

Performance-Schub für Siebel durch Künstliche Intelligenz

„KI statt KIWI“

 dimensio
informatics

Die Mächtigkeit von Oracle Siebel („the world’s most complete CRM solution“) kann zugleich Quelle vielfältiger Performance-Probleme sein. Die Flexibilität der Applikation fusst auf einem hochdynamischen, generischen Datenmodell, das Siebel-Anwendern erlaubt, eigene Geschäftsobjekte zu definieren, was in puncto Anfrageoptimierung nicht unproblematisch ist. Aber auch ohne individuelle Erweiterungen sind Siebel-Abfragen häufig sehr komplex und überfordern den Optimierer des Datenbankmanagementsystems. Antwortzeiten im Minuten- oder Stundenbereich hinterlassen dann enttäuschte Anwender. Der Versuch, den Performance-Problemen mit klassischen Tuningmassnahmen zu begegnen, fruchtet oft nicht, und auch schnellere Hardware (Stichwort: „kill it with iron“) kann die Lage nur bedingt verbessern.

Der Ausweg heißt „Künstliche Intelligenz“. Mit dem mehrdimensionalen semantischen DIMENSIO-Index bieten wir eine intelligente Lösung, die auf selbstlernenden neuronalen Netzen basiert. Damit lassen sich langlaufende Anfragen oft um das Tausendfache beschleunigen, ohne die Anwendung zu modifizieren. Die Software schaltet sich – von Siebel unbemerkt – dazwischen und reichert komplexe SQL-Anfragen mit Zusatzinformation an, die den DBMS-Optimierer bessere Ausführungspläne generieren lässt. Dass sich damit Siebel-Anfragen von Stunden auf Sekunden reduzieren lassen, zeigt das Anwendungsbeispiel einer Schweizer Versicherungsgesellschaft (siehe Rückseite: „Die Mobiliar“).

Lange Antwortzeiten – lange Gesichter

Die meisten Siebel-Anfragen resultieren in Joins relativ vieler Tabellen. Die Crux dabei: Die Restriktionen auf die einzelnen Tabellen sind nicht besonders selektiv, sodass das DBMS die Joins nacheinander „materialisieren“ muss und erst das entstandene Zwischenergebnis auf die relevante Ergebnismenge reduzieren kann. Das resultiert in Leseoperationen im Gigabyte-Bereich auf den Platten bzw. im Buffer-Cache und großen Datenmengen, die bewegt werden müssen, um am Ende nur wenige Treffer zurückzugeben. Die Erfahrung lehrt: Wenn die Durchlaufzeiten für Auswertungen wie eine „Segmentierung“ im Stundenbereich, oder die Antwortzeiten für vermeintlich einfache Abfragen im Minutenbereich liegen, dann ist Anwenderfrustration vorprogrammiert. Im schlimmsten Fall führt das dazu, dass die Anwender einzelne Funktionen nicht mehr nutzen oder gar das komplette System boykottieren.

Klassische Tuning-Massnahmen greifen oft nicht

Da die SQL-Anfragen aufgrund des generischen Datenmodells von Siebel zur Laufzeit dynamisch generiert werden, bleiben dem DBA

nur die klassischen Tuning-Massnahmen: Anlegen von Indizes und Bereitstellung von Statistikdaten. Indizes haben aber nicht nur Vorteile. Sie kosten viel Speicherplatz und der Overhead bei Datenänderungen steigt deutlich. Zudem müssten sie mit unterschiedlichsten Anordnungen der indizierten Spalten angelegt werden, um beliebige Anfrage-Szenarien zu unterstützen. Auch Statistikdaten sind hinsichtlich der Performance nicht unkritisch. Wenn sie nicht mehr aktuell sind, können gute Ausführungspläne „kippen“, und die Antwortzeiten verlängern sich ohne erkennbaren Grund.

„KIWI“ ist auch keine Option

Sind die klassischen Tuning-Massnahmen an ihre Grenzen gestoßen, bleibt nur noch der Versuch, die Performance-Probleme durch mehr „Blech“ zu lösen. Weitere CPUs würden aber nur dann helfen, wenn die Prozessoren (bzw. Cores) der begrenzende Faktor sind und das DBMS die Anfragen parallelisiert ausführen könnte. Hinzu kämen weitere Oracle-Datenbank-Lizenzen für zusätzliche Cores. In der Regel sind es die Leseoperationen, die die meisten Ressourcen beanspruchen. Ihnen könnten mehr Hauptspeicher oder schnellere Platten

helfen. Mit mehr RAM lässt sich zwar der Datenbank-Cache vergrößern, das nützt aber wenig, wenn das DBMS ganze Tabellen scannen oder Joins erst materialisieren muss. Und auch Flash-Speicher-basierende SSDs helfen nur bedingt, weil sie zwar die Plattenzugriffe beschleunigen, nicht aber das grundlegende Problem zu vieler Zugriffe lösen.

Die Lösung: Mehrdimensionale semantische Indizierung mit KI

Massnahmen zur Verbesserung der Performance sind nur dann zielführend, wenn sie das Problem angehen, dass komplexe Anfragen Einschränkungen auf viele Attribute in Dutzenden Tabellen vornehmen. Traditionelle Datenbankindizes sind eindimensional und können nur Spalten genau einer Tabelle abbilden. Sie helfen bei komplexen Anfragen dann nicht, wenn jeder für sich die Treffermenge aus einer einzelnen Tabelle nur in geringem Umfang einschränken kann, und erst die Kombination mehrerer Attribute ausreichend selektiv wirkt.

Die Zauberformel heisst „mehrdimensionale Indizierung“ beliebiger Spalten (ohne Berücksichtigung ihrer Anordnung) in beliebig vielen Tabellen. Klassische mehrdimensionale Indizes wurden ursprünglich für geocodierte Daten in zweidimensionalen Räumen entwickelt. Sie gliedern den Datenraum aber nach rein formalen Kriterien, d. h. ohne Wissen über die Datensemantik, und sind auf wenige Dimensionen begrenzt. Die formale Beschränkung kann durch „Wissen“ über die semantischen Datenbeziehungen überwunden werden, das mit Verfahren der „Künstlichen Intelligenz“ erlangt wird.

Der zum Patent angemeldete, mehrdimensionale semantische DIMENSIO-Index basiert auf einem selbstlernenden Verfahren, das auf neuronalen Netzen aufbaut. In der Lernphase wird zunächst das Wissen für die Konstruktion der Indexstruktur erworben. Hierzu werden die SQL-Anfragen analysiert und dabei die relevanten Attribute ausgefiltert, welche die Ergebnismenge einschränken. Im nächsten Schritt wird der Index aufgebaut, der ggf. Tausende von Dimensionen abbilden kann. Dazu werden die Anfragen gegen die Datenbank ausgeführt und die Primärschlüssel der betroffenen Tabellen extrahiert. Die gefundenen Attribut- und Schlüsselwerte werden dann an das selbstlernende Verfahren übergeben. Dieses maschinelle Lernverfahren clustert Daten entsprechend ihrer Ähnlichkeit in einer „semantischen Landkarte“, und speichert sie im DIMENSIO-Index. Dieser wird daraufhin – völlig transparent für die Anwendung – zur Optimierung der Datenbankabfragen verwendet.

Der Trick: Der Index kennt außer den Attributen, die für Selektionen verwendet werden, auch die Primärschlüsselwerte der relevanten Tabellen in der Datenbank. Jede Siebel-Anfrage wird nun analysiert und zunächst gegen den DIMENSIO-Index ausgeführt. Das Ergebnis ist eine Treffermenge in Form von Primärschlüsselwerten. Die ursprüngliche SQL-Anfrage wird dann durch diese Werteliste erweitert, wodurch der DBMS-Optimierer eine extrem selektive Anfrage erhält. Er wird daher einen Ausführungsplan wählen, der die Primärindizes verwendet und folglich mit wenigen Cache- oder Plattenzugriffen auskommt. Im Ergebnis verkürzen sich die Antwortzeiten von Siebel teilweise drastisch. Extrembeispiel ist eine „Kundensegmentierung“ mit 9 Ergebniszeilen, die von 3,5 Stunden auf 2 Sekunden reduziert werden konnte.

Das KI-basierte Verfahren passt sich an variierende Datenbestände an. Neues Wissen über hinzugefügte Datensätze oder geänderte Werte ergänzt den vorhandenen Index, ohne dass die einmal analysierten Datensätze noch einmal betrachtet werden müssten.

Minimal-invasive Arbeitsweise

DIMENSIO arbeitet transparent und „minimal-invasiv“. Die Applikationssoftware bleibt unangetastet und auf Datenbankseite sind lediglich minimale Eingriffe für den Synchronisationsservice erforderlich. In Siebel-Umgebungen wird der Applikationsserver als Integrationspunkt gewählt, wobei die Indizierungssoftware auf einem eigenen Server läuft. Zur Laufzeit schaltet sich die DIMENSIO-Software zwischen die Kommunikation von Applikationsserver und DBMS. Sie analysiert die SQL-Anfragen und bewertet jeweils den möglichen Nutzen, den der DIMENSIO-Index erzielen kann. Bei entsprechend großem Potenzial wird die Anfrage mit Daten aus dem Index angereichert. Kann die Anfrage keinen Nutzen aus dem Index erzielen, wird sie unverändert weitergeleitet. Der Overhead für die Analyse liegt im Nanosekundenbereich und ist praktisch nicht messbar.

Fazit

Die DIMENSIO-Software lässt sich in einer Siebel-Umgebung einfach einrichten. Von der ersten Analyse über einen Proof of Concept bis zum Going-Live vergehen in der Regel nur wenige Wochen. Im laufenden Betrieb arbeitet sie geräuschlos und erfordert nur minimalen Administrationsaufwand. Der nachhaltige Nutzen stellt sich unmittelbar ein: immense SQL-Beschleunigung, die Siebel-Abfragen in den Bereich „gefühlter Echtzeit“ bringt.

Referenzkunde: Die Mobiliar

Bei der **Schweizerische Mobiliar Versicherungsgesellschaft AG** betreuen rund 80 Generalagenturen über 1,2 Millionen Kunden. Die eingesetzte Siebel-Lösung war der Verwaltung von Kunden- und Vertragsdaten sowie der Unterstützung von Marketingaktivitäten jedoch nicht mehr gewachsen, lange Antwortzeiten waren die Folge. DIMENSIO reduzierte die Dauer von SQL-Abfragen von über 3 Stunden auf 2 Sekunden, was einer Beschleunigung um den Faktor 6.000 entspricht. Gleichzeitig konnte die Serverlast von 26 GB auf 5 MB – um das 6.300-fache – verringert werden.

„Bisher konnten wir das CRM aufgrund der immensen Datenmengen nur eingeschränkt nutzen. Durch DIMENSIO erfährt die Datenabfrage eine enorme Beschleunigung und leistet damit einen signifikanten Beitrag zur Effizienzsteigerung im Datenhandling“.

(Den kompletten Referenzbericht finden Sie auf www.diso.ch)

