

Anmerkung: Ein Teil dieses Übungsblattes ist bis 16.12. der Rest bis 23.12. abzugeben.

7.1 Fragen zur Vorlesung (30P)

7.1.1 Abstraktion der E/A

1. Mit welchen Funktionalitäten kann die Abstraktion der Ein-/Ausgabe verstärkt werden?
2. Welches Hauptproblem gibt es bei dieser erweiterten Abstraktion zu bewältigen?
3. Welche Charakteristika hat netCDF?
4. Wie kann man mit netCDF Daten lesen und schreiben?
5. Welche Charakteristika hat HDF5?
6. Was ist der Nutzen von HDF5 Daten-Chunks?

7.2 Evaluation der implementierten I/O-Varianten (570 Punkte)

In den vorangegangenen Übungen haben Sie verschiedene I/O-Operationen in das Jacobi-MPI-Programm eingebaut. Nun sollen die Resultate Ihrer Bestrebungen ausgewertet werden.

7.2.1 Wahl der Programmparameter (30 Punkte)

Um in angemessenem Zeitrahmen Messungen durchführen zu können, müssen wir zuerst die Programmparameter Matrixgröße und Iterationszahl so wählen, dass eine überschaubare Laufzeit zu erwarten ist.

Legen Sie sinnvolle Parameter für Ihr Programm so fest, dass ein einzelner Prozess ungefähr 400 Sekunden zur Bearbeitung in der langsamsten Umgebung (ohne I/O) benötigt. Wählen Sie nun sinnvolle Werte für die Iterationsanzahl des Checkpointings und der Visualisierung, so dass eine Verzögerung durch I/O von ungefähr 50% auftritt.

Um die Parameter festzulegen verwenden Sie die Ergebnisse Ihrer provisorischen Messungen aus den vorangegangenen Übungszetteln.

Geben Sie das Verhältnis der Anzahl an I/O Iterationen zu Gesamtiterationen an. Wozu kann dieser Wert dienen und wie ist er zu interpretieren?

7.2.2 Durchführung der Messungen (240 Punkte)

Als nächstes wollen wir aussagekräftige Messungen durchführen.

Dafür wählen wir die folgenden Knoten-Prozesse-Konfigurationen:

Messreihe I: (Knoten, Prozesse) = (x, x) für $x \in \{1, \dots, 9\}$

Messreihe II: (Knoten, Prozesse) = $(x, 2x)$ für $x \in \{1, \dots, 9\}$

Messreihe III: (Knoten, Prozesse) = $(2, 1), (4, 2), (6, 3), (8, 4)$

Messreihe IV: (Knoten, Prozesse) = $(2, x)$ für $x \in \{1, \dots, 6\}$

Dabei sollen die Prozesse immer so gut wie möglich auf die Knoten verteilt werden.

Führen sie nun Messungen mit den in der ersten Teilaufgabe ermittelten fixen Parametern für alle in den angegebenden Messreihen benötigten Konfigurationen für jeden der folgenden Fälle durch:

1. ohne I/O (mit dem POSIX-Programm)
2. mit aktiviertem Checkpointing
 - a) POSIX-I/O (`/dev/shm` (tmpfs))
 - b) POSIX-I/O (`/tmp`)
 - c) MPI-I/O auf PVFS2 (kollektiv)
 - d) MPI-I/O auf PVFS2 (unabhängig)
3. mit aktivierter Visualisierung
 - a) POSIX-I/O (`/dev/shm` (tmpfs))
 - b) POSIX-I/O (`/tmp`)
 - c) MPI-I/O auf PVFS2 (kollektiv)
 - d) MPI-I/O auf PVFS2 (unabhängig)

Um aussagekräftige Zahlen zu bekommen, müssen die Messungen mehrfach durchgeführt werden. Aus Zeitgründen entscheiden wir uns aber für ein einfaches Verfahren. Wir können das verantworten, da in der vom Ressourcen-Management kontrollierten Umgebung äußere Einflüsse minimiert sind und grobe Messfehler für unsere Zwecke ausreichend unwahrscheinlich sind.

Messen Sie jede Konfiguration mindestens zwei mal. Führen Sie Messungen, bei denen die Ergebnisse der beiden Läufe sich signifikant unterscheiden noch ein drittes mal durch.

Benutzung des Clusters mit Torque

Es gibt im Wiki einen neuen Abschnitt, der die Benutzung der Queue `ccio` beschreibt. Diese stellt automatisch ein PVFS2 zur Verfügung und soll für ALLE Messungen verwendet werden:

http://ludwig9.informatik.uni-heidelberg.de/wiki/index.php/Cluster:Benutzer-Quickstart#Nutzen_eines_automatisch_aufgesetzten_PVFS2

Hinweise

Abschätzungen zeigen das die Gesamtlaufzeit aller Messungen ca. 27 Stunden dauern wird. Bedenken Sie, dass Sie sich das Cluster mit allen anderen Gruppen teilen. Beachten sie darum bitte folgende Hinweise:

- Beginnen Sie frühzeitig.
- Planen Sie ihre Messungen so, dass sie möglichst früh mit den ersten Auswertungen beginnen können.
- Stellen Sie nicht alle Jobs gleichzeitig ein, so dass die anderen Gruppen auch zum Zuge kommen. (Das heißt, stellen Sie die Jobs ein, die Sie für ihren nächsten Auswertungsschritt benötigen.)

- Schätzen Sie die Laufzeit eines Job grob nach oben ab und setzen Sie die Walltime entsprechend. (Das verhindert, dass verklemmte Programme ewig die Ressourcen blockieren.)
- Testen Sie ihr Programm zuerst damit es möglichst fehlerfrei läuft.
- Sollten Sie Probleme auf dem Cluster feststellen oder beobachten, schreibe Sie eine Mail an die Liste.

7.2.3 Auswertung (240 Punkte)

Um die Messungen vergleichen und analysieren zu können, eignen sich Tabellen mit Zahlen ausgesprochen schlecht. Darum wollen wir die Ergebnisse in Speedup-Diagrammen visualisieren.

Erstellen Sie folgende Prozesse-Laufzeit-Diagramme:

Diagramm A: 1/I, 1/II, 1/III, 1/IV

Diagramm B: 1/I, 1/II, 2a/I, 2a/II, 3a/I, 3a/II

Diagramm C: 1/I, 1/II, 2b/I, 2b/II, 3b/I, 3b/II

Diagramm D: 2a/I, 2b/I, 2a/II, 2b/II, 2a/III, 2b/III, 2a/IV, 2b/IV

Diagramm E: 3a/I, 3b/I, 3a/II, 3b/II, 3a/III, 3b/III, 3a/IV, 3b/IV

Diagramm F: 1/I, 1/II, 2c/I, 2c/II, 3c/I, 3c/II

Diagramm G: 1/I, 1/II, 2d/I, 2d/II, 3d/I, 3d/II

Diagramm H: 2c/I, 2d/I, 2c/II, 2d/II, 2c/III, 2d/III, 2c/IV, 2d/IV

Diagramm I: 3c/I, 3d/I, 3c/II, 3d/II, 3c/III, 3d/III, 3c/IV, 3d/IV

Diagramm J: 1/II, 2b/II, 2c/II, 3b/II, 3c/II

Zeichnen sie neben dem Durchschnitt der Messungen auch die Werteschwankungen ein.

Ganz toll wäre es, wenn auch die genauen Durchschnittswerte im Diagramm direkt zu sehen sind, dies sollte aber nicht auf Kosten der Übersichtlichkeit gehen.

Auch die Größe der Diagramme sollte so gewählt werden, dass alles gut zu erkennen ist. Wir wollen hier eine selbstkritische Analyse durchführen und keine schlechten Werte in einer Veröffentlichung verschleiern. ;-)

Achten Sie auf eine sinnvolle Beschriftung der X- und Y-Achse.

Überlegen Sie sich den Sinn der einzelnen Diagramme und analysieren Sie sie entsprechend.

Versuchen Sie alle sichtbaren Effekte zu beschreiben und Erklärungen dafür zu finden.

Um zusätzlich ein besseres Bild von den Kosten der Visualisierung und des Checkpointings zu erhalten, erstellen Sie für diese beiden Zusatzfunktionalitäten jeweils ein Diagramm (Diagramm K und L) in dem der mittlere Laufzeitoverhead von MPI-IO (gegenüber der Version ohne I/O) in Sekunden für alle vier Messreihen dargestellt wird. Hierbei sollen sowohl kollektive als auch unabhängige Zugriffe in ein Diagramm eingetragen werden. Die Diagrammkonfiguration sieht also so aus:

Diagramm K*: 1/I, 1/II, 1/III, 1/IV, 2c/I, 2c/II, 2c/III, 2c/IV, 2d/I, 2d/II, 2d/III, 2d/IV

Diagramm L*: 1/I, 1/II, 1/III, 1/IV, 3c/I, 3c/II, 3c/III, 3c/IV, 3d/I, 3d/II, 3d/III, 3d/IV

Bedenken Sie das von den Laufzeiten der MPI-IO Varianten hierbei die Zeit des Laufes ohne I/O abziehen müssen.

Interpretieren Sie abschließend Ihre Ergebnisse in einem Fazit und fassen sie mögliche Ursachen und Lösungsideen für erkannte Probleme zusammen (ca. 1/2 Seite).

Abgabe (bis 16.12.2008)

Eine PDF-Datei mit

- den Diagrammen A bis E und dafür benötigten Messergebnissen
- der Analyse der Diagramme A bis E
- einem vorläufiges Fazit zum Thema POSIX-I/O

Abgabe (bis 23.12.2008)

Eine PDF-Datei mit

- allen Diagrammen (A bis L) und Messergebnissen
- der Analyse aller Diagramme
- dem abschließenden Fazit

7.2.4 Präsentation (60 Punkte)

Die Diagramme A bis D und ihre Analysen zu POSIX-I/O erwarten wir bis zur Übung am 16.12..

Für die restlichen Diagramme (E bis J) und Analysen haben sie eine Woche länger Zeit bis zum 23.12..

7.3 Rückmeldung

Gesamte Bearbeitungszeit			
Schwierigkeit	<input type="radio"/> zu leicht	<input type="radio"/> genau richtig	<input type="radio"/> zu schwer
Lehrreich	<input type="radio"/> wenig	<input type="radio"/> etwas	<input type="radio"/> sehr
Verständlichkeit	<input type="radio"/> großteils unklar	<input type="radio"/> teilweise unklar	<input type="radio"/> verständlich
Kommentar:			