

In-Memory für effiziente Business-Intelligence-Systeme

Reinhard Mense, areto consulting gmbh

Die Themen „Big Data“ und „Digitalisierung“ dominieren derzeit den BI-Markt. In-Memory-Datenbanken sind zunehmend in den Hintergrund gerückt und stehen daher nahezu unbemerkt in ihrem Schatten. Dabei spielen sie eine nicht weniger bedeutungsvolle Rolle, denn sie führen im Rahmen von BI-Systemen zu erheblichen positiven Veränderungen und bieten einen hohen Nutzen sowohl für die IT als auch für die Fachbereiche.

Die Endanwender der Fachbereiche erwarten von einem guten BI-System, dass sämtliche Berichte und Analysen möglichst performant ausgeführt werden – unabhängig davon, ob es sich um regelmäßig ausgeführte Standardberichte oder Ad-hoc-Berichte für individuelle Analysen handelt. Die Bedeutung flexibler Analysen wird durch den vermehrten Einsatz von Data-Science-Methoden für die Erkennung von Zusammenhängen und korrelierenden Daten oder die Erstellung von Prognosen weiter gesteigert.

Hoher Optimierungsaufwand klassischer Systeme enttäuscht Endanwender

In klassischen BI-Systemen ohne In-Memory-Lösung lässt sich eine hohe Performance nur durch umfangreiche Optimierungsmaßnahmen erreichen. Um effektiv zu sein, müssen Optimierungen zielgerichtet durchgeführt werden. Demzufolge sind Aufbau und Inhalt für Standardberichte im Vorfeld genau mit den Anwendern der Fachbereiche abgestimmt. Die Abfragen, Filter und Parameter sind bekannt und die erforderlichen Optimierungen in der Datenbank können von der IT exakt darauf ausgerichtet sein.

Bei Ad-hoc-Abfragen ist dies jedoch nicht der Fall, da das Spektrum an Anfragen hier deutlich umfangreicher ist und die exakten Abfragen vorher unbekannt bleiben. Die Optimierungsmaßnahmen können sich daher bestenfalls auf wahrscheinlich häufige Abfragemuster konzentrieren. Eine umfassend stabile Performance für die Gesamtheit der Ad-hoc-Abfragen lässt sich somit nicht sicherstellen. Für unerwartete Abfragen können immer wieder zusätzliche Optimie-

rungen des BI-Systems erforderlich sein. Die Endanwender werden damit in ihrer täglichen Analysearbeit behindert und sind häufig enttäuscht von der Leistungsfähigkeit des BI-Systems.

In-Memory-Lösungen reduzieren Optimierungsaufwand

In-Memory-Datenbanken versprechen eine durchgehend hohe Performance für BI-Systeme. Die Besonderheit dabei: Performante Abfragen können ohne aufwändige Optimierungen erzielt werden.

In-Memory-Datenbanken sind für analytische Anwendungen optimiert und ermöglichen somit eine sehr gute Performance bei der Analyse großer Datenmengen. Die hohen Abfragegeschwindigkeiten ergeben sich dabei nicht nur allein aus der Tatsache, dass die Daten im Hauptspeicher vorgehalten werden, sondern auch aus einer für den Hauptspeicher optimierten Speicherung der Daten [1].

Die In-Memory-Datenbanken der meisten Hersteller setzen auf eine effektive Kombination von spaltenorientierter Speicherung (Column Store) und leichtgewichtigen Komprimierungsverfahren [2]. Diese Techniken erlauben eine höchst effiziente Nutzung des Hauptspeichers, sodass mehr Daten im Hauptspeicher vorgehalten werden können. Sie bilden somit die Grundlage für eine nachhaltig hohe Abfrage-Performance ohne umfassende Optimierungsmaßnahmen. Darüber hinaus werden häufig eine massive Parallelisierung der Datenzugriffe und weitere auf In-Memory Column Stores abgestimmte Techniken eingesetzt, sodass ein möglichst schneller Zugriff auf die Daten gewährleistet werden kann.

Die Oracle In-Memory-Option

Etablierte Hersteller wie SAP und spezialisierte Hersteller wie EXASOL haben mit ihren In-Memory-Lösungen SAP HANA und EXASolution die traditionellen Datenbank-Hersteller unter Druck gesetzt. Während durch das Erscheinen von SAP HANA das Thema „In-Memory“ vor allem medial stärker in den Fokus rückte, erzielte EXASolution beim TPC-H-Benchmark, der die Performance von Ad-hoc-Abfragen in analytischen Systemen misst, regelmäßig Bestwerte [3]. Oracle hat auf diesen Umstand reagiert und im Jahr 2013 mit der In-Memory-Option eine eigene In-Memory-Lösung präsentiert.

Wie bei den Lösungen vieler anderer Hersteller kommen auch bei der Oracle In-Memory-Option ein Column Store und leichtgewichtige Komprimierungsverfahren zum Einsatz. Die Oracle In-Memory-Option wird als Ergänzung zur klassischen Oracle-Datenbank angeboten. Die persistente Speicherung der Daten erfolgt deshalb weiterhin in der klassischen, zeilenorientierten Form (Row Store), lediglich im Hauptspeicher werden die Daten in einem Column Store gespeichert, sodass man von einer Dual-Format-Architektur sprechen kann. Auf die aufwändige Erstellung von analytischen Indizes zur Performance-Optimierung kann für die Daten im Hauptspeicher verzichtet werden.

Einsatz der Oracle In-Memory-Option

Die von Oracle gewählte Dual-Format-Architektur erlaubt eine schnelle und einfache Einführung der Oracle In-Memory-Option für bereits bestehende Data-Warehouse-Systeme, die bisher die klassische Oracle-Datenbank genutzt haben. Da die persis-

tente Speicherung nicht verändert wird, ist die Datenbank nur darüber zu informieren, welche gewünschten Daten zusätzlich im Hauptspeicher vorzuhalten sind. Dies geschieht mit wenigen sehr einfachen SQL-Anweisungen und erfordert kaum Entwicklungsaufwand.

Eine weitere interessante Eigenschaft der von Oracle gewählten Architektur besteht darin, dass die In-Memory-Option auch im OLTP-Umfeld eingesetzt werden kann und so performante Analysen direkt auf diesen Systemen ermöglicht, ohne sie zusätzlich zu belasten. Zum einen lassen sich Real-Time-Auswertungen somit vereinfachen, zum anderen ist es möglich, kurzfristig erste und schnelle Analysen für Daten durchzuführen, die bisher noch gar nicht in ein Data-Warehouse-System integriert sind. An dieser Stelle sei jedoch darauf hingewiesen, dass der Aufbau eines Data-Warehouse-Systems damit nicht gegenstandslos wird, da es als Single Point of Truth (SPOT), in dem die Daten verschiedener Quellen integriert, homogenisiert und historisiert bereitgestellt werden, weiterhin erforderlich ist.

Grenzen der Oracle In-Memory-Option

Für die regelmäßige Beladung des Data Warehouse sind nicht nur performante Abfragen von Bedeutung, sondern auch schnell schreibende Zugriffe. Diese profitieren bei der Oracle In-Memory-Option davon, dass analytische Indizes nicht mehr nötig sind und somit beim Schreiben von Daten auch nicht mehr erzeugt beziehungsweise gepflegt werden müssen. Gleichzeitig erfordert die von Oracle gewählte Dual-Format-Architektur aber weiterhin Optimierungsmaßnahmen für das effiziente Schreiben großer Datenmengen in den Row Store.

Durch den Einsatz der Oracle In-Memory-Option lassen sich bei deutlich reduziertem Optimierungsaufwand Abfragen umfassend beschleunigen. Für schreibende Zugriffe fällt die Beschleunigung geringer aus, somit sind an dieser Stelle Optimierungsmaßnahmen weiterhin erforderlich. Hier punkten

In-Memory-Lösungen, die ausschließlich auf Column Stores setzen, da bei diesen auch für Schreibzugriffe eine gute Performance ohne umfassende Optimierungen erzielt werden kann. Allerdings können diese im Gegensatz zur Oracle In-Memory-Option nicht im OLTP-Umfeld eingesetzt werden.

Data Science profitiert von In-Memory-Lösungen

In den Fachbereichen werden vermehrt Data-Science-Methoden für umfassende Analysen und Prognosen eingesetzt. Die Erstellung von Analyse- und Prognosemodellen erweist sich jedoch oft als rechen- und ressourcenintensiv. Die Datenbank-Hersteller gehen deshalb dazu über, die Ausführung von Data-Science-Methoden direkt in ihren Datenbanken zu ermöglichen. Damit können auch Data-Science-Analysen von der guten Performance der In-Memory-Datenbanken profitieren und zum Beispiel mehr Modelle berechnen und prüfen, um so genauere Analyse-Ergebnisse und Prognosen zu erstellen.

Fachliche Anforderungen rücken wieder in den Fokus der IT

Durch die Reduzierung des Optimierungsaufwands kann sich das BI-Team der IT auf die Umsetzung der fachlichen Anforderungen konzentrieren – damit steht wieder das Wesentliche im Fokus. In der Praxis zeigt sich, dass dadurch Unklarheiten, Unstimmigkeiten oder Ungenauigkeiten in den fachlichen Anforderungen frühzeitiger vom Entwicklungsteam erkannt werden [4].

Risiko des Scheiterns von BI-Projekten reduziert sich

Die Fokussierung auf die fachlichen Anforderungen und die Reduzierung des Entwicklungsaufwands führen insgesamt zu einer schnelleren Umsetzung der Anforderungen. Auch von BI-Projekten wird zunehmend mehr Agilität gefordert. Unabhängig von den eingesetzten Projektmanagement-Methoden erweisen sich die BI-Projekte hier

noch oft als schwerfällig. Jede Aufwandsreduzierung ist an dieser Stelle somit willkommen, denn nur wenn Anforderungen von der IT schneller umgesetzt werden können, erhält das BI-Team auch frühzeitiger Feedback von den Endanwendern [5]. Das Feedback kann somit wiederum in die Entwicklung einfließen und dazu beitragen, dass die Qualität des BI-Systems verbessert wird. Die kurzen Kommunikationszyklen und Reaktionszeiten führen zu einer deutlich höheren Zufriedenheit der Nutzer und folglich zu einer höheren Akzeptanz des BI-Systems.

Fazit

Abschließend bleibt festzuhalten, dass In-Memory-Lösungen ohne umfassende Optimierungsmaßnahmen eine gute Performance von BI-Systemen gewährleisten können, nicht nur für Standardberichte, sondern auch für Ad-hoc-Abfragen. Die Endanwender können damit die Analyse-Aufgaben des Fachbereichs besser erfüllen und werden ermutigt, mehr Analysen durchzuführen, sodass daraus neue Impulse für das Geschäftsmodell des Unternehmens entstehen können.

Literatur

- [1] Hasso Plattner, Alexander Zeier: In-Memory Data Management – Technology and Applications, Second Edition, Springer-Verlag, ISBN 978-3-642-29574-4
- [2] Marktstudie In-Memory-Datenbanken: www.aretto-consulting.de/marktstudien
- [3] TPC-H-Benchmark: www.tpc.org/tpch
- [4] www.bi-labor.de/datenbanken/nutzen-von-in-memory-datenbanken-fuer-bi-systeme
- [5] Michael Lang (Hrsg.): Business Intelligence erfolgreich umsetzen – Von der Technologie zum Geschäftserfolg, Beitrag „Agile BI-Architekturen“ Reinhard Mense (Autor), Symposium, ISBN 978-3-86329-682-7
- [6] www.bi-labor.de/datenbanken/vorteile-in-memory-column-store-vor-row-store-datenbank

Reinhard Mense

reinhard.mense@aretto-consulting.de

Save
the Date

10. Primavera Community Day

15. Februar 2017