

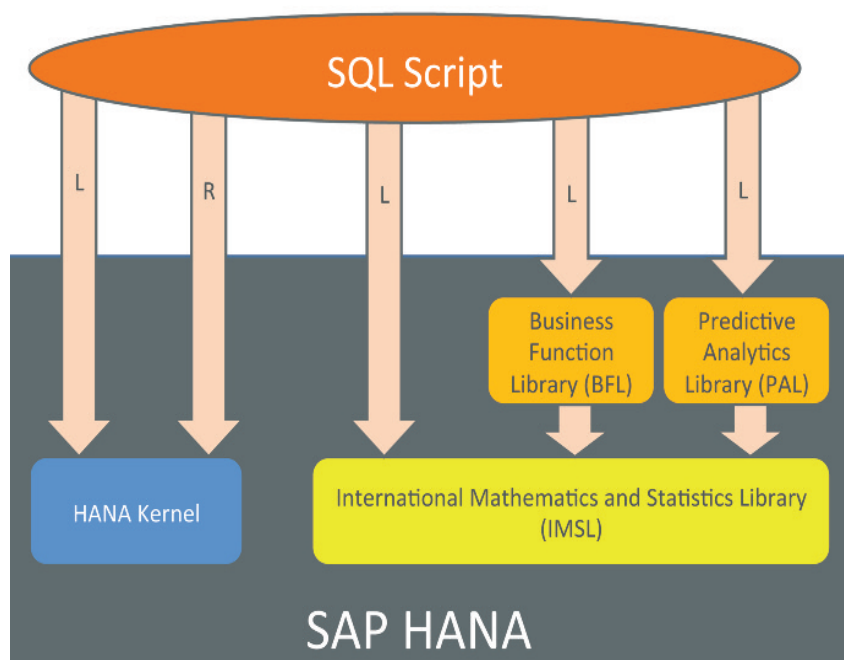
„SAP HANA“ bringt neues Potenzial für Anwendungen

Weit mehr als ein schneller Datenspeicher

Mit der Einführung der HANA-Technologie und durch deren schrittweise Integration in die eigenen Produkte hat SAP den Markt für Businesssoftware in Bewegung gebracht. Auch wenn die High-Performance Analytic Appliance – so der vollständige Name von HANA – das Thema Analyse bereits im Namen trägt, denken viele bei HANA in erste Linie an Performancegewinne durch den Einsatz von In-Memory-Technologie, spaltenbasierter Ablage und vielen weiteren technologischen Kniffen. Aber HANA ist mehr als ein schneller Datenspeicher. Sie enthält vielmehr auch eine Vielzahl wertvoller, in-memory-optimierter Funktionalitäten, die aufgrund der Nähe zu den Daten im Arbeitsspeicher auch sehr komplexe Analysen in Echtzeit erlauben und das Potenzial für Anwendungsklassen eröffnen, die lange Zeit undenkbar waren.

Von Thorsten Füg*
und Dr. Marcus Dill**

Während im Bereich Reporting und Interaktive Analyse die Vorteile von SAP HANA aufgrund der bislang ungekannten Verarbeitungsgeschwindigkeit unübersehbar sind, blieben die komplexen Analysefunktionen von SAP HANA bisher eher im Verborgenen, da es bisher nur wenige Anwendungen und Kunden gibt, die sie nutzen. Für viele Beobachter ist diese Entwicklung auch unerwartet angesichts der Produkthistorie von SAP, in der man nur zaghafte Versuche erkennen kann, wirklich brauchbare Funktionalität für komplexe Datenanalysen (Data Mining, Predictive Analytics) anzubieten. Aber ganz unverkennbar hat SAP nunmehr erkannt, dass moderne Geschäftsprozesse durchgesetzt sind von intelligenten Fragestellungen, die teilweise auch automatisiert ablaufen müssen. Hier muss auch SAP Lösungen bieten, will man nicht dieses äußerst spannende und lukrative Feld der Konkurrenz überlassen und Kun-



Analytische Funktionen und Schnittstellen von SAP HANA

den, die eigentlich eine klare SAP-Strategie haben, in gegebenenfalls teure und riskante Integrationsansätze zwingen.

Will man sich als Anwender der komplexen Analytik in SAP HANA bedienen, dann bewegt man sich heute noch über weite Strecken auf der Ebene von Programmierung, das heißt, man entwickelt selbst Anwendungen, die über die eine oder andere Schnittstelle mit SAP HANA kommunizieren und dabei nicht nur Daten bewegen, sondern auch analytische Funktionen von SAP HANA nutzen und ganze Blöcke von Anwendungslogik in SAP HANA verlagern. Der klassische Weg, über einzelne SQL-



*Thorsten Füg ist BI-Berater mit Schwerpunkt „SAP BW“ und „SAP HANA“ bei der mayato GmbH.



**Dr. Marcus Dill ist Geschäftsführer der mayato GmbH.

Befehle hinaus Programmabläufe in die Datenbank zu bringen, sind Stored Procedures. Auch SAP HANA bietet für diesen Zweck grundsätzlich zwei Optionen: „SQL Script“ und „L“. Nur erstere ist für die Entwicklung von Erweiterungen durch Kunden und Partner freigegeben, während „L“ – eine „C“ ähnliche Programmiersprache – bisher noch der Verwendung durch die SAP-eigenen Produktentwickler vorbehalten ist. Es ist jedoch eine native Einbettung von „L“ in SQL Script analog zu „R“ geplant, die damit dann auch allgemein freigegeben wäre.

SQL Script ist die wichtigste Programmierschnittstelle für Anwendungen auf SAP HANA. Es basiert auf einem erweiterten Standard-SQL und dient primär der Verlagerung von datenintensiver Anwendungslogik in die Datenbank. Alle in diesem Artikel behandelten analytischen Funktionalitäten (R, PAL, BFL) sind über SQL Script nutzbar oder werden dies über kurz oder lang sein.

Die objektorientierte Programmiersprache „R“ zur statistischen Analyse und grafischen Darstellung von Daten hat in den letzten Jahren längst über das akademische Umfeld hinaus in Unternehmen Einzug gehalten. Die Vielzahl an Analysefunktionen, Visualisierungsmöglichkeiten und Datenschnittstellen – ständig vergrößert von einer großen Open-Source-Community – hat auch viele Softwarehersteller auf die Idee gebracht, diesen Schatz an Werkzeugen ihren eigenen Anwendern zugänglich zu machen. Nicht zuletzt auch für die zunehmend relevanten Big-Data-Szenarien, beispielsweise die Analyse von Social-Media-Daten, bieten zahlreiche in „R“ über die Jahre entstandenen und ausgereiften APIs einen erheblichen Mehrwert.

SAP HANA erlaubt über zwei Ansätze eine Integration mit „R“. Im sogenannten Outside-In-Ansatz kann HANA wie jede andere Datenbank über JDBC oder ODBC angebunden werden. SAP selbst empfiehlt nicht die Verwendung dieses Weges und bringt noch in 2012 mit „RHANA“ einen eigenen Treiber für den Zugriff aus „R“ auf HANA-Tabellen. Insbesondere bei der Verwendung von „R“ als Beladungswerkzeug, z.B. für Big Data, bringt der RHANA-Treiber deutliche Performancevorteile.

Neben dieser noch eher klassischen Integration gibt es jedoch auch noch den sogenannten Inside-Out-Ansatz. In diesem Fall werden in SQL Script enthal-

tene „R“-Befehle von HANA in entsprechende Aufrufe eines verbundenen „R“-Server umgesetzt. Diesen Aufrufen werden die benötigten HANA-Daten direkt mitgegeben. Auch in diesem Fall werden also Daten in eine „R“-Umgebung übertragen und die „R“-Funktionen nicht direkt in HANA ausgeführt. Die viel beschworene enge Integration von Daten und Verarbeitungslogik stößt hier noch an architektonische Grenzen. SAP arbeitet intensiv daran, die Übertragungswege zwischen HANA und „R“-Server so kurz und effizient wie möglich zu machen, um den Performanceverlust durch die Datenübertragung zu minimieren. Eine Vollintegration im Sinne von innerhalb von SAP HANA laufenden „R“-Funktionen ist jedoch nicht zu erwarten.

SAP verlässt sich jedoch hinsichtlich analytischer Funktionen nicht nur auf die Open-Source-Community, sondern entwickelt auch eigene Analysefunktionen, die direkt in HANA ausgeführt werden. Diese Funktionen sind in zwei Libraries gebündelt:

- Die Predictive Analysis Library (PAL) enthält eine Sammlung von in-memory-optimierten Algorithmen zur automatisierten Erkennung komplexer Zusammenhänge in großen Datenmengen und zur Prognose zukünftiger Ereignisse, z.B. Kundenverhalten, Kreditrisiken, Betrug oder Ersatzteilbedarfen. Im Herbst 2012 umfasst die Bibliothek sieben Algorithmen in den drei Data-Mining-Kategorien „Cluster-“, „Klassifikation-“ und „Assoziationsanalyse“. Die PAL wird von einigen Anwendungen der SAP verwendet (aktuell vor allem im Bereich CRM und Retail), kann aber auch von eigenentwickelten Applikationen aufgerufen werden. SAP bietet mit „BusinessObjects Predictive Analysis 1.0“ auch einen vollwertigen Data-Mining-Client, der die PAL-Funktionen mit hilfreichen Visualisierungen und einer durchdachten Modellierung unterstützt. Aber auch Drittanbieter mit Best-of-Breed-Werkzeugen wie Ankhor bieten eine ausgezeichnete Integration mit HANA im Allgemeinen und der PAL im Besonderen.

- Die Business Function Library (BFL) besteht aus einer Reihe von typischerweise in Unternehmen verwendeten komplexen Berechnungsfunktionen (wie z.B. Abschreibungen oder Kapazitätenoptimierung, Bedarfsana-

lysen, Vorjahresvergleich). Diese sind in „C++“ entwickelt und können zur Laufzeit über Parameter gesteuert werden. Auch Kunden ohne Programmierkenntnisse bietet diese BFL über sogenannte „Calculation Views“ eine wesentlich performantere und einfachere Alternative zu SQL Script. Wesentliche Nutzer dieser Business Function Library werden SAP-eigene Anwendungen sein, allen voran die Planungsanwendungen im „SAP BW“. Einige der notorisch performancekritischen Planungsfunktionen, die bisher in der Applikationsschicht des SAP BW abliefen, können ab der Version 7.3 des SAP BW auch innerhalb von HANA ausgeführt werden. In Zukunft darf auch mit weiteren typischen BW-Funktionen wie z.B. der Währungsumrechnung innerhalb von HANA gerechnet werden.

SAP plant weiterhin die Integration der „IMSL“ (International Mathematics and Statistics Library), die eine Vielzahl an in nativem „C“ entwickelten und daher sehr performanten numerischen Funktionen bieten. Die Anbindung der IMSL wird ebenfalls über SQL Script und „L“ erfolgen. Intern setzen schon heute Teile von PAL und BFL auf der IMSL auf. Ursprünglich eine unabhängige Technologierichtung, haben Suchfunktionen über das vergangene Jahrzehnt hinweg Eingang in die Business-Intelligence-Welt gefunden. Innerhalb von SAP HANA sind sie eine besonders weitgehende Symbiose eingegangen. So können über die als SQL-Funktion integrierte Fuzzy Search Texte schnell und fehlertolerant durchsucht werden. Fehlertolerant bedeutet, dass Datenbankabfragen dank intelligenter String-Matching-Algorithmen auch mit falsch geschriebenen Wörtern umgehen können, wie es z.B. von Suchmaschinen bekannt ist.

Seit Kurzem bietet SAP HANA mit dem neuen Toolkit zusätzlich die Möglichkeit, eigene suchbasierte Anwendungen zu entwickeln. Die Anwendungen sollen den Kunden durch einen Echtzeitzugriff auf Informationen und facettierte Suchfunktionen bei der Analyse großer Mengen von strukturierten und unstrukturierten Textdaten unterstützen.

In Summe bietet HANA bereits heute einen umfangreichen Werkzeugkasten und vielfältige Möglichkeiten, auch allergrößte Datenmengen analytisch zu verarbeiten. Anwendern sind somit auch Fragestellungen erlaubt, die über einfache Aggregationen weit hinausgehen. (ap) ©