

von Otto Görlich

Die BI-Vision von IBM lässt sich wie folgt ausdrücken: „Ziel ist, BI-Funktionen in die Datenbank zu integrieren und sie ausschließlich über offene Standardschnittstellen als Teil einer integrierten BI-Plattform zugänglich zu machen. Für weitere Schichten der BI-Architektur arbeiten wir mit Geschäftspartnern zusammen.“

## Über den Autor

Otto Görlich ist Data Management Solution Consultant bei IBM Software Group, Central Region.

## BI-Funktionen in die Datenbank integrieren



Die Feststellung, dass IBM BI-Funktionalitäten in die Datenbank integriert, ist nicht so trivial wie sie klingt. Viele BI-Anwendungen extrahieren große Datenvolumina mit Hilfe einer einfachen SQL-Syntax aus dem Data Warehouse, bereiten sie in der Anwendungsschicht auf und filtern (oder bündeln) sie, um die gewünschte Granularität zu erreichen. Erst dann wenden sie die Funktionalität an, die die eigentliche Wertsteigerung darstellt. Client-Tools gehen oft genauso vor, ob der temporäre Cache nun ein Spreadsheet, OLAP-Würfel oder lokales Dateisystem ist. Viele Benutzeranwendungen ahmen dieses Modell nach, weil es das Vorbild ist, das die Entwickler am besten kennen. IBM DB2 UDB hat sich stetig in die Richtung weiterentwickelt, BI-Funktionen innerhalb der Datenbank zu unterstützen. Zu diesen BI-Fähigkeiten gehören Data Mining, OLAP, ETL, geografische Analysen und komplexere Statistik- und Analysefunktionen für Regression, Kovarianz, Stichproben, Ranking, gleitende Kennzahlen etc. – im Folgenden als „BI-Funktionen“ bezeichnet. OLAP-Tools eignen sich sehr gut für eine flexible Analyse des Unternehmens. Die Wirksamkeit von Marketing-Kampagnen hängt zum Beispiel von vielen Variablen ab, so dass ein vordefinierter Bericht vielleicht nicht alle benötigten Informationen zeigt. Aber wo beginnt man eine multidimensionale Analyse?

Um den OLAP-Würfel abzusuchen oder ein Drill-Down durchzuführen, muss man erst wissen, wie die Attribute zu definieren sind. Wenn zum Beispiel die geografische Lage ein wichtiger Faktor für die Einschätzung der Wirksamkeit einer Marketing-Kampagne ist, sollte dann die hierarchische Gliederung auf Städten, Postleitzahlen oder sonstige Untergliederungen aufgebaut werden? Und sollten diese dann nach oben zu Bezirken, Ländern oder Regionen führen? Die Mining-Funktionalität

von DB2 lässt den Benutzer wichtige Dimensionen und Attribute automatisch erkennen und hierarchisch gliedern. Wenn sich herausstellt, dass die Bevölkerungsdichte eine deutliche Korrelation zu den Faktoren Marketing-Kosten und Verkaufszahlen aufweist, dann kann in DB2 eine Analysefunktion aufgerufen werden, die ein Diagramm der verschiedenen Stufen der Bevölkerungsdichte erzeugt. Wenn diese Stufen Attribute eines OLAP-Würfels sind, können Client-Tools ein Drill-Down anhand der verschiedenen Bevölkerungsdichten durchführen, um so einen Eindruck von der Wirksamkeit von Marketing-Programmen zu gewinnen.

### **Die Vorteile von BI-Funktionen in DB2 sind deutlich und vielfältig:**

- Sie ermöglichen der Datenbank, genaue Daten mit der gewünschten Granularität an BI-Endpunkte (Anwendungen, Benutzer oder Tools) zu liefern.
- Sie verlagern einen größeren Teil der „Schwerarbeit“ - Scannen, Sortieren, Zusammenführen und Bündeln der Daten - auf die Ebene der Architektur, die speziell dafür geschaffen und optimiert wurde, den Warehouse-Datenbankserver.
- Sie verringern die Menge an Daten, die durch das Netzwerk fließen.
- Sie setzen nicht so viele Daten den weniger sicheren Bereichen außerhalb des Firewalls aus.
- BI-Funktionen auf der Basis derselben Daten, im selben DB2 Data Warehouse, geben viel eher „eine einzige Version der Wahrheit“ für das gesamte Unternehmen wieder - unabhängig vom Tool oder der Anwendung des Endbenutzers.

BI-Funktionen als Teil einer integrierten Plattform ausschließlich über offene Standardschnittstellen zugänglich machen. Wenn es offensichtlich sinnvoll ist, die Warehouse-Datenbank für BI-Funktionen zu nutzen, warum führen dann so viele Tools und Anwendungen einen Großteil der Arbeit lieber selber aus? Oder anders gefragt: Warum fungieren so viele BI-Endpunkte als Datengroßhändler und nicht als Verbraucher? Teilweise lässt sich dies auf die Entwicklungen aus einer Zeit zurückführen, in der sich sogar einfachste Hierarchien nicht in SQL ausdrücken ließen. Für multidimensionale Analysefunktionen wird seit langem eine spezielle Struktur bevorzugt, der MOLAP-Würfel (Multidimensional OnLine Analytical Processing) mit proprietären Speicherstrukturen und Schnittstellen. Data Mining-Algorithmen werden üblicherweise auf Dateien außerhalb der Datenbank angewendet.

In diesen Beispielen werden die Daten vom Warehouse für eine Zwischenplattform aufbereitet und dort umstrukturiert, ehe die BI-Funktionen eingesetzt werden. Dies führt zu mindestens drei verschiedenen, voneinander abhängigen Datenstrukturen, die drei unterschiedliche Zeitpunkte widerspiegeln und auf drei Serverplattformen verteilt sind- jede mit eigenen APIs (Application Programming Interface), Verwaltungsfunktionen und Optimierungstechniken. Die einzelnen Analyse-Silos spielen ihre Rolle zwar gut, das Verfolgen des Analysevorteils hat jedoch den grundlegenden Wert des Data Warehouse zunichte gemacht. Kritische Abweichungen, die sich in einem großen OLAP-Würfel verbergen, können recht mühsam zu finden sein. Mining-Algorithmen für die Entdeckung von Abweichungen sind in der Lage ungewöhnliche Werte automatisch hervorheben. In einigen Fällen können solche Unregelmäßigkeiten auf eine neue Geschäftschance hinweisen, in anderen wiederum auf ein Problem, das unbedingt gelöst werden muss.

Sie müssen sich als Abonnent anmelden um den hier fehlenden Teil des Inhalts zu sehen. Bitte [Login](#) für Zugriff.

Noch nicht Abonnent? [Sonderaktion nutzen](#).

- [7 Euro/Monat NEWSabo digital - sofort zugreifen & online bezahlen.](#)
- [13,5 Euro/Monat NEWSabo plus inkl. 5x Logins & Print-Ausgaben - sofort zugreifen & per Firmen-Rechnung bezahlen.](#)

[/private]