

IBM DB2 für i eröffnet neue Möglichkeiten zur Steigerung der BI-Performance

von Mike Cain

Darüber hinaus gibt er Hinweise für Tests, Überwachung und die Übergabe der Lösung in die Produktionsumgebung. Teil 1 des Beitrages ist in NEWSolutions, Ausgabe April 2009 erschienen.

Ein programmatischer Ansatz zum Datenbank-Parallelismus

Bei der Erstellung von Indizes und Queries kann sich das Feature DB2 für SMP als hilfreich erweisen. Für andere Arten von Workload, wie beispielsweise Non-SQL-Datenverarbeitung kann der Parallelismus über einen programmatischen Ansatz erreicht werden. So lässt sich z. B. während des ETL-Prozesses (Extraction/Transformation/Loading) ein höherer Durchsatz erreichen, indem man das ETL-Tool oder -Programm mehrere parallel ablaufende Verarbeitungs-Streams mit multiplen Datenbankverbindungen ausführen lässt. Dabei sollte aber bedacht werden, dass SQL Insert-, Update- und Delete-Operationen nicht parallel ausgeführt werden können. Wenn also mit einer großen Anzahl von Zeilen gearbeitet werden muss, erweist es sich als günstig, die Steuerung selbst zu übernehmen und programmtechnisch bestimmte Datenbereiche zu identifizieren, die dann in mehreren parallel ablaufenden Operationen abgearbeitet werden, jede für einen definierten Datenbereich.

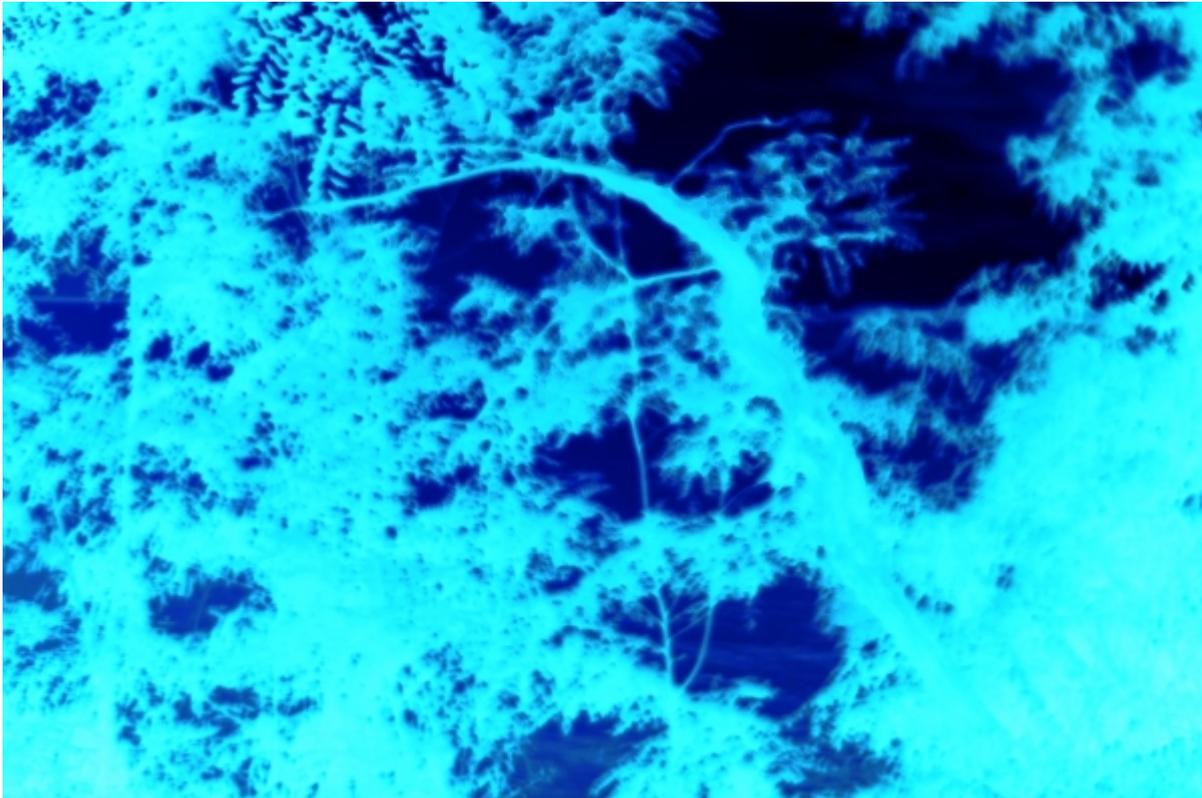
Der Autor

Mike Cain (mcaïn(ät)us.ibm.com) ist leitendes Mitglied des technischen IBM Stabes. Er befasst sich überwiegend mit der Untersuchung, Quantifizierung und Darstellung der i5/OS Unterstützung für sehr große Datenbanken, BI und SQL Query Performance. Übersetzt und für den deutschsprachigen Markt überarbeitet von Joachim Riener.

Isolation und Locking

Commitment Control und Isolations-Level bieten die Möglichkeit festzustellen, wann eine Transaktion abgeschlossen ist und wann die Zeilen, die in Zusammenhang mit einer bestimmten Transaktion stehen, wieder für andere Benutzer verfügbar sind. DB2 für i verwendet ein Locking auf Zeilenebene. Anders ausgedrückt werden Zeilen abhängig von dem im Job oder einer Datenbankverbindung und der SQL-Operation spezifizierten Isolations-Level gesperrt und sind für andere Benutzer nicht verfügbar. Mit dem Ansteigen gesperrter Zeilen bis in Millionenhöhe muss mehr Zeit und Mühe darauf verwandt werden, diese Locks zu überwachen und zu verwalten. An

dieser Stelle muss wohl nicht besonders darauf hingewiesen werden, dass die gesperrten Zeilen für die Dauer der Transaktion für andere Benutzer nicht verfügbar sind. Die entsprechenden Isolation-Level sollten vollständig verstanden sein und entsprechend der Situation eingesetzt werden. Es ist grundsätzlich empfehlenswert, den niedrigsten Isolations-Level zu verwenden, der ein angemessenes Anwendungsverhalten zulässt. Wird Journaling und Commitment Control eingesetzt, um die Wiederherstellbarkeit zu gewährleisten, sollten die Journal- und Journal Receiver-Objekte für bestmögliche Performance konfiguriert werden. Commit-Operationen sollten in angemessenen Intervallen erteilt werden, um die Anzahl der Locks für zusammengehörige I/O Operationen in sinnvollen Grenzen zu halten.



Query und Reporting Tools

SQL ist die strategische und am weitesten entwickelte Schnittstelle für Datendefinition und Datenmanipulation. Im Laufe der letzten fünf DB2 für i Releases haben SQL selbst, der Query Optimizer und die Database-Engine signifikante Verbesserungen erfahren. Die jüngsten Verbesserungen in Funktion, Performance und Tools stehen allerdings nur bei Nutzung der SQL Query Engine (SQE) zur Verfügung. DB2 für i verfügt über vielfältige Schnittstellen (z. B. SQL, OPNQRYF, Query/400, query API), aber nur SQL Requests sind in der Lage, SQE zu nutzen und die Vorteile aus den Verbesserungen zu ziehen. Es sollte sichergestellt werden, dass datenzentrierte Anwendungen, Query und Berichts-Tools beim Zugriff auf DB2 für i nur SQL verwenden. Wenn Sie SQE nicht nutzen, verzichten Sie darauf, DB2 für i die größtmögliche Leistung abzuverlangen und eine optimale BI-Performance zu erzielen. Ausgewogene Konfiguration - optimaler Set von Ressourcen Beim Zusammenstellen der Konfiguration eines Systems für Business Intelligence ist unbedingt darauf zu achten, dass ein angemessenes Volumen von Computing- und Speicher-Ressourcen eingeplant wird. Diese Ressourcen müssen über ein ausgewogenes Verhältnis zwischen CPU, Speicher und I/O Subsystemen verfügen. Die schnellste Systemkomponente ist die CPU, aber alle CPUs warten halt mit der gleichen Geschwindigkeit. Deshalb sollte für eine gute Performance stets ein angemessen ausgestattetes und konfiguriertes System bereitgestellt werden. Bezogen auf die abzusehende Query-Nutzung und das erwartete Antwortzeitverhalten sollte eine ausgewogene Konfiguration aus dedizierten POWER6 Prozessoren, 6 bis 12 GB Speicher pro CPU und 10 bis 20 Platteneinheiten pro CPU bestehen. Dies sind Minimalwerte basierend auf allgemeinen Beobachtungen und Erfahrungen. Natürlich können Ihre Anforderungen und Konfigurationen von diesen Angaben abweichen. Intensive Analysen, Planungen und Tests sind die letztlich ausschlaggebenden Erfolgsfaktoren. Vermeiden Sie auf jeden Fall unausgewogene Konfigurationen, bei denen zwar CPU-Leistung im Überfluss vorhanden ist, aber zu wenig I/O Subsysteme zur Verfügung stehen, um diese Leistung abzurufen.

Wird für Datenbank-Parallelismus DB2 SMP eingesetzt, müssen mehrere CPUs für die Ausführung mehrerer gleichzeitig ablaufender Threads oder Tasks vorhanden sein. Darüber hinaus ist es grundsätzlich sinnvoll, für eine BI-Umgebung ein dediziertes System oder eine logische Partition

(LPAR) bereitzustellen. Die gleichzeitige Abwicklung des Online-Geschäfts (OLTP - OnLine Transaction Processing) und der BI-Workload auf demselben System ist grundsätzlich als zumindest problematisch anzusehen. Ein effektives Tuning kann sich unter diesen Bedingungen äußerst schwierig gestalten. Ein separates BI-System oder eine logische Partition bieten eine weitaus bessere Nutzung der Ressourcen und wird den unterschiedlichen und teilweise weit auseinanderlaufenden Anforderungen von OLTP und Query/Reporting eher gerecht.

Ausgehend von der Tatsache, dass die IBM POWER Server und die dazugehörigen Betriebssysteme über die besten derzeit im Markt erhältlichen Virtualisierungsfunktionen verfügen, sollte man sich für eine Data Warehousing- oder Reporting-Infrastruktur der logischen Partitionierung bedienen. Wenn eine Data Warehouse LPAR sich in direkter Nähe der Transaktions-LPAR(s) befindet, lassen sich zur Unterstützung des Datentransports private Hochgeschwindigkeits-LANs einsetzen. Das System kann abhängig von den unterschiedlichen Anforderungen im Laufe eines Tages den einzelnen Aufgabengebieten dynamisch CPU- und Speicherressourcen zuweisen oder entziehen. Sollte sich überdies die Notwendigkeit ergeben, BI-Lösungen oder Middleware unter AIX oder Linux ablaufen zu lassen, lässt sich eine entsprechende LPAR in unmittelbarer Nähe der IBM i oder i5/OS LPARs einrichten.

Work Management

This content is available for purchase. Please select from available options.

- [7 Euro/Monat NEWSabo digital - sofort zugreifen.](#)
- [13,5 Euro/Monat NEWSabo plus inklusive 5x Login & Print-Ausgabe - sofort zugreifen.](#)

[Login & Purchase](#)