

Ace Crngarov

Evaluierung von Cluster-Dateisystemen für den Einsatz auf Parallelrechnern

Bachelorarbeit

BEI GRIN MACHT SICH IHR WISSEN BEZAHLT



- Wir veröffentlichen Ihre Hausarbeit, Bachelor- und Masterarbeit
- Ihr eigenes eBook und Buch - weltweit in allen wichtigen Shops
- Verdienen Sie an jedem Verkauf

Jetzt bei www.GRIN.com hochladen
und kostenlos publizieren



Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2006 Examicus Verlag
ISBN: 9783869438450

Ace Crngarov

Evaluierung von Cluster-Dateisystemen für den Einsatz auf Parallelrechnern

Examicus - Verlag für akademische Texte

Der Examicus Verlag mit Sitz in München hat sich auf die Veröffentlichung akademischer Texte spezialisiert.

Die Verlagswebseite www.examicus.de ist für Studenten, Hochschullehrer und andere Akademiker die ideale Plattform, ihre Fachtexte, Studienarbeiten, Abschlussarbeiten oder Dissertationen einem breiten Publikum zu präsentieren.



**Fachhochschule
Bonn-Rhein-Sieg**
University of Applied Sciences

Fachbereich Informatik
Department of Computer Science

Abschlussarbeit

im Bachelor Studiengang

Evaluierung von Cluster-Dateisystemen für den Einsatz auf Parallelrechnern

von
Ace Crngarov

Eingereicht am: 07.08.2006

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Die Arbeit wurde bisher keiner Prüfungsbehörde vorgelegt und noch nicht veröffentlicht.

Sankt Augustin, 07.08.2006

(Ace Crngarov)

Zusammenfassung

Zur Ein- und Ausgabe großer Datenmengen auf Parallelrechnern und Cluster-Systemen werden spezielle Dateisysteme eingesetzt, die den parallelen Zugriff von mehreren Rechnern gleichzeitig effizient unterstützen. Beispiele für solche Cluster-Dateisysteme sind Parallel Virtual File System (PVFS / PVFS2), Oracle Cluster File System (OCFS2), Red Hat Global File System (GFS), IBM General Parallel File System (GPFS) und Lustre.

Diese Arbeit evaluiert die oben genannten Produkte hinsichtlich ihrer Effizienz und praktischen Einsetzbarkeit in einer Parallelrechnerumgebung. Zu Beginn werden die verschiedenen Cluster-Dateisysteme vorgestellt, der Schwerpunkt liegt hier auf frei verfügbaren Dateisystemen. Weiter legt diese Arbeit geeignete Effizienzkriterien und Strategien zur Beurteilung solcher Dateisysteme fest. Zu den Kriterien zählen neben einer hohen Transferrate auch eine geringe Latenzzeit bei Zugriffen auf kleine Dateien. Ein Benchmark-Programm setzt die festgelegten Kriterien und die Strategie um. Dazu wird ein neues Benchmark-Programm entwickelt, da kein verfügbares Programm alle festgelegten Kriterien untersuchen kann.

Die vorgestellte Auswahl an Cluster-Dateisystemen wird mit dem entwickelten Benchmark-Programm untersucht. Diese Arbeit präsentiert und vergleicht die gewonnenen Ergebnisse. Abschließend werden die resultierenden Leistungsdaten analysiert und darauf aufbauend Empfehlungen gegeben.

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	Ziel und Motivation dieser Arbeit	1
1.2	Vorgehensweise.....	2
1.3	Abgrenzung	2
2	Dateisysteme für Parallelrechner-Cluster.....	3
2.1	Begriffsdefinitionen und Bedeutung	3
2.1.1	Cluster	3
2.1.2	Cluster-Dateisystem	4
2.1.3	Datei-intensive Anwendungen	4
2.2	Architekturen von Cluster-Dateisystemen.....	5
2.2.1	Shared Storage	5
2.2.2	Intelligente Server	5
2.3	Technische Übersicht	6
2.3.1	Parallel Virtual File System 2	6
2.3.2	Lustre.....	8
2.3.3	Oracle Cluster File System 2.....	9
2.3.4	General Parallel File System	9
2.3.5	Global File System.....	10
2.3.6	Weitere Cluster-Dateisysteme.....	11
2.3.7	Techniken zur Optimierung von Zugriffen	12
3	Kriterien zur Beurteilung von Cluster-Dateisystemen	14
3.1	Zugriffsmuster wissenschaftlicher Anwendungen	14
3.2	Festlegung der zu untersuchenden Leistungsdaten.....	16
3.3	Erhebung der festgelegten Leistungsdaten in Form von Testszenarien	17
3.3.1	Sequentieller Zugriff, gemeinsame Datei.....	18
3.3.2	Sequentieller Zugriff, unterschiedliche Dateien.....	18
3.3.3	Schrittweiser Zugriff	19
3.3.4	Metadaten.....	19
3.3.5	Pufferfähigkeit.....	20
3.3.6	Systemauslastung	20
4	Umsetzung der konzipierten Szenarien und Testvorbereitung.....	21
4.1	Benchmark-Programm zur Umsetzung der Szenarien	21
4.1.1	Anforderungen an ein Benchmark-Programm	21
4.1.2	Auswahl eines Benchmark-Programms	22
4.1.3	Entwicklung und Umsetzung eines Benchmark-Programms	23
4.2	Testvorbereitung	27
4.2.1	Testumgebung	27
4.2.2	Gruppierung der Rechner.....	28
4.2.3	Installation und Konfiguration	30
5	Ergebnisse der Testdurchführung.....	32
5.1	Testergebnisse der Szenarien.....	32
5.1.1	Sequentieller Zugriff, gemeinsame Datei.....	32
5.1.2	Sequentieller Zugriff, unterschiedliche Dateien.....	36
5.1.3	Schrittweiser Zugriff	42
5.1.4	Metadaten.....	49