

Many-Core-Architekturen zur Datenbankbeschleunigung – Tutorial –

Kai-Uwe Sattler¹, Jens Teubner², Felix Beier¹, Sebastian Breß²

¹ Technische Universität Ilmenau, Ilmenau, Germany
{first.last}@tu-ilmenau.de

² Technische Universität Dortmund, Dortmund, Germany
{first.last}@cs.tu-dortmund.de

Physikalische und technologische Grenzen bei der Erhöhung der Taktfrequenz von Prozessoren haben in den letzten Jahren die Entwicklung von Multi- und Many-Core-Architekturen forciert. Die Ausnutzung dieser Architekturen erfordert jedoch eine weitgehende Parallelisierung von Berechnungen. Für den Datenbankbereich bedeutet dies einerseits ein Überdenken etablierter Datenstrukturen und Verfahren, eröffnet aber gleichzeitig neue Möglichkeiten der Beschleunigung und Skalierung der Datenbankverarbeitung. Ziel des Tutoriums ist es daher, einen Überblick über den Stand der Forschung und die Einsatzmöglichkeiten von Many-Core-Architekturen in Datenbanksystemen zu geben. Neben Standard-Prozessoren stehen dabei insbesondere GPGPU-Architekturen im Mittelpunkt, die schon heute die Nutzung von Tausenden Cores ermöglichen. Ausgehend von einer Vorstellung aktueller Many-Core- und GPU-Architekturen sowie der darauf basierenden Programmiermodelle und APIs werden eine Reihe repräsentativer Einsatzfelder diskutiert. Neben Standard-Datenbankoperatoren, Indexierung von Daten sowie der Realisierung von benutzerdefinierten Funktionen (UDFs) werden auch Möglichkeiten der Unterstützung von DBMS-Komponenten wie dem Optimierer behandelt. Einen zweiten Schwerpunkt bilden Aspekte heterogener Plattformen aus CPU und GPU wie Fragen der Platzierung und des Datentransfers. Zielgruppe des Tutoriums sind Datenbank-Forscher und Entwickler, die sich für den Einsatz moderner Hardwareplattformen in Datenbanksystemen interessieren und sich über den Stand der Forschung informieren möchten. Sie erhalten einen Einblick in Programmiermodelle und -techniken sowie Entscheidungsunterstützung beim Einsatz von Many-Core- und GPU-Architekturen zur Datenbankbeschleunigung.

Kai-Uwe Sattler ist Professor für Datenbanken und Informationssysteme und leitet das gleichnamige Fachgebiet an der TU Ilmenau. Er promovierte 1998 an der Universität Magdeburg und habilitierte sich 2003 an der gleichen Universität. Zu seinen Forschungsgebieten gehören Anfrageverarbeitung mit Verteilungs- und Parallelisierungsaspekten, Datenstrommanagement sowie der Einsatz modernen Hardwarearchitekturen. Er ist Koautor mehrerer Lehrbücher zu Datenbankkonzepten sowie Algorithmen und Datenstrukturen.

Jens Teubner leitet seit April 2013 den Lehrstuhl für Datenbanken und Informationssysteme an der TU Dortmund. Zuvor forschte er als Postdoktorand bzw. Oberassistent am

IBM T.J. Watson Forschungslabor in den USA sowie in der Systems Group an der ETH Zürich. Er promovierte 2006 an der TU München bei Prof. Dr. Torsten Grust auf dem Gebiet der XML-Verarbeitung und diplomierte an der Fakultät für Physik der Universität Konstanz bei Prof. Dr. Ernst Bucher.

Forschungsschwerpunkt von Jens Teubner ist die Implementierung von Datenbanksystemen auf modernen Rechnerarchitekturen. Das von ihm gegründete Forschungsprojekt *Avalanche* an der ETH Zürich wurde schnell zur international führenden Gruppe was die Verwendung von Field-Programmable Gate Arrays (FPGAs) für die Beschleunigung von Datenbankaufgaben angeht. Er ist Mit-Entwickler der derzeit schnellsten Join-Implementation für moderne Hauptspeicher-Datenbanksysteme und zeigte, wie sich Netzwerkarten mit Hardwarebeschleunigung (sog. Remote Direct Memory Access, RDMA) besonders effizient für verteilte Datenbanken einsetzen lassen. Im Rahmen seiner Promotion beschäftigte Jens Teubner sich außerdem mit der hoch-skalierbaren Verarbeitung von XML, insbesondere unter Verwendung relationaler Datenbank-Back-Ends. Das Ergebnis seiner Promotion – der XQuery-Compiler *Pathfinder* – ist heute Teil des Open Source-Systems *MonetDB/XQuery*.

Jens Teubner erhielt für seine Dissertation *Pathfinder: XQuery Compilation Techniques for Relational Database Targets* im März 2007 den Dissertationspreis der Gesellschaft für Informatik, Fachgruppe Datenbanken (BTW-Dissertationspreis). Außerdem erhielt er einen Best Demo Award auf der BTW 2011, einen Best Paper Award auf der FCCM 2013 sowie ein Honorable Mention auf der ICDE 2013.

Felix Beier hat Informatik an der Technischen Universität Ilmenau studiert und promoviert dort seit 2010 am Fachgebiet Datenbanken und Informationssysteme. Sein Forschungsschwerpunkt ist die Parallelisierung von Indexoperationen auf hybriden CPU/GPU Systemen. 2013 wurde er für seine Forschungsarbeit mit dem IBM PhD Fellowship Award ausgezeichnet. Neben parallelen Algorithmen für moderne Prozessorarchitekturen gehören Datenstrommanagementsysteme und die Programmierung mit C++ zu seinen Interessen.

Sebastian Breß erhielt 2010 einen Bachelorabschluss und 2012 einen Masterabschluss in Informatik an der Universität Magdeburg. Aktuell arbeitet er an der Technischen Universität Dortmund und erforscht effiziente Techniken zur Anfrageoptimierung in heterogenen Prozessorsystemen. Dabei liegt der Fokus auf Ansätzen die keine Details der Prozessorarchitektur für die Optimierung benötigen. Während seiner Arbeit hat er zwei Systeme entwickelt: 1) Den lernenden Anfrageoptimierer *HyPE* und 2) *CoGaDB*, eine GPU-beschleunigte Hauptspeicherdatenbank für OLAP Szenarien. Er organisierte das „Deutsche Community Treffen für GPUs in Datenbanken“ 2014 in Dortmund und ist Mitorganisator des „International Workshop on Data (Co-)Processing on Heterogeneous Hardware“ (DAPHNE) bei der EDBT/ICDT 2015.