

Forschungsschwerpunkte – Prof. Dr. Thomas Neumann

Heute wird Fortschritt oft über Daten erzielt. Durch die Auswertung und Analyse großer Datenmengen können Erkenntnisse gewonnen werden, die rein analytisch schwer zu erreichen sind. Diese immense Bedeutung der Datenauswertung als neuer „Wirtschaftsmotor“ wird durch den Slogan „Data is the new oil“ betont. Dies bezieht sich nicht nur auf die wirtschaftliche Ausnutzung der Daten, sondern auch auf die Exploration von Daten in der Wissenschaft sowie für politische Entscheidungen. Gerade im wissenschaftlichen Bereich hat sich eScience, also die Erkenntnisgewinnung durch Datenauswertung, als viertes wissenschaftliches Paradigma neben den drei traditionellen Wissenschaftsmethoden der (1) Beobachtung und Messung, der (2) Modellbildung und der (3) Simulation etabliert.

Eine zentrale Herausforderung dabei ist die effiziente Durchführung komplexer Analysen. Die Datenmengen sind häufig so groß, dass nur durch eine geschickte Kombination effizienter Auswertungsalgorithmen, ermittelt durch kombinatorische Optimierung, sowie durch volle Ausnutzung aktueller Hardwarekapazitäten akzeptable Antwortzeiten erreicht werden können. Kurze Antwortzeiten sind gerade für eine explorative Datenanalyse enorm wichtig, da dabei Menschen auf das Ergebnis warten. Interaktive Antwortzeiten ermöglichen einen völlig anderen Umgang mit Daten als eine klassische Batchverarbeitung, stellen aber für das darunterliegende Datenverarbeitungssystem eine große Herausforderung dar. Die technische Entwicklung ist dabei sowohl ein Segen als auch eine Herausforderung: Einerseits stellen die großen Hauptspeicher und die verfügbare massive Rechenleistung die Grundlage für interaktive Auswertungen komplexer Analysen. Andererseits ist es sehr schwer, komplexe Auswertungsalgorithmen perfekt über Dutzende von Rechenkernen zu skalieren und dabei möglichst noch nebenläufige Änderungen zuzulassen. Wie man beispielsweise komplexe Anwendungslogik so in ein Datenbanksystem integriert, dass man sowohl Skalierbarkeit und Effizienz als auch die nötige Ausdruckstärke erreicht, ist eine ungelöste Herausforderung.

Wir gehen diese Herausforderungen an, indem wir komplexe Analysen möglichst gut auf die vorhandene Hardware abbilden und dabei alle verfügbaren Ressourcen nutzen. So bieten beispielsweise moderne CPU-Architekturen eine Vielzahl von optimierten Hardwareinstruktionen, die es ermöglichen, immer mehr Daten pro Prozessorzyklus zu verarbeiten. Die Verwendung dieser Spezial-Operationen ist wichtig, um kosten- und energieeffizient arbeiten zu können, die Operationen selbst sind aber nur winzige Bausteine, aus denen dann komplexe

Analysen zusammengesetzt werden müssen. Wir haben deshalb Techniken entwickelt, um die deklarative, hardwareunabhängige Anfragesprache SQL effizient und automatisiert in für die jeweiligen CPU-Architekturen optimierte Maschinencodes zu übersetzen. Das von unserer Gruppe entwickelte Datenbanksystem HyPer erreicht dabei für komplexe Anfragen Geschwindigkeiten, die an handgeschriebene Programme herankommen. Das System nutzt dabei alle verfügbaren Mittel, die auf der Hardware, auf der es ausgeführt wird, verfügbar sind, um hochperformante Datenanalyse auf einer großen Menge verschiedener Systeme zu ermöglichen – von einfachen handelsüblichen Single-Socket-Systemen mit einer einstelligen Anzahl an CPU-Kernen bis hin zu großen Serversystemen mit Hunderten von Prozessorkernen und Hauptspeichern, deren Größe sich im Terabyte-Bereich befindet.

Im Hinblick auf die Entwicklung von Prozessoren, die sich in den letzten Jahren weg von der Erhöhung der Taktrate immer stärker Richtung Vielkern-Systeme bewegt, ist die effiziente Auswertung komplexer Analysen unter nebenläufigen Änderungen unabdingbar. Auf einem Rechner mit 100 Rechenkernen sollte man Anfragen nicht einfach sequenziell hintereinander ausführen, man würde sonst 99 der 100 Rechenkerne ungenutzt lassen. Eine Ausführung auf allen 100 Rechenkernen gleichzeitig erfordert aber eine Synchronisation der Kerne, denn nebenläufige Änderungen dürfen laufende Analysen nicht beeinflussen. Vor allem die Transaktionssemantik relationaler Datenbanksysteme, die starke Konsistenz und Dauerhaftigkeit zusichert, erschwert eine lineare Skalierung der Performance mit der Anzahl der Prozessorkerne, da die Kerne sich untereinander absprechen müssen. Unser System unterstützt diesen gemischten Betrieb mit gleichzeitigem Lesen und Schreiben besonders gut, da wir Techniken entwickelt haben, um nebenläufige Operationen durch eine Kombination von Hardwareunterstützung und einer effizienten Realisierung von Mehrversionen-Serialisierbarkeit effizient voneinander abzuschirmen. Dadurch wird es möglich, gleichzeitig langwierige, analytische und kurze, transaktionale Anfragen effizient zu bearbeiten, womit die klassische Trennung zwischen Data-Warehouse-Systemen und Echtzeit-Transaktionssystemen aufgebrochen wird. Auch in diesem Bereich spielt die Ausnutzung moderner Hardware eine große Rolle: Transactional Memory ermöglicht eine hardwareunterstützte Isolierung von Transaktionen, wobei die Limitierungen der verfügbaren Hardware durch eine geschickte Koppelung an die Transaktionssemantik der Datenbank umgangen werden können.

Je mehr Nebenläufigkeit es bei der Anfrageverarbeitung gibt, desto wichtiger wird es, Arbeit fair und gleichmäßig auf alle verfügbaren Ressourcen zu verteilen. Wenn die Arbeit ungleich verteilt ist, wartet das System auf die langsamste Komponente. In unserer Forschung errei-

chen wir eine perfekt gleichmäßige Auslastung durch eine dynamische Parallelisierungsstrategie, die alle Anfragen in viele Arbeitspakete aufteilt, die dann je nach Auslastung des Systems auf beliebig viele Prozessorkerne aufgeteilt werden können. Ein weiterer Aspekt, der besonders bei Multi-Socket-Systemen beachtet werden muss, ist die Lokalität von Daten. So kann die Laufzeit einer Anfrage erheblich verringert werden, wenn Arbeitspakete auf den Prozessoren verarbeitet werden, die auch für den jeweils zugehörigen Hauptspeicher zuständig sind. Die gleichzeitige Berücksichtigung von Arbeitslast und Datenplatzierung kann die Anfragebearbeitung deutlich beschleunigen.

Neben diesen eher systemorientierten Bereichen ist die klassische Anfrageoptimierung nach wie vor ein wichtiges Forschungsgebiet. Dort gilt es, komplexe Optimierungsprobleme wie zum Beispiel das NP-schwere Problem der Bestimmung der optimalen Join-Reihenfolge für beliebige Anfragen zu lösen. Dazu entwickelte unsere Gruppe Algorithmen, die sich die zugrunde liegende Graphstruktur der Benutzeranfragen zunutze macht, um selbst für große Probleme beweisbar optimale Lösungen in einer kurzen Zeit zu finden. In seltenen Fällen sind die Probleme so groß, dass die Optimalität nicht garantiert werden kann, aber dann finden die Algorithmen immer noch sehr gute Lösungen. So können selbst für Anfragen mit Tausenden von Relationen immer noch Ausführungspläne gefunden werden, die nahezu optimal sind und somit alle in der Praxis auftretenden Anfragen beantworten können.

Die verfügbare Hardware entwickelt sich immer weiter, sodass unsere Gruppe immer wieder Neuerungen aufgreift, aktuell beispielsweise nicht flüchtige Hauptspeicher und sehr schnelle SSDs, um diese im Kontext von Datenbanksystemen zu nutzen. Darüber hinaus erweitert sich aber auch das Anwendungsfeld von klassischen Datenanalysen hin zu komplexen Anwendungen wie Clustering oder Maschine Learning. Wir entwickeln in unserer Gruppe deshalb unser Datenbanksystem hin zu einer „computational database“, bei der in einem integrierten System alles von klassischen SQL-Anfragen bis hin zu Maschine-Learning-Anfragen effizient und auf die jeweilige Hardware angepasst ausgewertet werden kann. Datenanalyse ist unerlässlich und die Bedeutung von Daten für den Erkenntnisgewinn wird in Zukunft noch weiterwachsen.