicklung eines grafischen FrontEnds zum Starten von verteilten Message–Passing Applikationen auf einem Cluster begonnen [DUE00]. Dieses Tool wurde *CHEMPI LaunchPad* genannt, ist jedoch auch für andere MPI Implementationen (insbesondere MPICH) einsetzbar.

Abbildung 2.2 zeigt beispielhaft das Hauptfenster dieses Tools.

Hauptziel dieser Entwicklung war es zunächst, dem Anwender ein Tool zu geben, mit dem er intuitiv verschiedene Rechner eines Clusters für seine verteilte Anwendung auswählen kann, ohne dabei manuell diverse Konfigurationsdateien editieren zu müssen.

Dieses grafische FrontEnd ist jedoch momentan noch nicht für große Clustersysteme geeignet.

Abbildung 2.2: Screenshot des CHEMPI LaunchPads



Literaturverzeichnis

Begutachete Literatur

[DRM99] Carsten Dinkelmann, Wolfgang Rehm und Marko Meyer: *Eine Konfigurations- und Managementlösung für ein dediziertes Linux-Cluster*, 2.Workshop Cluster Computing, 25./26. März 1999, Universität Karlsruhe. Chemnitzer Informatik-Berichte CSR-99-02 (157 S.). ISSN 0947-5125

Sonstige Arbeiten (Diplomarbeiten, Technical Reports etc.)

[DUE00] Jörg Dümmler: *Entwicklung eines grafischen Front-Ends für CHEMPI*, Studienarbeit, Technische Universität Chemnitz, Fakultät für Informatik, August 2000

2.5 Offene Fragen/Ausblick

Das Management von Clustern ist und bleibt ohne Frage ein wichtiger Aspekt. Dies betrifft beide Arten das Managements, wie sie auch hier besprochen wurden:

1. Wie lässt sich ein Cluster aus Sicht des Administrators verwalten?
   D.h. wie wird NFS–Last verteilt, wie wird Software verwaltet etc.

2. Wie benutzt ein Anwender den Cluster?
   D.h. wie und auf welchen Knoten startet er seine Applikation, inwiefern sieht

er den Cluster als “Single System Image”

Dies sind interessante Fragestellungen, welche für eine Verbesserung von Cluster–Systemen nach wie vor von Bedeutung sind.

In der Zukunft sollen diese Dinge jedoch nicht vordergründiges Forschungsthema der Arbeitsgruppe des Antragstellers sein. Nichtsdestotrotz sollen diese Fragestellungen nicht ganz vernachlässigt werden, da sie essentiell zum Betrieb eines Clusters sind — Auch, wenn es sich dabei um einen Forschungscluster handelt.

2.6 Cluster–Konferenzen

Obwohl dies nicht in unmittelbarem Zusammenhang mit dem konkreten Forschungsinhalt von Teilprojekt B7 steht, soll hier kurz auf die vom Antragsteller im Antragszeitraum durchgeführten Workshops bzw. Konferenzen eingegangen werden.

2.6.1 2. Workshop Cluster–Computing

Die im November 1997 vom Antragsteller initiierte deutschlandweite Workshop Serie *Cluster–Computing* wurde im März 1999 erfolgreich mit dem 2. Workshop Cluster–Computing als wichtigstes Austauschsforum auf diesem Arbeitsgebiet innerhalb Deutschlands weitergeführt [CC99, RUCC99, RUCC99+]. Der Workshop wurde dabei in Zusammenarbeit mit der Universität Karlsruhe organisiert, welche gleichzeitig als Tagungsort diente.

2.6.2 CLUSTER2000 — IEEE International Conference on Cluster Computing

Im November/Dezember 2000 konnte die Bedeutung der Technischen Universität Chemnitz auf dem Gebiet des Cluster Computing noch weiter hervorgehoben werden, in dem eine neue Konferenz von internationalem Rang — die *IEEE International Conference on Cluster Computing* organisiert und auch veranstaltet wurde [CC2000, BR2000].

Mit über 160 Teilnehmern aus 22 Ländern und 62 wissenschaftlichen Beiträgen war die CLUSTER2000 dabei eine der wichtigsten internationalen Konferenzen in diesem Themenbereich.

Literaturverzeichnis

[BR2000] Mark Baker* und Wolfgang Rehm**: *Proceedings of the CLUSTER2000 — IEEE International Conference on Cluster Computing*, 28.Nov.–1.Dez. 2000, Chemnitz, ISBN 0-7695-0896-0
Publication chair, **Co-Editor (General chair)

[CC2000] Homepage der *CLUSTER2000 — IEEE International Conference on Cluster Computing*:
`http://www.tu-chemnitz.de/cluster2000`

[CC99] Homepage des *2. Workshop Cluster–Computing*: `http://www.tu-chemnitz.de/informatik/RA/CC99/`

[RUCC99] Wolfgang Rehm und Theo Ungerer (Ed.): *Tagungsband zum 2. Workshop Cluster–Computing*, 25./26. März 1999, Universität Karlsruhe. Chemnitzer Informatik-Berichte CSR-99-02 (157 S.). ISSN 0947-5125

[RUCC99+] Wolfgang Rehm und Theo Ungerer (Ed.): *Ausgewählte Beiträge zum 2. Workshop Cluster–Computing*, 25./26. März 1999, Universität Karlsruhe. Preprint-Reihe SFB 393/97-22 (117 S.). 6 6 6 6 6