



Leseprobe

Stefan Müller, Christopher Keller

Pentaho und Jedox

Business Intelligence-Lösungen: Data Warehousing, Reporting, Analyse,
Planung

Herausgegeben von Sebastian Wenzky

ISBN (Buch): 978-3-446-43897-2

ISBN (E-Book): 978-3-446-44125-5

Weitere Informationen oder Bestellungen unter

<http://www.hanser-fachbuch.de/978-3-446-43897-2>

sowie im Buchhandel.

Inhalt

Vorwort	IX
1 Business Intelligence im Überblick	1
1.1 Business Intelligence: die Verwandlung von Daten in Informationen	1
1.2 Data Warehouse: das zentrale Datenlager des Unternehmens	4
1.3 Reporting und Dashboards: die maßgeschneiderte Verteilung von Informationen	11
1.4 Online Analytical Processing: mit Datenwürfeln das Business verstehen	14
1.5 Data Mining: unbekanntes Wissen in den Daten entdecken	20
1.6 Open Source trifft BI: vom Nischen- zum Mainstream-Player	21
1.7 Zusammenfassung	25
2 Pentaho: Business- und Big-Data-Analysen	27
2.1 Pentaho im Überblick	27
2.2 Warum Pentaho?	30
2.3 Community vs. Enterprise	32
2.4 Community oder Enterprise Edition im Unternehmenseinsatz	34
2.5 Zusammenfassung	35
3 Los geht's: Pentaho installieren und konfigurieren	37
3.1 Installation Pentaho BA-Server	37
3.2 Zugriff auf die User Console	43
3.3 Verwaltung der Lizenzen	46
3.4 Anlage von Datenverbindungen	46
3.5 Sicherheitseinstellungen	49
3.6 Einrichtung der E-Mail-Dienste	50
3.7 Zusammenfassung	50

4	Pentaho User Console	51
4.1	Überblick der Pentaho User Console	51
4.2	Reporting	56
4.3	Scheduling	58
4.4	Interactive Reporting	59
4.5	Analyzer Reports	62
4.6	Dashboards	73
4.7	Marketplace	77
4.8	Mobile	78
4.9	Zusammenfassung	82
5	Pentaho Data Integration	83
5.1	Pentaho Data Integration im Überblick	83
5.2	Aufbau eines Star-Schemas mit Pentaho Data Integration	92
5.3	Big Data und Pentaho Data Integration	114
5.4	Data Mining und Pentaho Data Integration	134
5.5	Zusammenfassung	136
6	Pentaho Reporting	137
6.1	Pentaho Report Designer	137
6.1.1	Überblick des Report Designers	138
6.2	Definition von Datenquellen und Abfragen im Report Designer	146
6.3	Entwicklung des ersten Berichts	150
6.4	Erweiterte Berichtsfunktionen am Beispiel eines MDX-Berichts	163
6.5	Berichtsverteilung mit Pentaho	171
6.6	Reportentwicklung für Big Data Stores	177
6.7	Pentaho Interactive Reporting	181
6.8	Metadaten für Reporting und Analysen	187
6.9	Zusammenfassung	193
7	Pentaho Analysis	195
7.1	Mondrian im Überblick	195
7.2	OLAP-Würfel mit der Pentaho Schema Workbench erstellen	199
7.3	Bessere Performance mit dem Aggregation Designer erreichen	213
7.4	Flexible Analysen für Big Data Analytics	219
7.5	Zusammenfassung	229
8	Pentaho Dashboards – von Self Service bis maßgeschneidert	231
8.1	Self Service Dashboards für Endanwender mit dem Pentaho Dashboard Designer	232
8.2	Maßgeschneiderte Dashboards mit den Pentaho CTools	241
8.3	Zusammenfassung	253

9	Jedox: Self Service BI und Performance Management	255
9.1	Jedox: Self Service BI und Performance Management	256
9.2	Versionsübersicht	259
10	Los geht's: Jedox installieren und konfigurieren	261
10.1	Installationsvarianten	261
10.2	Exemplarische Installation aller Komponenten	263
10.3	Konfiguration des Servers für den Zugriff externer Clients	269
11	Installationsverzeichnis & Ordnerstrukturen	273
12	Jedox for Excel	275
12.1	Add-in mit dem OLAP-Server verbinden	277
12.2	Der Dialog „Ansicht einfügen“	279
12.3	Datenfunktionen	284
12.4	Der Dialog „Elemente einfügen“	288
12.5	Entwicklung von Datenstrukturen – Modeller	292
12.5.1	Erstellen einer neuen Datenbank	293
12.5.2	Anlegen der Dimensionen	294
12.5.3	Erstellen eines Würfels	301
12.6	Dateneingabe	302
12.6.1	Splashing	302
12.6.2	Import-Assistent	304
13	Jedox Web	309
14	Jedox Web – Datei-Manager	313
14.1	Funktionsweise Datei-Manager	313
14.2	Funktionsweise Arbeitsmappe	315
14.3	Entwicklung von Berichten	319
14.3.1	Anlegen des Berichts (inkl. Überschriften)	321
14.3.2	Anlegen der Combo-Boxen	322
14.3.3	Anlegen des Datenbereichs	330
14.3.4	Platzieren des DynaRanges und des Datenbereichs	336
14.3.5	Einbinden des Diagramms	345
14.3.6	Einbinden der individuellen Bilder	349
14.4	Weitere Funktionen	349
14.4.1	Hilfreiche Zusatzfunktionen im Web	349
14.4.2	Verwendung von Variablen und benannten Bereichen	351
14.4.3	Macro-Engine	356

15 Berichts-Manager	359
15.1 Funktionsweise Berichts-Manager	359
15.2 Erweiterte Berichtserstellung	361
16 ETL-Manager	363
16.1 ETL-Menü	365
16.2 ETL-Komponenten	366
16.2.1 Verbindungen	366
16.2.2 Extraktion	368
16.2.3 Transformation	376
16.2.3.1 Transformationsschritt „Field Transform“	377
16.2.3.2 Transformationsschritt „Table Join“	381
16.2.3.3 Transformationsschritt „Table Transform“	382
16.2.3.4 Transformationsschritt „Table View“	387
16.2.3.5 Transformationsschritt „Tree FH“	387
16.2.3.6 Weitere Transformationen	390
16.2.4 Laden	391
16.2.5 Job	395
16.2.6 Variablen	397
16.3 Beispiel ETL-Prozess	400
17 Task- und OLAP-Manager	411
17.1 Task-Manager	411
17.1.1 Batch PDF	411
17.1.2 ETL Jobs	414
17.2 OLAP-Manager	415
18 System-Manager	417
18.1 Grundlegende Optionen	417
18.2 Benutzerberechtigungen	418
19 Jedox Mobile	425
20 Potenziale von Pentaho und Jedox zusammenführen	429
20.1 Pentaho und Jedox in der Gegenüberstellung	429
20.2 Pentaho und Jedox im Zusammenspiel	431
20.3 Zusammenfassung	432
Index	433

Vorwort

Beschäftigt man sich mit Business Intelligence und durchsucht das Internet, wird man schnell mit Phrasen wie „Daten sind das neue Öl“ oder „Wer die Daten hat, hat die Macht“ konfrontiert. Daten bzw. die enthaltenen Informationen werden immer wichtiger für Unternehmen. Wer seine Kunden am besten kennt oder am schnellsten auf geänderte Marktbedingungen reagieren kann, generiert Wettbewerbsvorteile und kann sich von seinen Marktbegleitern absetzen. Voraussetzung ist, die Daten schnell und zuverlässig zu erheben und auszuwerten. Nur so kann man die Schätze aus den eigenen Datenbanken und IT-Systemen heben.

Als Berater für Business Intelligence und Big Data führen wir entsprechende Lösungen seit mehreren Jahren in einer Vielzahl von Projekten ein. Als Software verwenden wir in diesen Projekten die Tools von Pentaho oder Jedox, um Daten zu extrahieren, zu veredeln und somit in wertvolle Informationen zu verwandeln. Dieses Buch stellt die beiden Lösungen vor. Nach einer kurzen allgemeinen Einführung in das Thema Business Intelligence wird im ersten Teil des Buches die Software von Pentaho vorgestellt. Pentaho bietet einen vollständigen Ansatz für Business Intelligence und Big Data Analytics. Der zweite Teil dieses Buches stellt die BI-Software von Jedox vor. Jedox ist ebenfalls eine vollständige BI-Plattform und bietet zudem die Möglichkeit, umfangreichen Planungsapplikationen zu entwickeln.

Neben der Vorstellung der beiden Software-Suiten sind in den jeweiligen Kapiteln auch praktische Beispiele hinterlegt.



www.downloads.hanser.de

Hier steht ein Teil der Beispiele zum Download bereit, entsprechende Hinweise sind in den Kapiteln hinterlegt. So können Sie direkt mit den Tools arbeiten und die Beispiele besser nachvollziehen.

Sowohl Pentaho als auch Jedox haben sich in unseren Projekten mehr als bewährt und die Anwender überzeugt. In diesem Sinne wünschen wir Ihnen viel Spaß beim Lesen und Ausprobieren – und vor allem viel Erfolg für Ihre eigenen Business Intelligence-Vorhaben.

Viele Grüße

Stefan Müller und Christopher Keller

Fulda, im Dezember 2014

2

Pentaho: Business- und Big-Data- Analysen

Pentaho ist der weltweit führende Anbieter von Open-Source-Business- und Big-Data-Analytics-Software. Pentaho ist eine flexible und offene Business-Intelligence-Plattform. Sie unterstützt Unternehmen dabei, die richtigen Entscheidungen auf operativer und strategischer Ebene zu treffen. Innerhalb der Plattform existieren Lösungen für Datenintegration, Analyse, Reporting, Data Mining und Mobile BI. Weiterhin existieren Schnittstellen zu allen gängigen Big Data Stores. Durch die enge Kopplung von den Datenintegrations- und Analysefunktionen können auch sehr große Datenmengen mit Pentaho und den neuen Big-Data-Technologien analysiert werden.

Kapitel 2 ist der Einstieg in den Bereich Pentaho dieses Buches. In Abschnitt 1 wird ein erster Überblick der Software mit seinen einzelnen Modulen gegeben. Weiterhin lesen Sie, warum Pentaho als Business-Intelligence-Software im Auswahlprozess zu berücksichtigen ist und wie sich die Community- von der Enterprise-Version abgrenzt. Schließlich wird noch das Geschäftsmodell der Enterprise-Version von Pentaho erläutert.

■ 2.1 Pentaho im Überblick

Das Unternehmen Pentaho wurde im Jahr 2004 von fünf Gründern ins Leben gerufen. Der Hauptsitz des Unternehmens ist in Orlando in den USA. Von Beginn an war es die Idee, eine Commercial-Open-Source-BI-Plattform als Alternative zu den bestehenden proprietären Systemen ins Leben zu rufen. Seit dieser Zeit hat Pentaho eine steile Entwicklung durchlaufen und offeriert eine moderne, integrierte Plattform für Business und Big Data Analytics. Pentaho beheimatet und verantwortet die populären Open-Source-BI-Projekte Kettle, Mondrian, JFreeReport und Weka. Pentaho verfolgt ein subscriptionsbasiertes Geschäftsmodell. Die Software von Pentaho wird in 185 Ländern von über 1000 Enterprise-Kunden genutzt. Berücksichtigt man die Zahl der Community-Editionen, ist die Zahl der Anwender sicherlich noch deutlich größer.

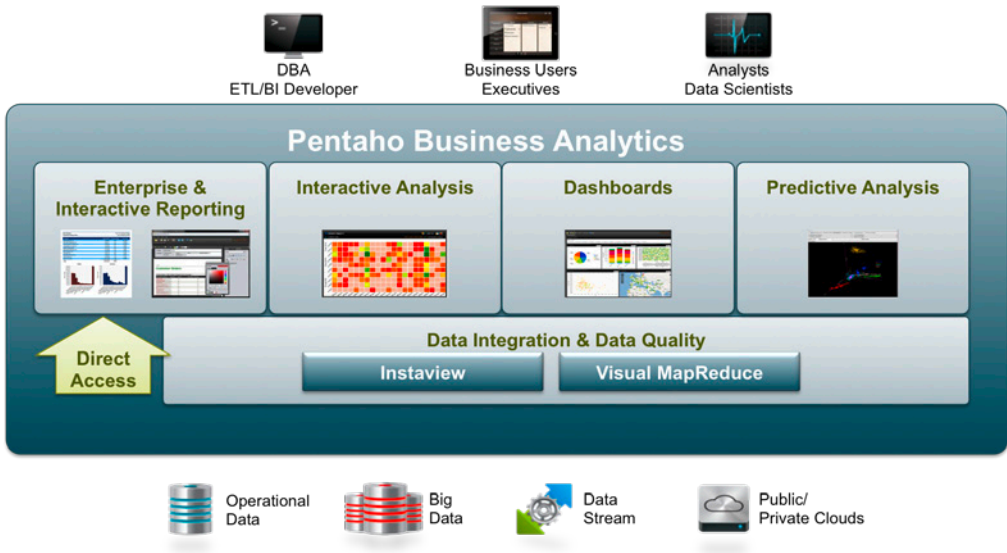


Bild 2.1 Übersicht Pentaho

Die Pentaho-Plattform ist höchst modular aufgebaut und erweiterbar. Nachstehend ist eine Übersicht der einzelnen Komponenten zu finden. Die folgenden Kapitel stellen die einzelnen Bausteine theoretisch und praktisch vor.

Pentaho Data Integration

Hinter Pentaho Data Integration (PDI) steht das Open-Source-Projekt Kettle. Mit PDI lassen sich Jobs und Transformationen für den Aufbau eines Data Warehouse entwickeln. Es existieren zahlreiche Schnittstellen zu den unterschiedlichen Datenformaten und Datenbanken. Neben den klassischen Datenquellen können mit PDI auch Informationen aus Big Data Stores ausgelesen und verarbeitet werden. PDI ist ein mächtiges Werkzeug für den Zugriff, die Aufbereitung und die Bereitstellung von Daten aus den unterschiedlichsten Quellen innerhalb einer BI-Architektur.

Pentaho Reporting

Pentaho bietet unterschiedliche Funktionen im Bereich Reporting von interaktiven „Self-Service“-Berichten bis hin zu hochformatierten Standardberichten. Mit Pentaho Interactive Reporting steht eine webbasierte Drag-and-Drop-Oberfläche für Endanwender für die Erstellung von Ad-hoc-Berichten zur Verfügung. Der Ursprung von Pentaho Reporting ist das Open-Source-Projekt JFreeReports. Mit dem Report Designer werden pixelgenaue Berichte erstellt und in unterschiedlichen Formaten an die Empfänger im Unternehmen verteilt.

Pentaho Analysis

Zur Bereitstellung von mehrdimensionalen Analysen nutzt Pentaho den open-source-basierten OLAP-Server Mondrian. Mondrian nutzt relationales OLAP und setzt somit auf ein Starschema in einer relationalen Datenbank auf. In Abhängigkeit der eingesetzten Edition

stehen unterschiedliche OLAP-Oberflächen zur Verfügung. Der Pentaho Analyzer ist Bestandteil der Enterprise Edition. Die webbasierte Drag-and-Drop-OLAP-Oberfläche ermöglicht die Navigation durch die Datenwürfel, die Visualisierung der Informationen sowie deren Export in unterschiedliche Formate. Die Community Edition nutzt im Standard JPivot als Oberfläche. Alternativ kann bspw. auch SAIKU aus dem dem Pentaho-Community-Umfeld genutzt werden.

Pentaho Dashboards

Auch bei der Erstellung von Dashboards gibt es Unterschiede hinsichtlich der genutzten Edition. Anwender der Enterprise Edition von Pentaho haben mit dem Dashboard Designer die Möglichkeit, Cockpits eigenständig in einer Oberfläche zusammenzustellen. Für die Erstellung der Dashboards befüllen die User Templates mit Tabellen, Charts, Reports oder OLAP-Ansichten. Bei Nutzung der Community Edition oder bei hohen Ansprüchen an die Gestaltung bleibt die Erstellung von Dashboards Entwicklern vorbehalten. Diese Dashboards lassen sich maßgeschneidert gestalten und werden mit den Open-Source-Erweiterungen der CTools erstellt.

Pentaho Predictive Analysis

Das Open-Source-Projekt Waikato Environment for Knowledge Analysis (Weka) kann für die Suche nach Mustern in den Datenbeständen genutzt werden. Weka stellt eine Vielzahl von Algorithmen für Data-Mining-Aufgaben bereit. Hierzu gehören bspw. Funktionen zur Regressions- oder Clusteranalyse. Basierend auf den entdeckten Mustern lassen sich zukünftige Trends aus den Daten ableiten.

Pentaho Business Analytics Server

Der Pentaho Business Analytics (BA) Server ist der Kern der Architektur und eine Eigenentwicklung des Herstellers. Der Server hostet zum einen die auf der Plattform erstellten Inhalte und zum anderen den von den Designwerkzeugen publizierten Content. Über den BA-Server werden Reports, OLAP-Analysen und Dashboards den Anwendern zur Verfügung gestellt. Weiterhin ermöglicht die BA-Plattform die zeitgesteuerte Erstellung von Berichten, die Verwaltung der Security sowie der genutzten Datenquellen.

Pentaho Mobile

Pentaho Mobile ist eine weitere Möglichkeit für den Zugriff auf die Inhalte des BA-Servers über ein Apple iPad. Es handelt sich um eine Web-App zum Abruf der Reports, aber auch für die interaktive Erstellung von OLAP-Analysen. Die Bedienung ist für die Nutzung mit einem Tablet-PC optimiert. Mit einer bestehenden Internetverbindung kann ein mobiler Zugriff von überall erfolgen.

Pentaho Open Source Plug-ins

Die Pentaho-Plattform ist durch ihren modularen Aufbau erweiterbar. Rund um das Unternehmen bzw. das Produkt Pentaho hat sich eine Community entwickelt, die derartige Erweiterungen unter Open-Source-Lizenz zur Verfügung stellt. Ab der Pentaho-Version 5 können diese Plug-ins über einen Marketplace per Knopfdruck installiert werden. Prominente Ver-

treter sind die CTools und Saiku. CTools steht für Community Tools und eine Sammlung von Plug-ins des Unternehmens Webdetails. Weit verbreitet für die Erstellung von Dashboards sind das Community Dashboard Framework (CDF), der Community Dashboard Editor (CDE) und die Community Charting Components (CCC). Saiku steht zum einen für einen OLAP-Client und zum anderen für ein Ad-hoc-Reportingwerkzeug.

■ 2.2 Warum Pentaho?

Der Markt für Business-Intelligence-Software wird von einer Vielzahl von Teilnehmern geprägt. Evaluiert man eine solche Software für den Einsatz im eigenen Unternehmen, wird man mit einem äußerst vielfältigen Angebot konfrontiert. Auf der einen Seite trifft man das stellenweise unüberschaubare Angebot der Megavendors, auf der anderen Seite gibt es die Anbieter mit hochspezialisierten Lösungen. Warum sollte man sich in diesem hoch differenzierten Markt für Pentaho als BI-Plattform im Unternehmen unterscheiden? Nachstehend finden Sie fünf Gründe, warum Pentaho bei der Evaluierung berücksichtigt werden sollte:

Integrierte BI-Plattform

Wo Pentaho drauf steht, ist auch Pentaho drin. Die vollständige Plattform stammt komplett von Pentaho bzw. besteht aus Bestandteilen unter der Kontrolle von Pentaho. Pentaho bietet eine umfassende Lösung für Ihre Business-Intelligence-Applikation aus einer Hand. Beide Aspekte sind keineswegs Standard bei derartiger Software und bringen eine Reihe von Vorteilen mit sich. Die Vollständigkeit des Angebots bezieht sich auf die Möglichkeit, die Architektur komplett mit Pentaho-Mitteln bewirtschaften zu können. Es gibt also nicht nur eine ETL-Komponente, sondern auch Werkzeuge für die Darstellung und Verteilung der Informationen. Man muss sich also keiner Drittsoftware bedienen, um einzelne Funktionsbereiche abdecken zu können. Dieses Angebot aus einer Hand kann bspw. bei Supportfällen von Vorteil sein. Die Verantwortung für das korrekte Zusammenspiel der Komponenten liegt nur bei Pentaho. Als Anwender hat man nur einen Ansprechpartner, auf welchen man bei Problemen zugehen kann. Weiterer Vorteil eines derart integrierten Ansatzes sind die minimalen Reibungsverluste an den Schnittstellen zwischen den einzelnen Komponenten. Als Anwender merkt man durch die Integration der einzelnen Werkzeuge an vielen Stellen, dass die gesamte Software von einem Hersteller stammt. Viele Prozesse lassen sich durchgängig gestalten, sodass etliche Abläufe erleichtert werden. Ein gutes Beispiel ist sicherlich Pentaho Data Integration, welches die Basis für eine Vielzahl an Funktionen innerhalb der Plattform ist. So lassen sich die aufbereiteten Daten zum Beispiel direkt innerhalb von PDI visualisieren und auswerten.

Einfache Erweiterbarkeit

Pentaho bietet eine offene, skalierbare Architektur für vielfältige Erweiterungen der BI-Lösungen. Pentaho ist eine moderne, zu 100% aus Java bestehende Plattform. Es werden offene Industriestandards genutzt, um die Software nahtlos in die eigenen Systeme integrieren zu können. Bestehende Security Frameworks können von dem BI-System ebenso

genutzt werden, wie die Anwendung durch offene APIs um Charts und Grafiken von Drittanbietern erweitert werden kann. Nicht nur Desktop-Anwendungen, sondern auch mobile Applikationen auf Smartphones und Tablet-PCs können um analytische Komponenten erweitert werden. Die Offenheit und Flexibilität von Pentaho ermöglicht die Entwicklung passgenauer analytischer Systeme im Unternehmen und stellt auch die Umsetzbarkeit künftiger Anforderungen an diese Lösungen sicher. Durch den Open-Source-Ansatz ist Pentaho weiterhin nicht nur eine hervorragende Option für Stand-alone-, sondern auch für eingebettete Business Intelligence. Pentaho lässt sich hierbei mit unterschiedlicher Tiefe in bestehende Applikationen integrieren. Vom Customizing der User-Oberfläche bis hin zur Einbettung des Source-Codes in die eigene Applikation sind viele Varianten denkbar.

Breites Spektrum an Analysemöglichkeiten

Die Anwender einer Business-Intelligence-Lösung haben höchst unterschiedliche Anforderungen an den Zugriff auf Informationen. Pentaho bietet vielfältige Frontends für die unterschiedlichen Anwendergruppen im Unternehmen. Mit Pentaho Reporting lässt sich ein Standardberichtswesen umsetzen. Mit diesen eher statischen Auswertungen wird die Mehrheit der Anwender automatisiert mit Informationen versorgt. Pentaho Analysis ist das Werkzeug für Poweruser. Pentaho arbeitet im OLAP-Umfeld mit unterschiedlichen relationalen Datenbanken zusammen. Unter anderem können auch spezielle analytische Datenbanken genutzt werden. Diese Gattung an Datenbanken nutzt verschiedene Technologien für die Beschleunigung der Abfragen gegen den OLAP-Server. Durch die Erhöhung der Abfrageperformance können mit Pentaho sehr große Datenmengen mit den leistungsstarken, anwenderfreundlichen Analyseoberflächen ausgewertet werden. Pentaho Analysis kann in diesen Szenarien sehr gut skalieren. Weiterhin kann die Analyseoberfläche durch beschriebene Erweiterbarkeit um Visualisierungen erweitert werden. Hierbei können bspw. auch Spezial-Charts für die Auswertung großer Datenmengen genutzt werden. Zum Beispiel können Heatmaps oder Chord-Diagramme Analysten dabei unterstützen, Muster und Auffälligkeiten zu entdecken.

Innovativität durch Open Source

Neben der Flexibilität und Erweiterbarkeit von Open Source profitiert Pentaho von der Innovativität von open-source-basierter Software. Pentaho verfügt über eine aktive Community, welche die Software testet und weiterentwickelt. Auch hier ist PDI wieder ein gutes Beispiel. Seit vielen Jahren wird das ETL-Werkzeug um Komponenten und Schnittstellen aus der Community erweitert. Anwender entwickeln spezielle Schnittstellen für das Auslesen verschiedener Datenquellen oder Komponenten für die Verarbeitung der Daten. Durch den Einsatz in den unterschiedlichsten Szenarien ergibt sich eine Vielzahl von Erweiterungen, die in die Software einfließen und von allen Anwendern genutzt werden kann. Auch in den übrigen Bereichen von Pentaho existieren interessante Open-Source-Erweiterungen des BI-Servers. Die CTools ermöglichen bspw. die Entwicklung hoch individueller Dashboards, sind Open Source und stehen zur freien Verfügung. Ein weiteres Beispiel für eine innovative Ergänzung des Standard-Pentaho ist der App Builder Sparkl. Hierbei handelt es sich um ein Instrument zur Entwicklung von Plug-ins basierend auf CDE und PDI. Ohne Programmierkenntnisse kann eine Vielzahl von Applikationen für die Pentaho-Plattform entwickelt werden, welche mit anderen Anwendern geteilt werden können. Für das Teilen

der Plug-ins und Erweiterungen hat Pentaho sowohl für PDI als auch für den BA-Server einen Marketplace eingerichtet. Dieser Marketplace erlaubt die Installation der Open-Source-Komponenten auf Knopfdruck und erleichtert deren Verteilung unter den Pentaho-Anwendern. Auf diese Weise kann Pentaho seine Stärken hinsichtlich Innovativität und Erweiterbarkeit voll ausspielen.

Business und Big Data Analytics mit einer Plattform

Ein weiterer Beleg für die Innovativität von Pentaho sind die Funktionalitäten im Bereich Big Data Analytics. Auf einer Plattform können die Daten aus den Big Data Stores extrahiert, verarbeitet und visualisiert werden. Es existieren Schnittstellen zu allen gängigen Hadoop-Distributionen, NoSQL- und analytischen Datenbanken. Mit Pentaho steht eine visuelle Entwicklungsumgebung für Big Data Analytics bereit. Somit können die Entwickler auch auf bestehendes Know-how aus dem klassischen BI-Umfeld zurückgreifen. Mit diesen Werkzeugen kann eine erhebliche Beschleunigung der Generierung von Wissen aus Daten erreicht werden. Auch die Verarbeitung der Daten an sich kann mit sehr guter Performance erfolgen. Die mit PDI erstellten ETL-Prozesse können innerhalb eines Hadoop-Clusters unter Nutzung der dort vorhandenen Performance ausgeführt werden. Mit Instaview steht ein leistungsstarkes Werkzeug für Data Scientists zur Verfügung. Durch die enge Kopplung von Datenzugriff und Analyse können Informationen aus den Big-Data-Quellen in einem Self-Service-Modus ausgewertet werden. Big Data Analytics ist ein wesentlicher Bestandteil der Ausrichtung von Pentaho, sodass ein weiterer Ausbau dieser Features auch zukünftig zu erwarten ist.

■ 2.3 Community vs. Enterprise

Pentaho unterscheidet zwischen der Enterprise Edition und der Community Edition der Software. Die CE ist frei verfügbar und kann unter <http://www.community.pentaho.com> bezogen werden. Rund um die Pentaho CE existiert eine aktive Community, welche die Software testet und weiterentwickelt. Die EE von Pentaho wird von dem kommerziellen Unternehmen entwickelt und qualitätsgesichert. Die Software kann als 30-Tage-Evaluierungsversion unter <http://www.pentaho.com> heruntergeladen werden. Nach dieser Zeit muss die Software durch Pentaho lizenziert werden.

Im Wesentlichen lassen sich die folgenden Unterschiede feststellen:

Kosten

Die Community-Version steht kostenfrei zu Verfügung. Die Software steht unter einer Open-Source-GPL-Lizenz und kann frei verwendet werden. Im Betrieb können unter Umständen im Vergleich zur Enterprise-Version höhere Kosten anfallen, da mehr interne Ressourcen benötigt werden, um die Lösungen zu supporten. Für die Pentaho EE ist eine jährliche Subskriptionsgebühr zu zahlen. Die Höhe der Gebühr ist unabhängig von der Anzahl der User und richtet sich nach genutzter Funktionalität und den verwendeten Server-Cores. Der Subskription-Ansatz bietet den Vorteil geringer Initialkosten. Es sind keine größeren Kos-

tenblöcke für einen Lizenzkauf zu Beginn eines Projektes zu tragen. Die interne Ressourcenbindung ist vergleichsweise sicherlich geringer, da zum einen eine Reihe von Komfortfunktionen existiert und bzgl. Support auf Pentaho zurückgegriffen werden kann.

Support

Support ist ein wesentliches Differenzierungsmerkmal zwischen CE und EE bei Pentaho. Als Anwender der CE ist man auf Foren und ähnliche Einrichtungen bei Problemen angewiesen. Kunden der Pentaho EE bekommen Support durch den Hersteller. Dieser Service ist sicherlich ein zentraler Aspekt bei der Entscheidung, eine Subskription zu kaufen. Treten Schwierigkeiten beim Betrieb der Pentaho-Lösung auf, kann der Hersteller-Support genutzt werden. Der Support unterstützt bei der Fehlersuche und hilft, die Problemlösung zu beschleunigen. Je nach Anforderungen sind unterschiedliche Supportpläne verfügbar. Des Weiteren steht EE-Kunden Zugriff auf spezielle Hilfeseiten und eine professionelle Dokumentation zur Verfügung.

Wartung

Auch die Wartung der Software ist Bestandteil der EE. Pentaho stellt Fixes zeitnah seinen Kunden bereit, sodass diese deutlich früher zur Verfügung stehen als in künftigen CE-Releases. Bei Nutzung der CE können Probleme natürlich auch selbst gefixt werden.

Erweiterte Funktionalität

Die Pentaho EE stellt gewisse Funktionen exklusiv seinen Anwendern bereit. Mit dem Analyzer und dem Dashboard Designer sind bereits Beispiele aufgeführt worden. Weiterhin bietet die EE einige Komfortfunktionen bei der Installation und dem Betrieb der Plattform. Anwender der CE müssen somit stellenweise auf Plug-ins ausweichen oder können einzelne Funktionen nicht nutzen.

Nachstehende Tabelle zeigt die funktionalen Unterschiede zwischen der Community und der Enterprise Edition von Pentaho.

Tabelle 2.1 Gegenüberstellung Community und Enterprise Edition

Funktionalität	Community Edition	Enterprise Edition
Automatisierte Installer	-	x
Pentaho Analyzer	-	x
Self Service Dashboard Designer	-	x
Hadoop-Integration für Big Data Reporting	-	x
Support Mobile BI	-	x
Shared Repository	-	x
Versionierung des Repository Contents	-	x
Repository Content Locking	-	x
Automatisches Ablaufen des Contents	-	x
Audit Reports	-	x

(Fortsetzung nächste Seite)

Tabelle 2.1 Gegenüberstellung Community und Enterprise Edition (*Fortsetzung*)

Funktionalität	Community Edition	Enterprise Edition
Anbindung an LDAP und Active Directory	–	x
Single Sign-on	–	x
Automatisiertes Scheduling von ETL-Prozessen	–	x
Zentralisierte Administration und Monitoring	–	x
Clustering	–	x
Job Restartability – PDI	–	x
Transaction Control – PDI	–	x
Connection Security – PDI	–	x
Data Federation – JDBC, Data Services	–	x

Pentaho ist auf einem „Open-Core“-Modell aufgebaut. Dies bedeutet, dass die Entwicklungen der Core-Engines der BI-Plattform alle in der Community-Version stattfinden. Funktionalitäten für den Einsatz im professionellen Unternehmensumfeld, wie bspw. Admin- oder Security-Features, werden auf dieser Basis entwickelt und stehen in der Enterprise-Version zur Verfügung.

■ 2.4 Community oder Enterprise Edition im Unternehmenseinsatz

Kosten sind sicherlich ein zentraler Faktor in jedem Projekt. Den Einsatz einer Software ohne die Zahlung von Lizenzkosten ist sicherlich für eine Vielzahl von Unternehmen sehr interessant.

Je nach Anwendungsfall und Kritikalität der geplanten Lösung sollten Sie aber sehr sorgfältig überlegen, ob der Einsatz von Pentaho ohne professionellen Support wirklich eine Option ist. Soll eine unternehmensweite Business-Intelligence-Plattform aufgebaut werden, existieren sicherlich zahlreiche Anforderungen an die Betriebssicherheit. Längere Ausfälle oder Fehler sind gewiss problematisch. Fachanwender nutzen das Business-Intelligence-System für ihre tägliche Arbeit und derartige Schwierigkeiten werden die Akzeptanz einer Lösung nicht gerade steigern. Folglich ist Zugriff auf einen professionellen Support bei einer produktiven Umgebung eigentlich unerlässlich, um Probleme schnell und zuverlässig lösen zu können. Dies gilt im Übrigen nicht nur für Open-Source-, sondern auch für Closed-Source-Software.

Kosten fallen nicht nur für Lizenzen, sondern auch für die Betreuung der Lösung an. Eine Pentaho CE wird im laufenden Betrieb sicherlich mehr interne Ressourcen und Know-how binden als eine EE. Aus Sicht eines IT-Leiters gilt es also zu überlegen, entsprechende Ressourcen für die Administration zu allokkieren oder die Funktionen der Enterprise Edition zu nutzen.

Insgesamt lässt sich also festhalten, dass für den Betrieb von Pentaho im produktiven Einsatz eine Enterprise Edition sicherlich empfehlenswert ist. Ein BI-System wie Pentaho ist sicherlich nicht in gleichem Maße mission-critical wie bspw. ein ERP-System. Nichtsdestotrotz werden auch BI-Systeme immer kritischer für den Unternehmenserfolg, sodass ein entsprechendes Umfeld geschaffen werden sollte. Die Community Edition ist sicherlich eine interessante Option für die Evaluierung der Software oder den Einsatz jenseits der Produktion. Voraussetzung für den Einsatz der CE ist sicherlich die Verfügbarkeit von entsprechend qualifiziertem Personal. Selbstverständlich ist die Community Edition auch für solche Unternehmen interessant, welche sonst aus Budgetgründen niemals eine Business-Intelligence-Software einführen würden.

■ 2.5 Zusammenfassung

Pentaho ist eine umfassende Plattform für Datenintegration, Business Intelligence und Big Data Analytics. Die Pentaho-Plattform bietet umfangreiche Funktionen zur Integration von Daten aus den unterschiedlichsten Quellen ebenso wie Werkzeuge für Berichtswesen, Analysen und Visualisierungen. Pentaho ist offen und erweiterbar und bietet somit eine ideale Ausgangslage zur Gestaltung passgenauer Business-Intelligence-Anwendungen, die Daten in wertvolle Informationen verwandeln. Pentaho lässt sich für die folgenden Themen einsetzen:

- **Datenintegration**
Extraktion, Verknüpfung und Aufbereitung von Daten aus den unterschiedlichsten Quellen, um diese für Business Intelligence und Big Data Analytics nutzbar zu machen
- **Business Intelligence**
Analysen und Berichte für eine zielgerichtete, faktenbasierte Entscheidungsfindung im Unternehmen
- **Big Data Analytics**
Analysen von großen, polystrukturierten Datenmengen zur Generierung von neuem Wissen
- **Integration in Drittapplikationen**
Nahtlose Integration analytischer Funktionen in Drittapplikationen

9

Jedox: Self Service BI und Performance Management

Jedox ist eine Business-Intelligence-Lösung zur systematischen Analyse von Unternehmensdaten, die über eigenständige Mechanismen berechnet und den Anwendern zur Verfügung gestellt werden können.

Jedox verfolgt dabei das Self-Service-BI-Prinzip, indem insbesondere die Planung, die Analyse sowie das Reporting im Vordergrund stehen und von den Anwendern individuell gestaltet werden können.

Planung

Erfolgreiche Unternehmensplanung basiert auf exakten und aussagekräftigen Prognosen. Jedoch erschwert die interne Aufgabenkomplexität der einzelnen Fachbereiche häufig den Abstimmungsprozess und mindert so die Effizienz des Planungszyklus. Mit den web- und Excel-basierten Planungstools von Jedox hingegen können Fachanwender Reports modellieren und selbst komplexe Planungskonstellationen selbstständig berechnen und analysieren.

Sämtliche Eingabedaten und qualitativen Ergebnisse werden zentral auf der Jedox In-Memory OLAP-Datenbank hinterlegt, sodass auch mehrere Fachbereiche koordiniert und ressourcensparend zusammenarbeiten. Besonders Unternehmen mit dezentralen Organisationsstrukturen profitieren von Jedox-Lösungen, denn transparente Planungsprozesse können schnell und einfach im System hinterlegt werden. So können Sie auch bei komplexen Planungsprozessen den Überblick behalten und Optimierungspotenziale auf einen Blick erkennen.

Analyse

Die Analyse komplexer Fragestellungen verlangt nach leistungsstarken Tools, mit denen Fachanwender das Beste aus ihren Daten holen können. Die Ad-hoc-Analyse von Jedox arbeitet dabei in einer anwenderfreundlichen Excel-Umgebung. Der Nutzer hat dort die Möglichkeit, unmittelbar in den Datenbereichen zu navigieren und auf alle Daten bis zur untersten Belegebene („Drill-through“) zuzugreifen. Diese Verbindung von hocheffizienter Jedox-Funktionalität und intuitiver Excel-Bedienoberfläche erlaubt es Anwendern aus allen Fachabteilungen, wichtige Entscheidungen in Echtzeit zu treffen. So können Sie etwa Abweichungen Ihrer Kostenarten je Kostenstelle betrachten und bei Bedarf auf einzelne Buchungsbelege abspringen. Mit dem Jedox Web Browser lassen sich zudem funktionale

und visuell ansprechende Analysepfade hinterlegen. Ein kurzer Klick auf eine Jahreskennzahl etwa öffnet in der effizienten Visualisierung von Jedox Web eine Monatsansicht im Wasserfallmodell. Neben der individuellen Ad-hoc-Analyse unterstützt Jedox wichtige Entscheidungsprozesse mit vordefinierten, standardisierten und dadurch besonders nutzerfreundlichen Analysepfaden. Analyseergebnisse und Kommentare können über alle Frontends (Excel, Web, Mobile) in die integrierte OLAP-Datenbank zurückgeschrieben und auf diese Weise konsistent mit den Daten abgelegt werden. So sind Ihre Geschäftsdaten mit Jedox immer auf dem aktuellen Stand.

Reporting

Fachanwender stehen heute verstärkt vor der Aufgabe, Kennzahlen und Berichte flexibel und tagesaktuell aufbereiten und anpassen zu müssen. Ständig wachsende, heterogene Datenmengen machen die manuelle Datenverarbeitung mit Excel jedoch beschwerlich und fehleranfällig. Jedox erleichtert diesen Reportingprozess und bietet Ihnen die Möglichkeit, individuell modellierbare Berichte in einer webbasierten, Excel-nahen Oberfläche zu erstellen. Standardisierte Reportings werden in der konsistenten und hochbeschleunigten Datenumgebung des Jedox OLAP-Servers besonders zeitsparend und sicher aufbereitet und können automatisch nach Zeitplan an definierbare Nutzergruppen verteilt werden. Neben Excel können Anwender auch über den Jedox Web Browser oder Jedox Mobile jederzeit von ihrem Smartphone oder Tablet auf die zentral hinterlegten Berichte („Single Point of Truth“) zugreifen und diese in Echtzeit anzeigen lassen. Dazu generieren Sie mit Jedox State-of-the-Art-Visualisierungen, mit dynamischen Grafiken oder nach dem Vorbild von SUCCESS-Charts.

Auch externe Datenquellen werden über die Jedox ETL-Funktion schnell und einfach in das Datenmodellgerüst eingebunden. Ganz im Sinne des „Self-Service“-Prinzips von Jedox können Fachanwender so Reports ohne Rückgriff auf die IT individuell anpassen oder Berichte in der gewohnten, Excel-affinen Oberfläche eigenständig erstellen. Überwinden Sie mit Jedox das Excel-Chaos und profitieren Sie von einem nutzerfreundlichen und transparenten Berichtswesen. Vgl. [Jedox]

■ 9.1 Jedox: Self Service BI und Performance Management

Jedox ist ein deutsches Unternehmen mit Hauptsitz in Freiburg im Breisgau. Das Unternehmen wurde 2002 von Kristian Raue gegründet und entwickelt seit dieser Zeit Software für Business Intelligence und Performance-Management. Zentrale Idee der Lösung von Jedox ist die Bereitstellung von „Self-Service“-BI. Fachanwender sollen also in der Lage sein, ohne tieferes IT-Know-how mit der Software zu arbeiten. Arbeiten Controller oder Business-Analysten bislang mit Excel, PowerPoint & Co, um Berichte und Analysen zu erstellen, bietet Jedox die Möglichkeit, bestehendes Wissen weiter zu nutzen und gleichzeitig die Vorteile eines OLAP-Systems anzuwenden. Durch die Kombination von Spreadsheets mit einem BI-System können mächtige, betriebswirtschaftliche Applikationen schnell und einfach ent-

wickelt werden. Durch die Rückschreibefunktion kann Jedox auch für die Unterstützung von Planungs- und Budgetierungsprozessen genutzt werden.

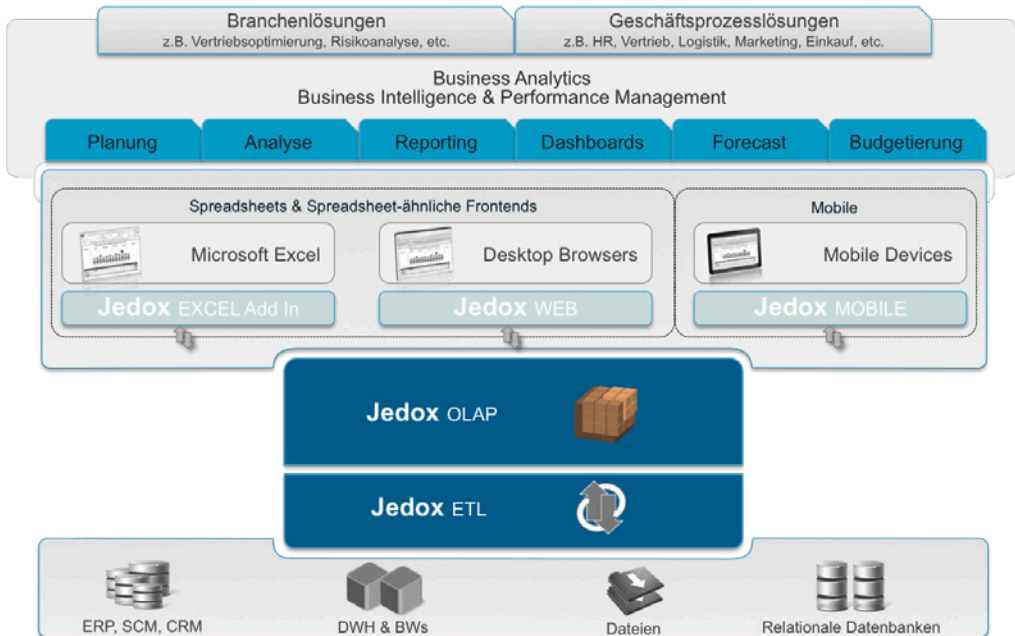


Bild 9.1 Übersicht Jedox

Nachstehend werden die Bausteine der Jedox-Architektur vorgestellt.

Jedox ETL

Jedox ETL ist ein Werkzeug zur Extraktion, Transformation und zum Laden von Daten. Die Bedienung und Modellierung der ETL-Prozesse erfolgt über eine Weboberfläche. Daten lassen sich aus unterschiedlichen Quellen und Formaten lesen. Beispielsweise existieren Schnittstellen zur Extraktion von Daten aus SAP ERP und BW. Nach dem Auslesen der Daten kann eine Reihe von Funktionen für die Aufbereitung der Daten genutzt werden. Schließlich können die transformierten Daten in die Würfel der OLAP-Datenbank oder auch in relationale Datenbanken sowie Dateien geladen werden. Nach Definition der ETL-Strecken können diese zeitgesteuert und automatisiert auf dem Server ausgeführt werden.

Jedox OLAP

Der Jedox nutzt multidimensionales OLAP (MOLAP). Der OLAP-Server ist eine multidimensionale Datenbank und speichert die Daten in Würfeln und Dimensionen. Durch die Nutzung von In-Memory-Technologien wird eine gute Performance erreicht. Die Möglichkeit zum Zurückschreiben von Werten in die OLAP-Würfel ist eine weitere Funktion dieses Ansatzes. Eingaben können hierbei nicht nur auf der Detailebene des Würfels, sondern auch in Aggregationen erfolgen. Durch das sogenannte „Splashing“ werden die erfassten Werte nach vorgegebenen Schlüsseln auf die zugrunde liegenden Hierarchieebenen verteilt. Die

eingeegebenen Werte können beispielsweise auch für Echtzeitberechnungen genutzt werden. Erfasst ein Anwender Werte, können diese in Echtzeit für Berechnungen auf der Datenbank genutzt werden. Auch auf die Modellierung der Würfel kann der Anwender bei Bedarf direkt aus der Excel-Oberfläche angepasst werden, um beispielweise neue Planelemente anlegen zu können. Ebenfalls im OLAP-Server erfolgt die Administration der Berechtigungen für die Würfel.

Jedox Excel-Add-in

Excel und Jedox OLAP sind eine altbewährte Kombination. Von Beginn an setzte Jedox auf die Tabellenkalkulation als Frontend für seine BI-Suite. Fachanwender können in vertrauter Umgebung arbeiten und auf den OLAP-Server zugreifen. Zum einen können die Anwender die Würfel nach Belieben kombinieren und filtern. Auf diese Weise lassen sich ohne IT-Kenntnisse Ad-hoc-Berichte und Analysen erstellen. Neben diesen dynamischen Ansichten lassen sich mit dem Excel-Add-in auch formatierte Berichte erstellen. Die Daten werden auch an diesem Szenario aus der OLAP-Datenbank abgefragt. Die Formatierung, Grafiken usw. werden mit den Funktionen von Excel erstellt. Ebenso wie Excel arbeitet Jedox zellbezogen. Durch die Zellorientierung kann die volle Flexibilität von Excel für die Darstellung der Berichte genutzt werden und auch bestehende Excel-Dateien können – die entsprechende Datenqualität vorausgesetzt – schnell an einen OLAP-Würfel angebunden werden. Entsprechend berechnete Anwender können auch administrative Funktionen, wie bspw. die Anlage neuer Dimensionen und Würfel, über das Excel-Add-in vornehmen.

Jedox Web

Mit Jedox Web steht eine weitere Oberfläche für die BI-Suite zur Verfügung. Jedox ist vollständig webfähig und kann somit ohne jegliche Installation auf Clientseite betrieben werden. Zentrale Komponente ist Jedox Spreadsheet. Hierbei handelt es sich um eine an Excel angelehnte Oberfläche im Webbrowser. Mit Jedox Spreadsheet wird ein analoger Ansatz wie beim Excel-Add-in verfolgt. Mit Office-Know-how lassen sich auch im Browser schnell und einfach Analysen und Berichte erstellen. Mit dem Jedox Analyzer kann eine weitere Oberfläche für Ad-hoc-Abfragen ähnlich einer Pivottabelle genutzt werden. Ebenfalls Bestandteil von Jedox Web ist der ETL-Server und der Task-Manager für alle zeitgesteuerten Funktionen auf der Plattform. Mittels des OLAP-Managers kann die Datenbank administriert werden. Der System-Manager ermöglicht die Verwaltung der User und der Lizenzen.

Jedox Mobile

Für den mobilen Zugriff stehen bei Jedox mehrere Optionen zur Verfügung. Zum einen kann mit dem Browser des mobilen Endgerätes mittels einer Web-App zugegriffen werden. Über eine optimierte Benutzerführung können die Informationen mit Jedox Mobile Touch abgerufen und auch Daten eingegeben werden. Zum anderen stehen für iOS und Android native Apps zur Verfügung, welche aus dem Appstore bzw. Marketplace geladen werden können. Die Apps bietet eine eigene Oberfläche für die Erstellung von Analysen oder Dashboards.

■ 9.2 Versionsübersicht

Wie bereits in der Marktübersicht geschildert, ist gegenwärtig keine vollständige Open-Source-Version von Jedox verfügbar. Lediglich der OLAP-Server steht unter einer Open-Source-Lizenz. Die zuvor verfügbare Community Edition war unter dem Namen Jedox Palo verfügbar. Die Version 3.2 ist die letzte verfügbare Open-Source-Version von Jedox und beinhaltet alle aufgeführten Komponenten mit Ausnahme von Jedox Mobile. Allerdings ist die Community- auch bei Jedox funktional von der Enterprise-Version abgegrenzt. Beispielsweise nutzt die Community Edition (Jedox Base) nur einen Server-Core, während die Enterprise Edition (Jedox Premium) die Last auf mehrere Server-Cores zur Steigerung der Performance verteilen kann.

Nachfolgend sind alle Komponenten dargestellt, die in den beiden Versionen zur Verfügung stehen:

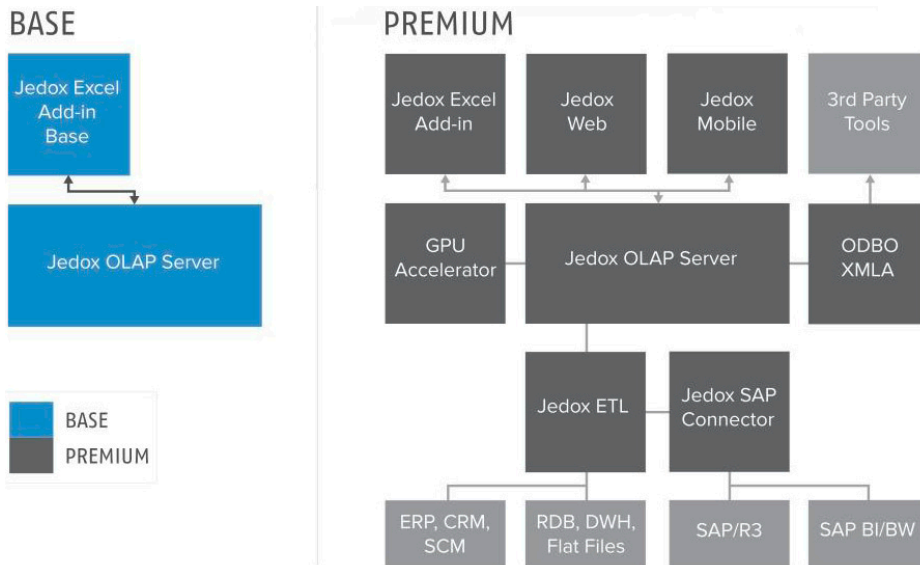


Bild 9.2 Abgrenzung der Jedox-Versionen

Während in der Basisversion nur die grundlegenden Komponenten verfügbar sind, die zur Erstellung und Bearbeitung von Datenmodellen sowie zur Visualisierung benötigt werden, kann in der Premium-Version aus einer ganzen Reihe an Modulen gewählt werden, die beispielsweise die automatisierte Beladung der Datenbank aus Vorksystemen ermöglichen. Auch die Berichtsgestaltung und -verteilung über das Web-Frontend sind dabei Features der Premium-Version.

Die Entscheidung, welche von den beiden Versionen im Unternehmen eingesetzt wird, hängt also stark davon ab, welche Aufgaben mit dem gewünschten BI-System erfüllt werden sollen und welche Komponenten benötigt werden. Dabei spielt vor allem der Blick auf die Quelldaten eine Rolle, die nur dann automatisiert angebunden werden können, wenn eine Premium-Version zum Einsatz kommt.

Support

Jedox bietet seinen Nutzern einen ausgiebigen Support, der über das Jedox-Support-Portal kontaktiert werden kann. Die Nutzer der Premium-Version können diesen Dienst in Anspruch nehmen.

Die Community Edition (bis Release 3.2) besitzt keine Unterstützung seitens Jedox.

Neben dem Support-Portal stehen noch weitere Foren und Hilfestellungen wie z. B. Handbücher oder verschiedene Beispieldateien zur Verfügung, die bei Fragen und Problemen im täglichen Einsatz verwendet werden können.

Wartung

Auch die Wartung der Software ist Bestandteil der Jedox-Versionen. Neben dem Major Release, welches einmal im Jahr erscheint, stellt Jedox im Laufe dieser Periode noch weitere Service-Releases zur Verfügung, die Fehlerbehebungen beinhalten, jedoch auch sehr häufig mit neuen Funktionen bestückt sind.

Index

A

Ad-hoc-Berichte 12
Aggregation Designer 213
Analyzer 62, 207

B

bedingte Formatierung 165
Benutzerdefinierte Kennzahlen 68
Berichts-Manager 359
Big Data 8, 114, 177, 219
Business Intelligence 2

C

Carte-Server 222
CDA 252
CDE 242
Community Edition 32
Content-Linking 76, 240
CTools 29, 76, 241

D

Dashboard Designer 29, 232
Dashboards 13, 73
Data Blending 221
Data-Driven Modelling 276
Data Integration 28, 83, 86
Data Marts 6
Data-Mining 20, 29, 134
Data Warehouse 3, 4
der Enterprise Edition 32
Drill-down 17
DynaRange 336

E

ETL 3, 7
ETL-Prozesse 364

H

Hadoop 116
Hops 90

I

Import-Assistent 277
Instaview 86, 131
Interactive Reporting 28, 59, 181, 192

J

Jedox ETL 257
Jedox Excel-Add-in 258, 261, 265, 275
Jedox OLAP 257
Jedox Web 258, 262, 277, 309, 313
Job Entries 91
Jobs 90

K

Kettle 28

L

Lizenzen 46

M

Macro-Engine 356
MapReduce 117
Marketplace 77

MDX 163, 197
Metadata Editor 187
Mobile 29
Mobile BI 78
Modeller 276
Mondrian 28, 195
MongoDB 126, 177, 226

O

OLAP 4, 14, 164, 195, 256f., 363

P

Planung 255

R

Report Designer 28, 138
Report-Wizard 177
Reporting 28, 56

S

Saiku 72
Schedules 58, 89

Scheduling 113, 171
Schema Workbench 199
Self-Service-BI 255
Slice and Dice 18
Splashing 302
Spoon 83
Standardberichte 11
Star-Schema 93, 197
Steps 90
System-Manager 417
Systemvoraussetzungen 37

T

Task-Manager 411
Thin Kettle JDBC Driver 222
Transformationen 90

U

User Console 43, 51

W

Weka 29, 134