

MapReduce für die Kopplung einer biomechanischen und einer systembiologischen Simulation

MapReduce to couple a bio-mechanical and a systems-biological simulation

Art der Arbeit: Bachelorarbeit
Bearbeiter: Alexander Gessler
Betreuer: Dipl.-Inf. Peter Reimann
Prüfer: PD Dr. rer. nat. habil. Holger Schwarz

Hintergrund:

„Big Data“ bezeichnet die massiv-parallele Verarbeitung sehr großer Datenmengen aus einer Vielzahl unterschiedlicher Datenquellen. Hierbei sind traditionelle Datenhaltungssysteme wie z.B. SQL-Datenbanksysteme häufig nicht mehr ausreichend. Stattdessen werden neuartige Konzepte wie MapReduce oder das PACT-Modell eingesetzt [1,2]. Auch bei computerbasierten Simulationen kann die Größe der zu verarbeitenden Daten zwischen mehreren GBs und zahlreichen TBs liegen, was für eine Nutzung dieser „Big Data“-Konzepte sprechen kann. Hauptgegenstand dieser Arbeit ist eine multi-skalare Simulation zu Strukturänderungen in Knochen, die zwei Simulationsmodelle aus den Anwendungsgebieten Biomechanik und Systembiologie koppelt [3]. Das biomechanische Modell wird dabei mit dem FEM-basierten Rahmenwerk Pandas¹ und das systembiologische Modell mit GNU Octave² numerisch implementiert, wobei beide Tools unterschiedliche Datenformate für ihre Ein- bzw. Ausgabedaten voraussetzen. Multi-skalar bedeutet zudem, dass die Tools Pandas und Octave mit unterschiedlichen Genauigkeiten der Raum- und Zeitdimensionen arbeiten, was die Komplexität der Datenintegration zwischen diesen Tools noch zusätzlich erhöht. Die gesamte Simulation wurde mit Hilfe der Workflow-Sprache Business Process Execution Language (BPEL, [4]) als Simulationsworkflows umgesetzt [5,6]. Zur Realisierung der Datenhaltung und Datenverarbeitung in den Workflows kommen bisher traditionelle Datenhaltungssysteme, z.B. ein SQL-Datenbanksystem, zum Einsatz.

Aufgabenstellung:

In dieser Arbeit soll untersucht werden, inwieweit in den betrachteten Simulationsworkflows Konzepte und Systeme aus dem Themenbereich „Big Data“ sinnvoll eingesetzt werden können. Dabei sollen Lösungsstrategien erarbeitet und diskutiert werden, wie sowohl die Datenintegration zwischen den Tools Pandas und Octave als auch die systembiologischen Simulationsberechnungen in Octave selbst mithilfe des MapReduce-Konzepts und auf Basis von Apache Hadoop³ oder verwandter Systeme [7] umgesetzt werden können. Erarbeitete Lösungsstrategien sollen prototypisch realisiert und in die Simulationsworkflows integriert werden. Anschließend soll der resultierende Prototyp evaluiert und dabei insbesondere mit der alten, rein workflow-basierten Umsetzung verglichen werden [5,6]. Die wichtigsten Evaluationskriterien hierbei sind (1) die Effizienz der Datenverarbeitung, (2) Flexibilitätsaspekte wie z.B. Möglichkeiten zur dynamischen Lastbalancierung und (3) die Komplexität der resultierenden Workflow-Modelle. Schließlich soll auf Basis dieser Evaluation noch diskutiert werden, ob in den betrachteten Simulationsworkflows andere „Big Data“-Konzepte eingesetzt werden können bzw. sollten, z.B. das PACT-Modell [2].

¹ <http://www.mechbau.uni-stuttgart.de/pandas/index.html>

² <http://www.gnu.org/software/octave/>

³ <http://hadoop.apache.org/>

Notwendige Abgaben/Leistungen:

- Vollständige Dokumentation des Lösungsansatzes und der Ergebnisse der Arbeit.
- Vollständige, gut dokumentierte und weiterverwendbare Implementierung des Prototyps.
- Vorstellung des Lösungsansatzes und der Ergebnisse der Arbeit in einem Vortrag im Rahmen des Abteilungskolloquiums.

Literatur:

- [1] Dean, J.; Ghemawat, S.: MapReduce: Simplified Data Processing on Large Clusters. In: Proc. of the 6th Symposium on Operating Systems Design and Implementation, San Francisco, California, 2004.
- [2] Alexandrov, A. et al.: Massively Parallel Data Analysis with PACTs on Nephelie. Proc. of the VLDB Endowment, Band 3, Nr. 2, 2010.
- [3] Krause, R. et al.: Bone Remodelling: A Combined Biomechanical and Systems-Biological Challenge. Applied Mathematics and Mechanics, 11(1), 2011.
- [4] Jordan, D.; Evdemon, J.: Web Services Business Process Execution Language Version 2.0, OASIS Standard. Organization for the Advancement of Structured Information Standards (OASIS), April 2007.
<http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.pdf>
- [5] Krause, R. et al.: Scientific Workflows for Bone Remodelling Simulations. Applied Mathematics and Mechanics, 13(1), 2013.
- [6] Reimann, P.; Schwarz, H.: Datenmanagementpatterns in Simulationsworkflows. In: Gesellschaft für Informatik (ed.): Datenbanksysteme für Business, Technologie und Web (BTW), Magdeburg, 2013.
- [7] Biesinger, S.; Pitterle, M.; Canko, M.: MapReduce – Eine vergleichende Analyse aktueller Implementierungen. Fachstudie Softwaretechnik Nr. 148, Universität Stuttgart, 2012.

Kontakt:

Peter Reimann
Institut für Parallele und Verteilte Systeme
Abteilung Anwendersoftware
Peter.Reimann@ipvs.uni-stuttgart.de
Tel.: 0711 685 88445