

IBM DB2 10.5

Multi Workload Database for Big Data & Business Analytics

■ **IBM DB2 10.5 with BLU Acceleration - Analytics at the speed of thought!**

Einen markigen Text hat sich die IBM für das neue DB2 Release ausgedacht. Wir haben uns gefragt welche Aussage die IBM damit verbindet und vor allem, was sich hinter dem Kürzel BLU verbirgt. Wissen Sie es?

BLINK Ultimate ist die Lösung, ein Forschungsprojekt der IBM um Data Warehouse Abfragen auf die Geschwindigkeit eines Augenzwinkerns zu bringen.

Damit wäre anstelle von Kepler „**Bezaubernde Jeannie**“ der treffendere Arbeitsname für das neue Release gewesen. Die Botschaft frei übersetzt: „**Ein ultimatives Zwinkern - und analytische Auswertungen rauschen mit der Geschwindigkeit der Gedanken dahin ...**“, ... wenn das nicht zauberhaft ist.

Aber zurück auf den Boden der Tatsachen, dass es sich bei BLU Acceleration um grundsolide Technik handelt und nicht um Zauberei, erfahren Sie in diesem Artikel.



DB2 10.5 und BLINK Ultimate

Für Data Warehouse-, Business Analytics- und Big Data-Anwendungen, stellt die neue DB2 Version mit BLU Acceleration einen echten Meilenstein in der DB2 Historie dar. Mit DB2 BLU Acceleration kombiniert die IBM zeilenorientierte OLTP Datenhaltung mit spaltenorientierter und im Memory angelegter OLAP Datenhaltung in einer Version.

Die Grundlagen hierfür wurden in den letzten Jahren federführend vom IBM Almaden Research Center (San Jose, CA) im Rahmen des Projektes BLINK geschaffen. Die Forschungsgruppe um Guy M. Lohman beschäftigte sich hauptsächlich mit der Frage, wie die typischerweise lang laufenden Business Analytics Abfragen, durch neue Algorithmen und Techniken ein Antwortniveau von Sekunden erreichen können. Die Ergebnisse dieser Forschungen sind nun im neuen DB2 Release umgesetzt und bringen zweistellige Performanzverbesserungen und noch einmal eine deutlich erhöhte Datenkompression. Ansätze der BLINK Technologie kommen übrigens auch beim IBM Informix Warehouse Accelerator zum Einsatz.

Hintergrund dieser Bemühungen ist nicht nur die allgemeine Verbesserung der Performanz von Warehouse Anwendungen. Primäres Ziel der IBM dürfte es sein, die Marktführerschaft im Bereich Big Data Analytics zu besetzen. DB2 10.5 BLU Acceleration liefert die notwendige Technologie für die Speicherung und Auswertung großer Datenmengen und ist damit ein strategischer Baustein im Big Data Lösungsportfolio der IBM.

Bis 40-fach schnellere OLAP Auswertungen

25 bis 40 fache Beschleunigung gibt IBM offiziell als mögliche Performanzverbesserung an. Dies sind ganz erhebliche Werte, zur Verdeutlichung: ein Abfrage von 12 Minuten - mit dem Faktor 25 gerechnet - ist in weniger als 30 Sekunden erstellt. Wie werden diese Werte erreicht?

In-memory-technology

In-memory-technologie ist schnell, aber teuer und kommt für Warehouse Anwendungen bisher kaum zum Einsatz. Die sinkenden Preise für Hauptspeicher machen diesen Ansatz für OLAP Anwendungen zunehmend zu einer Alternative.

Sind die Daten erst einmal geladen, ist in-memory Verarbeitung deutlich schneller. Dabei wird der Vorgang der Datenübernahme für eine Umsetzung in die, für OLAP Auswertungen optimierte, spaltenbezogene Organisation der Daten genutzt.

Die spaltenbezogenen Organisation bietet zusätzlich eine Vereinfachung für die Formulierung der Warehouse Auswertungen. Eine besondere Anpassung und Optimierung der SQL-Abfragen für die Analyse wird überflüssig, die Administration entlastet.

Begrenzende Faktor für den Einsatz von in-memory Technologien bleiben aber die Speicherkosten. Der Schlüssel, um die Kosten und die Geschwindigkeit der Verarbeitung in den Griff zu bekommen, liegt in der Selektion und der Kompression der Daten.



Selektion, der Hybrid Ansatz von DB2 BLU Acceleration

DB2 BLU Acceleration lädt nicht die gesamte Datenbank sondern nur die, für die Warehouse Auswertung relevanten Tabellen in den Hauptspeicher. Dies ist ein wesentlicher Aspekt um Hauptspeicher einzusparen. Der Hybrid Ansatz von DB2 BLU Acceleration vereint damit zeilen- und spaltenorientierte Organisation in einer Datenbank und bietet die besten Voraussetzungen OLTP und OLAP Anwendungen gemeinsam bedienen zu können.

Compression

DB2 BLU Acceleration setzt auf massive Datenkompression mit neuen Algorithmen und in mehreren Stufen. Die Ausgangsdaten werden um ca. 90% komprimiert und in dieser Form für den Aufbau der spaltenorientierten Verarbeitung vorgehalten. Das Laden in den Hauptspeicher und der Aufbau der Spalten beinhaltet eine weitere Kompression, erneut um 80% - 90%. Die Kompressionsalgorithmen sind direkt im DB2 Kernel implementiert und machen diese Verarbeitungsphase extrem schnell. Dabei bleibt die Sortierreihenfolge unverändert und Vergleichsoperationen auf den komprimierten Daten liefern die gleichen Ergebnisse, wie in den umkomprimierten Vorstufen.

Data Skipping

Weitere Reduzierung des Volumens erfolgt durch „Data Skipping“. Dieser Schritt filtert quasi alle nicht relevanten Daten aus, sodass zur Verarbeitung nur die relevanten Daten angelegt werden. Auch hier sind weitere deutliche Reduzierungen möglich.

Smart-Caching und SIMD

Die Daten werden hardwarenah aufbereitet und an die vorhandenen CPU-Cores verteilt angelegt. Dabei nutzt DB2 BLU Acceleration die Prozesseigenschaften zur Parallelverarbeitung. Wenn es die Daten zulassen, kommt die SIMD Technologie (SIMD single instruction, multiple data) zum Einsatz. Dabei wird die gleiche CPU Operation auf mehrere Datenströme gleichzeitig angewendet. Durch die Optimierung von BLU für IBM Power Systeme, die übrigens von einem deutschen Forschungs- & Entwicklungsteam der IBM durchgeführt wurde, ist es gelungen die Geschwindigkeit wichtiger Datenbankoperationen, dem so genannten Hashing und Scanning, zusätzlich um den Faktor 2 bis 3 zu beschleunigen.

Verfügbarkeit

DB2 BLU Acceleration steht nur für die „Warehouse Editionen“, d.h. die Advanced Enterprise und Advanced Workgroup Edition, bzw. zusätzlich für die Developer Edition, zur Verfügung. Portierungen erfolgen für die gängigen UNIX, Linux und Windows Plattformen und aktuelle Prozessorbaureihen, Details sind den Announcements der IBM zu entnehmen.

Zusammenfassung - *Analytics at the speed of thought*

DB2 BLU Acceleration ist eine

- in-memory hybrid Technologie,
- mit starker Datenkompression und
- selektiver Organisation der Daten in Spalten.
- Hardwarenah aufbereitete Datenhaltung mit Optimierung für Parallelverarbeitung.

DB2 BLU Acceleration bietet

- in der Summe schnellere Analysen für Data Warehouse Anwendungen,
- mit einfacheren Abfragen und niedrigeren administrative Aufwendungen,
- ohne die hohen Kosten oder Begrenzungen von reinen in-memory Systemen.

Weitere Informationen auf den Seiten der CURSOR Distribution www.cursor-distribution.de:

IBM Pressemitteilung vom 15.04.2013

[IBM stellt neue DB2 Version mit BLU-Technologie und IBM PureData System for Hadoop vor](#)

IBM Announcement vom 23.04.2013

[IBM DB2 10.5 with BLU Acceleration the multi-workload database software for the era of big data](#)

IBM Data Magazine

[DB2 10.5 BLU Excelleration
Super Analytics - Super Easy](#)

Clabby Analytics

[IBM's New Move in Big Data: Acceleration, Optimization, and an Open Source Alternative](#)

BIRTE 2011

[Background to DB2 10.5 BLU Excelleration
Blink: Not Your Father's Database!](#)



■ **Über die IBM Distribution der CURSOR Software AG**

Die CURSOR Software AG ist Premier Business Partner der IBM und Großhändler für ausgewählte IBM-Software.

In unserer Eigenschaft als **IBM-ASL-Master-Distributor** vertreiben wir applikationsspezifisch lizenzierte IBM Software mit den Schwerpunkten Information Management und Business Intelligence.

Auf Datenbanken spezialisiert, bieten wir leistungsstarke Produkte zusammen mit umfassender Beratung und einem erstklassigen technischen Support an.

Weitere Informationen unter www.cursor-distribution.de

Ihr Ansprechpartner für diese Vertriebsinformation:



Dipl. Inf. (FH) Jürgen Storch
(Geschäftsbereich IBM Distribution)

CURSOR Software AG
Friedrich-List-Straße 31
D-35398 Gießen
juergen.storch@cursor.de

Bildmaterial Fotolia
Copyright (© Viktor Kuryan – © Stauke / Fotolia.com © 2013)

