

Pentaho, BIRT und JasperReports im Vergleich

Open Source Business Intelligence

Text: Samir Mimouh und Ruth Heidingsfelder

Längst haben sich Open-Source-Lösungen auch im Bereich der Business Intelligence etabliert. Inzwischen lautet die Frage nicht mehr, ob es kostengünstige Alternativen zu traditionellen BIANbietern wie Business Objects, Cognos, SAS etc. gibt, sondern welche Open-Source-Lösung gewählt werden soll. Da in den letzten Jahren immer mehr kommerzielle Anbieter quelloffener BIProdukte den Markt erobern, können sich diese vor allem im Bereich Berichtswesen bzw. Reporting über zunehmende Nachfrage freuen. Da für dieses Segment inzwischen recht viele Open-Source-Lösungen existieren, konzentriert sich der vorliegende Artikel auf die derzeit wichtigsten Reporting-Anbieter. Ist im Unternehmen erst einmal die Entscheidung gefallen, Open Source für Reporting-Aufgaben einzusetzen, stellt sich sofort die Frage nach der richtigen Report Engine. Oft soll eine bestehende Report-Implementierung mit möglichst geringem Aufwand abgelöst werden. Die neue Lösung muss zukunftssicher sein und in die eigene IT-Infrastruktur passen. Es gibt drei Anbieter, die im Moment den Open-Source-Markt für Reporting dominieren: Die Pentaho BI Suite, das von der Firma Actuate initiierte Eclipse-Projekt BIRT sowie JasperReports von JasperSoft. Die drei Lösungen werden im Folgenden vorgestellt und, soweit möglich, bezüglich der wichtigsten Eigenschaften miteinander verglichen.

Hinweis: Der vorliegende Artikel ist in Ausgabe 5.2008 das Java Magazins erschienen. In dieser Online-Version wurden jedoch einige Korrekturen in der Vergleichsmatrix (siehe Teil 3) im Zusammenhang mit BIRT vorgenommen.

Pentaho

Pentaho ist einer der wenigen Anbieter, die im Bereich Open Source Business Intelligence eine komplette Software Suite anbieten. Die Suite liegt im Moment in der Version 1.6 vor. Neben dem Reporting in Pentaho sind OLAP-Funktionalitäten, Dashboarding und mehrere Tools zu Erstellung und Verwaltung von Auswertungen enthalten. Pentaho ist aus vielen einzelnen Open-Source-Projekten entstanden und damit aus technisch sehr ausgereiften Komponenten. Im Reporting-Bereich kommt bei Pentaho das Reporting-Projekt JFreeReports zum Einsatz. Die Reporting Engine wird durch die Einbindung in Pentaho durch weitere Features erweitert. Standardfunktionalitäten im Reporting-Bereich sind die Gruppierung von Daten, die Anzeige von Header und Footer sowie die Generierung von Berichten im HTML- und PDF-Format. Pentaho bietet daneben die Verschachtelung von Reports mit Teilreports, die Parameterisierung von Berichten und die Verwendung mehrerer Datenquellen innerhalb eines Berichts. Diese Funktionalitäten finden sich jedoch auch in den anderen Reporting-Lösungen. Ein neuer Aspekt ist hingegen die Erstellung von Reports auf der Weboberfläche mittels Ajax. Das bedeutet einen weiteren Schritt hin zu einer einheitlichen BI-Webapplikation, den BIRT und JasperReports noch nicht gemacht haben.

Durch die Integration in die Pentaho Suite profitiert die Report Engine in weiteren Bereichen: Das Versenden von Reports ist durch eine E-Mails-Server-Anbindung möglich; es existiert ein

umfangreiches Auditing zur personalisierten Protokollierung von Reportaufrufen; das dafür nötige umfangreiche Sicherheitsmanagement ermöglicht die Verwaltung der Zugriffsrechte. All das erspart die Anschaffung und Integration zusätzlicher Komponenten, die nachträglich implementiert oder integriert werden müssten.

Im Bereich der Dateneinbindung bietet Pentaho ebenfalls viele Möglichkeiten. Zum einen kann eine direkte Verbindung mit jeder JDBC-fähigen Datenbank hergestellt werden, was heute dem Standard entspricht. Zum anderen kann auch ohne direkte Verbindung zu operationalen Datenbanken eine Datenkonnexion hergestellt werden, z.B. die Verbindung zu OLAP-Cube-Daten über einen MDX-Abfragestring.

Pentaho legt bei der Gestaltung seiner Suite generell viel Wert darauf, dass die Erstellung von Berichten auch für Nicht-Entwickler und IT-Profis einfach ist. So wird z.B. bei der Auswahl der Daten ein SQL Query Designer-Tool zum einfacheren Auswählen der Daten mitgeliefert, das dann die benötigten SQL Statements generiert. Für die Erstellung von MDX Queries existiert zwar noch keine grafische Oberfläche. Allerdings unterstützt Pentaho in der Pivot-OLAP-Analyse der Weboberfläche die Funktion, zu jeder Datenansicht den zugehörigen MDX-Abfragestring zu erhalten. Dadurch ist es möglich, Reports auch auf Basis bestehender OLAP Cubes zu generieren.

Das bedeutet für den Benutzer maximalen Komfort, da auf die bereits im Cube existierenden Metadaten zurückgegriffen werden und der Anwender seine Daten aus den aussagekräftigen Bezeichnungen der Daten zusammenstellen kann. Der Zugriff auf Metadaten aus Datenbanken ist aber auch ohne OLAP Cube möglich. Die Verbindung zwischen dynamischen Auswertungen und mehr oder weniger statischen Reports bedeutet überall dort leichte Bedienbarkeit, wo eine OLAP-Architektur existiert.

Ein Beispiel: Ein Benutzer drillt sich über die OLAP-Pivot-Tabellen durch seine Daten und möchte einen Bericht exakt von der aktuellen Zusammenstellung der Daten erstellen, der täglich mit aktuellen Daten zu Verfügung stehen soll. Da alle Informationen über die Datenauswahl bereits im MDX-Abfragestring enthalten sind, würde es ausreichen, den Abfragestring im Pentaho Report Design Wizard anzugeben. Dadurch wäre der komplette Bericht, sofern keine Formatanpassungen erfolgen sollen, vollständig definiert und könnte in der Applikation zu Veröffentlichung gespeichert werden. Der Report würde sich dann bei jedem Aufruf automatisch aus aktuellen Daten generieren. Diese Verschmelzung von voll dynamischen und statischen Reports wird heute immer stärker gefordert. Eine Stärke von Pentaho ist es, diese Funktion bereits zu unterstützen. Für die Erstellung und Pflege von Reports bietet Pentaho Entwicklern und Anwendern eine Vielzahl von Tools. Zur Unterstützung der genannten Metadaten-Funktionalität verfügt die Pentaho Suite über eine Metadaten-Layer, die dem Anwender nicht nur die oft wenig aussagekräftigen Datenbanktabellen und -spalten durch vertraute Begriffe ersetzt, sondern auch die zentrale Internationalisierung von Texten und die Steuerung von Formatierungen und Zahlenformaten erlaubt. Ein grafischer, auf Eclipse basierender Metadaten-Editor ermöglicht die Bearbeitung, Verwaltung und Veröffentlichung von Metadaten-Objekten im Pentaho Server und zeigt Beziehungen sowie das Mapping auf physikalische Tabellen auf.

Für die Erstellung eines Berichts bietet Pentaho zwei Tools an: Den Report Design Wizard und den Report Designer. Mit dem Report Design Wizard kann ein Bericht leicht und schnell in

maximal sieben Schritten definiert werden. Der Report Designer bietet die Möglichkeit, Reports mit Links, dynamischen Feldern, Parameterisierungen, Objektanordnung, grafischen Inhalten etc. pixelgenau zu definieren. Dabei kann man mit dem Wizard bestehende Berichte aus dem Report Designer als Template verwenden oder im Report Designer nachbearbeiten. Das Report Design Studio rundet das Reporting-Paket ab. Das Studio hilft bei der Einstellung der Art und des Zeitpunkts der Publizierung des Reports (z.B. als E-Mail-Anhang) und bei der Konfiguration der Interaktion mit dem Benutzer.

Um Reporting mit Pentaho betreiben zu können, müssen folgende Voraussetzungen bestehen: Die Pentaho Suite muss entweder auf einem J2EE- bzw. JEE-Applikationsserver oder einem Servletcontainer wie Tomcat aufgesetzt werden. Diese müssen für die später verfügbaren Datenquellen konfiguriert werden. Die verfügbaren Reporting-Tools müssen nicht angepasst werden und können als Standalone-Lösung auf jeder beliebigen Plattform abgelegt werden. Über eine Server-Authentifizierung können mit den Tools erstellte Reportdefinitionen auf den BI-Server hochgeladen werden. Diese minimale Installation reicht aus, um Reporting mit Pentaho zu betreiben. Soll jedoch beim Reporting auf OLAP-Daten zurückgegriffen werden, braucht es dafür natürlich eine Pentaho-OLAP-Architektur, bei der der Implementierungsaufwand weit höher ist. Auch für das Zugreifen auf die Metadatenschicht ist zuerst eine Definition der bestehenden Objekte nötig.

Von Vorteil ist, dass nur eine Webapplikation konfiguriert und gewartet werden muss, da sowohl die Tools als auch die Ausgabe-Clients auf die Pentaho-Webapplikation zugreifen. Wie bei den meisten anderen Report-Lösungen können auch mit Pentaho Berichte aus bestehenden operationalen Datenquellen generiert werden, ohne vorher Daten in andere Datenbanken exportieren zu müssen.

Es kommt häufig vor, dass bestehende Reporting-Lösungen ersetzt werden sollen. Da es keine Möglichkeit gibt, die Reportdefinitionen anderer Anbieter zu importieren, müssen alle Definitionen neu erstellt werden. Hier ist es von Vorteil, wenn die Berichtsdefinition lesbar ist, d.h. beispielsweise in XML geschrieben (wie auch bei Pentaho). Die Übernahme der Einstellungen und Datenquellen wird dadurch erheblich vereinfacht.

Um Zukunftssicherheit zu gewährleisten, ist es unumgänglich, dass die Reporting-Lösung offene Standards aufweist. Im Bereich Reporting ist Pentaho hier vorbildlich: Die komplette Pentaho Suite ist als JEE-Architektur erstellt und damit sowohl server- als auch clientseitig plattformunabhängig.

Dadurch lässt sich Pentaho leichter in bestehende Architekturen implementieren. Erweiterungen und individuelle Anpassungen der Applikation sind mit Java und JSP leicht durchführbar.

BIRT

Mit BIRT (Business Intelligence and Reporting Tools) bietet auch die Eclipse Foundation eine Komplettlösung zum Erstellen von Berichten an. Im Gegensatz zu Pentaho ist BIRT keine komplette BI-Suite, sondern konzentriert sich auf Report-Funktionalitäten, die in bestehende Applikationen integriert werden können. Das Besondere bei BIRT ist, dass es sich nicht nur als eigene Applikation verwenden lässt, sondern vielmehr durch die BIRT Report Engine API direkt in den Java-Code einer anderer Anwendungen integriert werden kann. Ein weiteres

Alleinstellungsmerkmal ist die Bereitstellung einer eigenen Ajax-basierten JEE-Webapplikation zur Visualisierung der Daten, der BIRT Report Viewer. Bei der Ausgabe der Auswertungen im Browser können so dynamische Komponenten wie Drill Downs und interaktive Parametrisierungen optisch ansprechend umgesetzt werden.

Die Anbindung an Datenquellen kann bei BIRT über XML-, JDBC- oder POJO/JavaBeans-Schnittstelle erfolgen. Falls BIRT in bestehende Java-Anwendungen integriert werden soll, ist es damit möglich, bereits von der Applikation geladene Daten mit in die Auswertungen zu integrieren. Das BIRT Report Designer Tool ermöglicht ähnlich wie beim Pentaho Designer die Erstellung von Report-Definitionen auf Eclipse-Basis mit einer grafischen Bedienoberfläche. Der Report Designer kann auch als Plug-in in eine bestehende Eclipse-Installation integriert werden. Einzelne Reportdefinitionen oder Charts können so mit einem bestehenden Eclipse-Projekt gespeichert werden. Gerade für Entwickler ist die Bedienung des Designers einfach, da der Umgang mit Eclipse schon bekannt ist. Zudem hilft die ausführliche Dokumentation.

Die Umsetzung einer BIRT-Reportlösung erfordert die Implementierung mindestens einer Webapplikation, in der die Report Engine Verwendung findet. Entscheidet man sich für die Integration von BIRT in bestehende Java-Anwendungen, muss man einen zusätzlichen Aufwand für ihre Anpassung berücksichtigen.

Wie auch Pentaho orientiert sich BIRT als Open-Source-Projekt an offenen Standards wie XML (für Reportdefinitionen), JDBC und JavaEE. Allerdings werden Standards wie MDX noch nicht berücksichtigt.

JasperReports

JasperSoft ist mit seiner Reportlösung JasperReports ein weiterer wichtiger Anbieter im Open-Source-Reporting-Segment. Wie Pentaho bietet JasperSoft auch eine komplette BI Suite mit hohem Funktionsumfang. Der Hauptunterschied zu Pentaho ist der hohe Anteil an Eigenentwicklungen. Nur wenige Komponenten wurden, wie bei Pentaho, dazugekauft.

Einer der besonderen Merkmale von JasperReports ist die Möglichkeit, lange Berichte seitenweise aufzubereiten. So können überlange Reportergebnisse optimal dargestellt werden. Eine weitere Besonderheit ist die hohe Anzahl an Ausgabeformaten, die zur Auswahl stehen. Neben den Standardformaten HTML und PDF, die auch von Pentaho und BIRT unterstützt werden, bietet JasperReports neben Excel auch eine Ausgabe im OpenOffice-Format (ODF).

Bei JasperReports besteht neben den konventionellen Arten der Datenanbindung zusätzlich die Möglichkeit, auf EJBs und die Daten einer Hibernate-Implementierung zurückzugreifen. Letzteres birgt durch die Nutzung des Hibernate Caches erhebliche Vorteile für die Leistungsfähigkeit der Anwendung. Das in der JasperSoft BI Suite enthaltene JasperStudio zum Erstellen von Berichten bietet ähnlich den BIRT und Pentaho Report Designern eine komfortable grafische Oberfläche zur Report-Gestaltung. Wie bei Pentaho werden hier die pixelgenaue Anordnung der Elemente sowie ein Query Designer zur Auswahl der Daten und eine Reportvorschau unterstützt. Ähnlich wie BIRT kann JasperReports als eigenständige Webanwendung laufen oder in eine bestehende Applikation integriert werden. Aus mehreren Gründen empfiehlt es sich aber, für die Reporting Engine eine eigene Webapplikation zu verwenden. Zum einen lässt sich eine Ausgliederung der Anwendung auf einen eigenen Server zur Vermeidung möglicher Leistungsengpässe leichter realisieren, zum anderen sinkt das Risiko, bei einem Release-Wechsel ein zeit- und kostenintensives Refactoring durchführen zu müssen.

