

Komplex- bzw. Forschungspraktikum

Implementierung ausgewählter Datenbankoperatoren auf moderner Hardware

Jede SQL-Anfrage in einem Datenbanksystem wird in einen Ausführungsplan bestehend aus einer Vielzahl von Operatoren (Join, Aggregation, Gruppierung etc.) überführt. Die Effizienz der Anfrageverarbeitung hängt damit von der Effizienz der einzelnen Operatoren ab. Im Rahmen dieses Komplex- bzw. Forschungspraktikums wollen wir uns daher mit der Effizienz der einzelnen Operatoren im Detail beschäftigen. Dazu werden wir uns mit der Vektorisierung ausgewählter Operatoren wie beispielsweise Join oder Gruppierung mittels SIMD („Single Instruction Multiple Data“) beschäftigen.

Viele moderne Prozessorarchitekturen beinhalten inzwischen SIMD-Erweiterungen, das heißt spezielle zusätzliche Befehlsätze (Instruktionen), die mit einem Befehlsaufruf gleichzeitig mehrere gleichartige Werte verarbeiten. Der Funktionsumfang gängiger SIMD-Erweiterungen ermöglichte bisher oft nur eine naive Vektorisierung. Dabei wird eine skalare Implementierung so verändert, dass sie dieselben Instruktionen auf mehrere Werte gleichzeitig ausführt. Die Instruktionen selbst bleiben dabei dieselben.

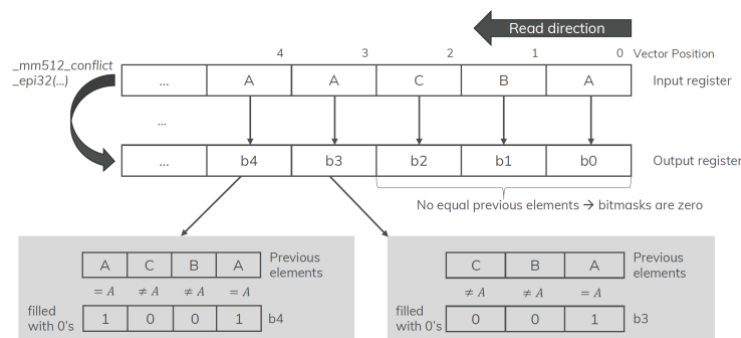
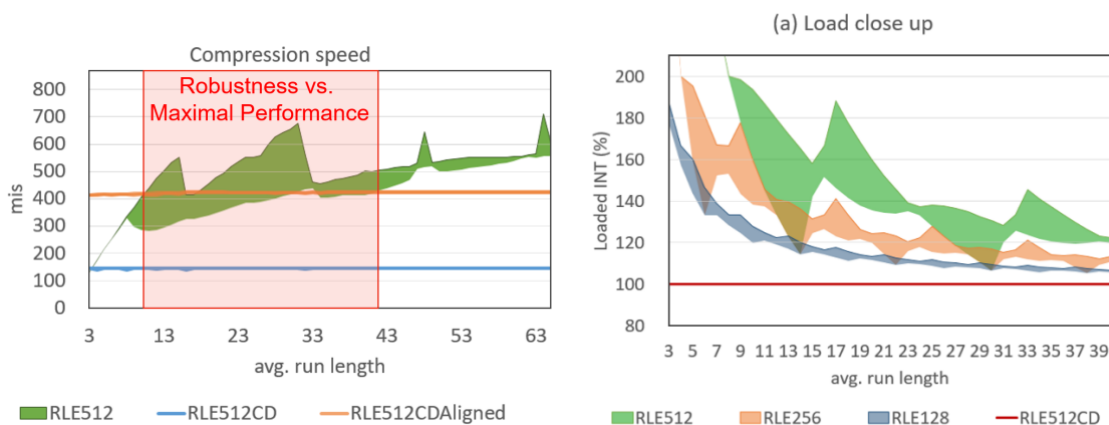


Abbildung 1: Erstellung konfliktfreier Untermengen

Mit der Intel-Erweiterung AVX-512 gibt es nun neue Instruktionen, die auch Abhängigkeiten zwischen den Werten in einem Vektor betrachten, z. B. das Erstellen konfliktfreier Untermengen wie in der Abbildung 1 dargestellt ist. Das eröffnet neue Möglichkeiten für die Implementierung von Datenbankoperatoren. In diesem Praktikum soll an einigen Beispieloperatoren untersucht werden, wie die neuen Instruktionen sinnvoll genutzt werden können. Dies beinhaltet auch die Implementierung, Messung und Auswertung der überarbeiteten Operatoren.



Ablauf:

Am Anfang des Praktikums wird eine Einführung in die SIMD-Programmierung und insbesondere in AVX-512 Erweiterung gegeben, damit alle Teilnehmer auf gleichem Niveau sind. Danach wird eine Auswahl von Operatoren präsentiert und wie deren klassische Vektorisierung funktioniert. Auf Basis dessen, sollen dann neue Implementierungskonzepte für die Operatoren unter Ausnutzung der AVX-512 Funktionalitäten konzipiert und schlussendlich implementiert werden. Als Testplattform steht ein aktueller Intel Xeon Phi Knights Landing Prozessor zur Verfügung, der über die neue AVX-512 Erweiterung verfügt.

Voraussetzungen:

- Erfahrung mit C/C++
- Datenbanken Grundlagen

Bei Interesse bitte bei Annett Ungethüm (annett.ungethuem@tu-dresden.de) melden.