

Tutorium: Neue Hardwarearchitekturen für das Datenmanagement (DPMH)

Cagri Balkesen, Louis Woods, Jens Teubner

Über Jahrzehnte hinweg gelang es außerordentlich gut, aus der "Dividende" von Moore's Law immer schnellere Computersysteme zu bauen, die Anwendungssoftware quasi ganz automatisch beschleunigten. Die Grenzen dieses Ansatzes werden immer deutlicher und es ist zwischenzeitlich klar, dass signifikante Leistungssteigerungen in Zukunft nur noch durch einen hohen Grad an Hardware-Spezialisierung möglich sein werden. Für Fließkomma- oder Grafikoperationen hat sich dieser Trend bereits erfolgreich in der Praxis durchgesetzt.

Im Tutorium werden daher die Konsequenzen diskutiert, die neue Hardwaretechnologie auf datenintensive Anwendungen hat; allen voran diskutieren wir die Konsequenzen für Datenbankmanagementsysteme. Zu Beginn des Tutoriums werden wir Erneuerungen in der Computer- Systemarchitektur betrachten, die sich bereits in der Praxis durchgesetzt haben, von Konzepten wie tiefen Cache-Hierarchien über Parallelität und Multi-Core-Systeme bis hin zu sogenannten NUMA-Systemen. Dabei wird deutlich werden, wie wichtig ein geeignetes Zugriffsmuster (zum Beispiel auf den Hauptspeicher des Systems, aber auch zwischen parallelen Einheiten des Systems) für die Laufzeit von Algorithmen sein kann und es wird gezeigt, wie Datenbankalgorithmen konstruiert werden können, damit sie geeignete Zugriffsmuster aufweisen.

Wir werden dann Co-Prozessoren als echt spezialisierte Hardware diskutieren. Dabei werden wir im besonderen auf Grafikprozessoren (GPUs) und programmierbare Logikbausteine (FPGAs) eingehen, da diese sich im Kontext eines Datenbankeinsatzes als besonders nutzbringend erwiesen haben. Aufbauend auf der internen Architektur der jeweiligen Technologie werden wir zeigen, wie Co-Prozessoren programmiert werden und wie ihre Funktionalität in ein Gesamtsystem eingebunden werden kann

Ziel des Tutoriums wird es sein, die spezifischen Charakteristika der verschiedenen Technologien zu vermitteln. Dadurch werden die Teilnehmer nachher in der Lage sein, Aufwand und Nutzen des Einsatzes spezieller Hardware für einen gegebenen Anwendungskontext zu bewerten.

Über die Autoren

Cagri Balkesen ist wissenschaftlicher Mitarbeiter in der Systems Group an der ETH Zürich. Im Rahmen seiner Promotion beschäftigt er sich dort mit Datenbanktechnologien für den Einsatz im Hauptspeicher sowie mit Datenstromverarbeitung. Er hält einen MSc (Informatik) der ETH Zürich sowie einen BSc (Computer Engineering) der Middle East Technical University in der Türkei.

Louis Woods ist wissenschaftlicher Mitarbeiter in der Systems Group an der ETH Zürich. Sein Forschungsgebiet umfasst FPGAs im Kontext von Datenbanksystemen, parallele Algorithmen, Datenstromverarbeitung und Echtzeit-Mustererkennung. Er hält einen BSc und MSc der ETH Zürich in Informatik.

Jens Teubner ist Oberassistent und Leiter des Avalanche-Forschungsprojekts an der ETH Zürich. Das Projekt ist führend in der Verwendung von FPGA für den Datenbankeinsatz. Weitere Forschungsinteressen betreffen die Datenverarbeitung im Hauptspeicher sowie Hochgeschwindigkeits-Netzwerke. Jens Teubner promovierte 2006 an der TU München auf dem Gebiet der XML-Verarbeitung und arbeitete von 2007 bis 2008 im IBM-Forschungslabor in Hawthorne, New York.