

DCA-IO: A Dynamic Control Scheme for Parallel And Distributed File Systems

Seminar Supercomputer

Camila Chavez Lopez

Arbeitsbereich Wissenschaftliches Rechnen
Fachbereich Informatik
Fakultät für Mathematik, Informatik und Naturwissenschaften
Universität Hamburg

2022-12-13

Gliederung (Agenda)

- 1 Themenvorstellung
- 2 Lustre
- 3 DCA-IO
- 4 Zusammenfassung
- 5 Literatur

Abbildung: Wetterkarte [Wet]

Abbildung: Urknallsimulation [ntv10]

Kernaussagen des Papers

- I/O ist ein Bottleneck
- Konfigurationen in Dateisystemen werden nicht vorgenommen
- Entwicklung eines Algorithmus, der keine Nutzerinteraktion benötigt
- Algorithmus verändert die Parameter Stripe Size und Stripe Count

Ergebnisse des Papers

- Konfigurationen in Dateisystemen:
 - in einem isolierten System eine Steigerung der Leistung um 75%
 - in CORI eine Steigerung der Leistung um 50%

Komponenten von Lustre

Abbildung: Architektur von Lustre [KSW19]

Datenstruktur und Datei I/O

Abbildung: File Striping [Rie]

Komponenten von Lustre II

- Metadatenserver (MDS)
 - Zuständig für Metadaten im Dateisystem
 - Regelt Zuweisung von Speicherorten und Dateien
- Metadatentarget (MDT)
 - Speicherung der eigentlichen Metadaten
 - Jedes Dateisystem hat mind. einen MDT
- Quelle: [Rie]

Komponenten von Lustre III

- Objekt Storage Server (OSS)
 - Verantwortlich für OSTs
 - Behandelt Netzwerkanfragen sowie I/O Zugri
- Objekt Storage Target (OST)
 - Plattenspeicher, der Daten der User speichert
 - Dateien werden objekbasiert gespeichert
- Quelle: [Rie]

CORI

- Cori ist Platz 45 auf der Top 500 Liste [500]
- HPC System von NERSC
- Steht in Berkley, California, USA
- Ist ein XC40 System mit einer Höchstleistung von 30 Peta ops
- Quelle: [Doc]

Analysieren der Logs

- Darshan
 - Tool, um das I/O Verhalten von Systemen besser zu analysieren [SCH[†] 16]
 - Darshan default I/O Analysetool in Cori
 - Wenn eine Anwendung läuft zeichnet Darshan I/O Funktionsaufrufe auf
 - Gesammelte Information:
 - Anzahl I/O Operationen
 - Typ von I/O Operationen
 - Weitere statistische Daten
- Extrahieren der Informationen aus Darshan Logs

Analysieren der Logs

Abbildung: Log File [KSW19]

Anlegen einer Datenbank

- Speichern in einer Datenbank erlaubt das Erkennen von Zusammenhängen

Abbildung: Tabelle mit Datenbankinformationen [KSW19]

Ergebnisse der Analyse

- Die meisten Nutzer benutzen Lustre default Konfigurationen
 - 99.317% der Ausführungen benutzen den default Stripe Count
 - 99.948% der Ausführungen benutzen die default Stripe Size
- Es gab 1284643 Ausführungen von 1163 Anwendungen
- ! Nutzer schöpfen die Kapazitäten die sie in einer HPC Umgebung bekommen nicht richtig aus

Initial Execution

- Es existiert kein Wissen über die auszuführende Anwendung
- DCA-IO versucht Einstellungen zu erraten, da I/O Verhalten unbekannt
- Verwendung der Anzahl der Prozesse als Entscheidungsmerkmal

Initial Execution II

Abbildung: Pseudocode Initial Execution [KSW9]

Recurring Execution

- Es gibt Einträge zu dieser Anwendung in der Datenbank
- Informationen aus den Logs zur Optimierung der Einstellungen
- Rule Based Phase
 - Optimierung anhand von existierenden Regeln
- Heuristic Phase
 - Stripe Count wird linear erhöht

Recurring Execution II

Abbildung: Pseudocode Recurring Execution [KSW0]

Evaluieren von DCA-IO

- Benchmarks: FIO, PLOK, ...
- Analyse der Performance von DCA-IO in lokaler Umgebung
- Analyse der Performance von DCA-IO in CORI

Evaluieren von DCA-IO II

Abbildung: VPIC-IO Performance in CORI [KSW9]

Evaluationsgebniss von DCA-IO

- Die Verwendung von DCA-IO kann die Performance in einer lokalen Umgebung um bis zu ~~75~~⁷⁵ verbessern
- Die Verwendung von DCA-IO in einem großen HPC System, wie Cori, kann die Performance um bis zu ~~50~~⁵⁰ verbessern.

Zusammenfassung

- Konfigurationen in parallelen und verteilten Dateisystemen sind notwendig zur Steigerung der I/O Performance.

Zusammenfassung

- Konfigurationen in parallelen und verteilten Dateisystemen sind notwendig zur Steigerung der I/O Performance.
- DCA-IO ist ein Verfahren, dass Lustre Konfigurationen dynamisch anpasst

Zusammenfassung

- Konfigurationen in parallelen und verteilten Dateisystemen sind notwendig zur Steigerung der I/O Performance.
- DCA-IO ist ein Verfahren, das Lustre Konfigurationen dynamisch anpasst
- DCA-IO verwendet Informationen aus bestehenden Logs eines HPC Systems um die I/O Performance zu verbessern

Zusammenfassung

- Konfigurationen in parallelen und verteilten Dateisystemen sind notwendig zur Steigerung der I/O Performance.
- DCA-IO ist ein Verfahren, das Lustre Konfigurationen dynamisch anpasst
- DCA-IO verwendet Informationen aus bestehenden Logs eines HPC Systems um die I/O Performance zu verbessern
- Konfigurationen müssen nicht vor der Ausführung einer Anwendung getestet werden
- Kein Wissen über die Anwendung nötig

Literatur

- [500] TOP 500. November 2022. [Internet](#).
- [Doc] NERSC Documentation. Cori. [Internet](#).
- [KSW⁺19] Sunggon Kim, Alex Sim, Kesheng Wu, Suren Byna, Teng Wang, Yongseok Son, and Hyeonsang Eom. Dca-io: A dynamic i/o control scheme for parallel and distributed file systems. In *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 351–360. IEEE, 2019.
- [ntv10] ntv. Urknall-simulation erfolgreich, 2010.
- [Rie] Benedikt Johannes Riehm. Das parallele dateisystem lustre.
- [SCH⁺16] Shane Snyder, Philip Carns, Kevin Harms, Robert Ross, Glenn K Lockwood, and Nicholas J Wright. Modular hpc i/o characterization with darshan. In *2016 5th workshop on extreme-scale programming tools (ESPT)*, pages 9–17. IEEE, 2016.
- [Wet] Deutscher Wetterdienst. Wetter- und klimalexikon.