

Machen In-Memory-Datenbanken das Data Warehouse überflüssig?

In Memory-Datenbanken in BI-Systemen

Mit der Einführung der 64-Bit-Technologie war eine der wichtigsten und am meisten erwarteten Neuerungen speziell im Bereich Data Warehousing/Business Intelligence sicherlich das Aufkommen von In-Memory-Datenbanken auch für wirklich große Datenmengen. Die spaltenorientierte Form wurde speziell für Analysensysteme optimiert und durch ihren Einsatz wird eines der wichtigsten Ziele eines BI-Systems, die sehr gute Verarbeitungs- und Abfrageperformance, fast spielend erreicht und dies bei einer stark vereinfachten Handhabung.

Das Package hilft dem Anwender online eine Tabelle zu kopieren. Ergänzend dazu werden weitere Objekte der Tabelle (Constraints, Indizes etc.) automatisch erzeugt. Die Funktionsweise wird anhand des folgenden Beispiels beschrieben:

Als Ausgangspunkt wird eine Tabelle betrachtet, die nachträglich partitioniert werden soll.

Stärken und Schwächen der verschiedenen Systeme

Es gibt aktuell die Tendenz, auch in transaktionalen Systemen vermehrt die In-Memory-Technologie zum Einsatz zu bringen. Dies haben einige Hersteller für die Ankündigung zum Anlass genommen, auch tiefergehende Analysen (für die andernfalls ein BI-System aufgebaut würde) in Zukunft direkt auf den ERP-Strukturen stattfinden zu lassen. Lässt sich daraus ableiten, dass BI-Systeme überflüssig werden?

Nun sind beim Aufbau eines Data Warehouse-Systems neben der Performance auch noch weitere wichtige Themen zu berücksichtigen, beispielsweise die Konsolidierung verschiedener Quellen, Datenqualität, Einfachheit der Auswertungsstrukturen, Sicherheitseinschränkungen (diese sind meistens zwar an die der ERP-Systeme gekoppelt, aber nicht identisch!) sowie die Kontrollierbarkeit der Prozesse. All dies ist nicht automatisch erfüllt, wenn Analysen in der ERP-Datenbank stattfinden.

Ist die Ankündigung der ERP-Hersteller also völlig falsch?

Nein, natürlich nicht, es lohnt sich aber, die Stärken und Schwächen der verschiedenen Systeme etwas nüchterner zu betrachten.

Betrachten wir zunächst die großen Stärken der DWH-Systeme: Verschiedene Systeme miteinander integrieren und konsolidieren, das alles in einem definierten, robusten und kontrollierbaren Ladeprozess, übergreifende Auswertungen, eine präzise Historisierung der Stammdaten inklusive Bereinigung, langfristige Auswertungen inklusive statistischer Methoden, eine genaue Zuweisung von Rechten für die Anwender und natürlich das alles in der benötigten guten Performance, sowohl, was die Beladung, als auch, was die Analysen betrifft.

Es gibt aber auch einige Anforderungen wie umfangreiche Auswertungen über Einzelsätze, Real- und Near-Realtime-Auswertungen und schnelle Ergänzungen der Datenstrukturen, für die es zwar durchaus produktiv eingesetzte Lösungen gibt, diese erfordern aber bisweilen deutlich mehr Aufwand:

- Es ist unbestritten, dass das DWH die Informationen auf der granularsten Ebene der Einzelsätze benötigt, sowohl um die Konsolidierung zu erzeugen und zu prüfen, als auch um Detailauswertungen zu ermöglichen, die Übernahme wirklich aller Detailattribute und technischer Schlüssel, ganz zu schweigen von angehängten Daten (wie z.B. eingescannten Rechnungen) kann das DWH aber überfrachten, zumal viele der Informationen wirklich nur für wenige Detailanalysen benötigt werden, wenn überhaupt.
- Der Ladeprozess kann von jedem geschriebenen oder veränderten Einzelsatz im ERP-System oder wenigstens sehr viel häufiger als täglich angestoßen werden, um Real- und Near-Realtime-Auswertungen zu ermöglichen, dies kann aber zu größerer Prozesskomplexität für die Schnittstellen und den Ladeaufwand führen.



Machen In-Memory-Datenbanken das Data Warehouse überflüssig?

Kombinierter Einsatz für eine künftige BI-Landschaft

Genau hier können die ERP-Systeme unter Einsatz der In-Memory-Technologie ihre Stärken ausspielen:

- Wenn alle Detailsätze im Quellsystem vorhanden sind und dort sinnvoll abgefragt werden können, so sind in den Einzelsätzen im DWH nur noch die Attribute nötig, die auch in Auswertungen relevant sind.
- Für Realtime-Auswertungen könnten die aktuellen Daten aus dem ERP mit den konsolidierten langfristigen aus dem DWH kombiniert werden. Im Front-End werden die Daten für die Endanwender als ein gemeinsames, einfach strukturiertes Analysesystem dargestellt. Dadurch entfällt die Entwicklung von aufwendigen Real-Time-ETL-Prozessen, die Latenzzeiten werden gering (es handelt sich also um wirkliches Real-Time) und auch die Belastung der ERP-Datenbanken ist durch den Einsatz der In-Memory-Technologie relativ gering.
- Natürlich werden hierbei keine Prüfungen der Datenqualität durchgeführt, eine solche ist in einem Real-Time-Prozess faktisch kaum darstellbar. Dies und auch die Konsolidierung der verschiedenen Systeme, die Historisierung etc. ist weiterhin die Aufgabe des gewohnten täglich laufenden ETL-Prozess des BI-Systems, in dem die Daten ins DWH übernommen werden.

Die Idee ist also nicht ein „Entweder oder“, sondern eine Kombination der Stärken beider Ansätze. Das DWH ist nicht überflüssig, sondern hat weiterhin seinen Platz in der Konsolidierung und Optimierung der Auswertungssysteme. Die In-Memory-ERP-Systeme können aber eine extrem wertvolle Ergänzung sein und insbesondere Prozesse, die schnelle Reaktionen erfordern, massiv vereinfachen.

