

Big-Data-Bundle von SAP für Hadoop

Dafür hat SAP jetzt ein neues [Big-Data-Bundle](#) geschnürt, das Echtzeit-Data-Warehouse-Umgebungen, sprich SAP HANA, mit Apache Hadoop verbindet. Dazu hat SAP jetzt eine gemeinsame Marktstrategie mit den Hadoop-Anbietern [Cloudera](#), Hitachi Data Systems, Hortonworks, HP und IBM angekündigt.

Neben SAP HANA kombiniert der Anbieter auch den Analyseserver SAP Sybase IQ, das Datenmanagement SAP Data Integrator und verschiedene Business-Intelligence-Lösungen aus dem BusinessObjects-Portfolio. In dem Bundle bietet SAP eine Lösung für Data Warehousing über die auch in größeren Datenmengen Echtzeitanalysen möglich sind. Das neue Bundle stellte SAP auf der Veranstaltung O'Reilly Strata Conference + Hadoop World vor, die vom 23. bis 25. Oktober in New York stattfindet.

Um die ständig steigenden Datenmengen zu speichern, zu analysieren und zu verarbeiten greifen Unternehmen immer häufiger zu Hadoop-basierten Lösungen. Die Schwierigkeit dabei ist jedoch häufig die Integration mit bestehenden Umgebungen für Datawarehousing oder Business Analytics. Hier will SAP mit dem neuen Angebot eine Brücke schlagen, die Hadoop-Technologien mit der Geschwindigkeit von In-Memory-Datenbanken verbinden sollen. So lassen sich die Informationen nicht nur in Feldern oder Zeilen, sondern auch in Spalten vorhalten, teilt SAP mit.

Als beispielhafte Anwendungen zitiert SAP das Unternehmen Mitsui Knowledge Industry, das sich auf Bioinformatik und Genomanalyse spezialisiert hat. Die analytische Echtzeitdatenbank dieses Unternehmens nutzt komplexe iterative Algorithmen von SAP HANA, um mit Hadoop vorverarbeitete Genomsequenzen zu untersuchen. Die für die Genomanalyse benötigte Zeit reduziert sich so von Tagen auf Minuten.

Ählich verfährt auch comScore, ein international tätiges Internet-Marktforschungsunternehmen, das sich auf die Analyse des Webs spezialisiert hat. Der Marktforscher verarbeitet mit Hadoop täglich Datenmengen im Terabyte-Bereich. Die Ergebnisse lädt das Unternehmen in sein Enterprise Data Warehouse, das auf SAP Sybase IQ basiert. Hier stehen die Ergebnisse Tausenden von comScore-Kunden zur Verfügung, die die Ergebnisse mit Self-Service-Werkzeugen analysieren können.

Kunden von SAP können Hadoop auf unterschiedliche Art und Weise in ihre bereits existierenden BI- und Datawarehouse-Umgebungen integrieren und an ihre Bedürfnisse anpassen. Mit der Software SAP Data Integrator sind sie in der Lage, Daten aus Hadoop Distributed File Systems (HDFS) oder Hive-Datenbanken auszulesen und relevante Informationen schnell in SAP HANA oder SAP Sybase IQ zu laden.

So stellt SAP sicher, dass BI-Anwender bestehende Berichts- und Analysewerkzeuge weiter nutzen können. Darüber hinaus lassen sich Abfragen über SAP Sybase IQ und Hadoop-Umgebungen kombinieren. Alternativ ist es Kunden möglich, MapReduce-Jobs über eine SAP Sybase IQ MPP-Umgebung mittels eingebauter Funktionalität laufen zu lassen. Anwender von SAP BusinessObjects BI können Abfragen innerhalb von Hive-Umgebungen durchführen. Analytiker haben so die Möglichkeit, Hadoop-Umgebungen direkt zu durchsuchen.

Durch die Zusammenarbeit mit weiteren Herstellern versucht SAP durchgängigen Support für die Big Data-Lösungen durchzusetzen.