

BARC Software-Evaluation

## Data Warehousing und Datenintegration

Studienauszug

IBM InfoSphere Warehouse und IBM InfoSphere Information Server



In Kooperation mit:

Business eKnowledge  
Solutions GmbH **BeKS**



## Autoren

### Jacqueline Bloemen

Jacqueline Bloemen ist seit 2003 Geschäftsführerin der BeKS (Business eKnowledge Solutions - [www.beks-consulting.com](http://www.beks-consulting.com)) sowie seit 2005 Senior-Analystin Data Warehousing und Business Intelligence bei BARC. Im Rahmen von BI und DWH Projekten unterstützt sie als Beraterin seit über 20 Jahren Unternehmen verschiedener Branchen und Größen bei der SW Strategie- und Architekturentwicklung. Ihr besonderer Fokus liegt auf komplexe Datenmanagementaufgaben, wo vor allem große Datenvolumina eine große Rolle spielen.



### Timm Grosser

Dipl.-Kfm. Timm Grosser ist Analyst am Business Application Research Center (BARC) im Bereich Business Intelligence. Seine Spezialgebiete sind entscheidungsunterstützende Informationssysteme mit besonderem Fokus auf die Bereiche Datenintegration, Datenqualität und Data Warehousing.



Dieses Dokument ist ein Auszug aus der BARC-Studie „Data Warehousing und Datenintegration“. Die Studie wurde völlig unabhängig vom neutralen Marktanalysten BARC erstellt. Weitere Informationen zur Studie sind unter <http://www.barc.de/dw> erhältlich.

## Vorwort

Gerade in wirtschaftlich unsicheren Zeiten spielt Business Intelligence eine entscheidende Rolle, die Wettbewerbskraft von Unternehmen zu steigern. Transparenz über das Geschehen im Unternehmen und seinem Umfeld, flexible Forecasting-Möglichkeiten und die Erfüllung gesetzlicher Auflagen (Compliance) gewährleisten die Steuerungsmöglichkeit von Unternehmen auch in turbulenten Zeiten. Business Intelligence und Data Warehousing bleiben damit wesentliche Erfolgsfaktoren für die Steuerbarkeit und erfolgreiche Ausrichtung von Unternehmen. Nicht umsonst ist Business Intelligence somit auch seit vier Jahren in Folge laut Befragungen die Top-Priorität des CIO.

Ausdruck der gestiegenen Bedeutung von Data Warehousing und Business Intelligence sind auch die deutlich verstärkten Anstrengungen, diese Themen sowohl in der Geschäfts- als auch der IT-Strategie zu verankern. Immer deutlicher wird, dass die analytische Informationsverarbeitung zu komparativen Wettbewerbsvorteilen führt und weniger die operative Transaktionsabwicklung auf Basis von Standardanwendungssoftware.

Grundlage jedes zuverlässigen Informationssystems sind seine Daten, und gerade Business Intelligence-Systeme benötigen eine konsolidierte und qualitätsgesicherte Datenbasis, die auch große Datenmengen schnell für analytische Zwecke verfügbar machen kann. Genau an dieser Stelle setzen Datenintegrations- und Data-Warehouse-Lösungen an, die in dieser Studie untersucht werden.

Nach der Konsolidierungswelle im Business-Intelligence-Markt haben die großen Anbieter IBM, Oracle, SAP und Microsoft inzwischen ein vertikal integriertes Portfolio geschaffen, das sowohl die Datenversorgungskette, Data-Warehouse-Plattformen als auch Auswertungswerkzeuge aller Couleur beinhaltet. Doch auch innovative und kleine Anbieter finden noch ihre Nische auf allen Ebenen der Systemarchitektur, in der sie eine ergänzende oder komplementäre Lösung bieten. Die Entwicklung bleibt dynamisch und BARC-Studien helfen Ihnen, aus neutraler Sicht die Übersicht zu behalten.

*Dr. Carsten Bange*



Dr. Carsten Bange  
Geschäftsführer BARC

## Inhaltsverzeichnis

IBM Information Management Produkte in der BARC Software-Evaluation „Data Warehousing und Datenintegration“ .....	4
IBM InfoSphere Warehouse V9.5.1.....	6
IBM InfoSphere Information Server 8.1 .....	13
Bewertung .....	21
IBM InfoSphere Warehouse .....	23
IBM InfoSphere Information Server .....	25

## IBM Information Management Produkte in der BARC Software-Evaluation „Data Warehousing und Datenintegration“

Die Software-Evaluation Data Warehousing und Datenintegration vergleicht und bewertet die führenden Werkzeuge für Data Warehousing und Datenintegration im deutschen Markt. Grundlage der Bewertungen ist ein umfassender Kriterienkatalog. Dieser Kriterienkatalog basiert auf den Erfahrungen der BARC-Mitarbeiter aus über zehn Jahren Softwaretests sowie Softwareauswahl- und Beratungsprojekten.

Derzeit verfügbar sind folgende Produkte:

Data-Warehouse-Plattformen	
IBM	InfoSphere Warehouse V9.5.1
Oracle	11g R1 Database und OWB
SAP	NetWeaver BI 7
SAS	SPDS / Enterprise Data Integration Server 9.1
Teradata	Enterprise Data Warehousing / Database 12.0

Datenintegrationswerkzeuge	
Business Objects	Data Integrator 11.5
IBM	InfoSphere Information Server 8.1
Informatica	Powercenter 8.5
Syncsort	DMExpress 4.1

Dieser Auszug aus der BARC Software Evaluation Data Warehousing und Datenintegration umfasst die Beschreibungen und Bewertungen der Produkte:

**IBM InfoSphere Warehouse** als Datenspeicher- und Werkzeuglösung zur Verwaltung analytischer Unternehmensdaten und

**IBM InfoSphere Information Server** als Datenintegrations- und Datenqualitätsplattform.

Neben diesen beiden Kernangeboten für Data Warehousing und Datenintegration werden von der IBM Software Group unter der Marke Information Management die Themen Data Management, Enterprise Content Management, Master Data Management, Business Intelligence und Express Middleware adressiert, die in dieser Studie nicht betrachtet worden sind. IBM Information Management ist mit weltweit rund 60 Millionen Benutzern und 645.000 Kunden der größte Bereich der IBM Software Group. Lösungsentwicklung und -vermarktung richtet sich an der „Information On Demand“ Strategie aus. Als wesentliches Ziel gilt hier auf der vorhandenen Informationsinfrastruktur in Unternehmen aufzubauen und die vorhandenen Informationsressourcen durch Integration und Virtualisierung als Service im Rahmen einer serviceorientierten Architektur (SOA) verfügbar zu machen. Hierfür sieht die IBM in ihrer SOA-Referenzarchitektur entsprechende informationszentrierte Service-

typen vor mit denen Informationen aus ihren Anwendungssilos herausgelöst und unternehmensweit als strategische Ressource verwaltet und zur Verfügung gestellt werden können. In diesem Zuge wurde 2008 die „Information Agenda“ vorgestellt, die Hardware, Software und Services bündelt, um eine prozessübergreifende und nicht applikationsorientierten Sicht auf das Unternehmen und seine Informationen zu ermöglichen. Als Teil der Information Agenda finden auch die Foundation Tools in den bewerteten Produkten der Studie Berücksichtigung. Hierunter werden Werkzeuge aus dem IBM-Portfolio verstanden, die Kunden mit einer heterogenen, auch nicht-IBM Infrastruktur helfen sollen, aus Informationen einen strategischen Vermögenswert zu machen. Dabei handelt es sich um die Komponenten:

- InfoSphere Information Analyzer,
- InfoSphere Business Glossary,
- InfoSphere Metadata Workbench,
- InfoSphere Information Server FastTrack,
- InfoSphere Data Architect,

die auch unabhängig vom InfoSphere Information Server als Erweiterung des InfoSphere Warehouse Pakets oder auch anderer Lösungen angeboten werden.

## IBM InfoSphere Warehouse V9.5.1

Mit InfoSphere Warehouse (IW) bietet IBM eine Werkzeugplattform zur Unterstützung des Data Warehousings auf Basis der eigenen relationalen Datenbank DB2 für Linux, UNIX, und Windows. Das Softwarepaket unterstützt Business Analysten, Modellierer, Entwickler und Administratoren mit folgenden Komponenten:

- **DB2 9.5 for Linux, Unix, Windows**  
Die Datenbankengine ist ein hybrider Datenserver, der relationales, dimensionales und XML-Datenmanagement bietet. InfoSphere Warehouse enthält die DB2 Enterprise Server Edition sowie weitere, für das Data Warehousing wichtige Funktionen wie Table Partitioning, Datenkomprimierung, Multidimensional Clustering and Materialized Query Tables (MQTs).
- **DB2 Workload Management (WLM)**  
Der Nachfolger von Query Patroller ermöglicht eine Echt-Zeit-Steuerung und Überwachung des Ressourcenverbrauchs in Data-Warehouse-Umgebungen. Es ermöglicht damit den gleichzeitigen Betrieb unterschiedlicher Lastprofile in einem System und verhindert, dass Ausreißer-Abfragen das System für andere Anwender blockieren.
- **DB2 Database Partitioning Feature (DPF)**  
DPF erlaubt die Partitionierung einer Datenbank innerhalb eines einzigen oder über einen Cluster mit mehreren Servern hinweg, um insbesondere große Datenbanken und komplexe Nutzlast zu unterstützen.
- **Design Studio**  
Die Eclipse-basierte Entwicklungsumgebung liefert Datenmodellierung für das Enterprise Data Warehouse (EDW) sowie für dimensionale Data Marts mittels integrierter InfoSphere Data Architect Komponenten sowie Entwicklung von Datenintegrations-Prozessen (Data & Control Flows inklusive Mining Flows) mittels des SQL Warehousing Tools (SQW).
- **OLAP Optimization**  
Cubing Services ermöglichen einen optimierten OLAP-Zugriff auf relationale DB2 Daten mittels etablierter Standards (OLE DB for OLAP, XMLA und MDX) und liefern einen zuschaltbaren Cache für Abfrageergebnisse. Cube Views unterstützen die Gestaltung von und Performance-Tuning für dimensionale Datenstrukturen zur Unterstützung von sowohl SQL- als auch MDX-basierten Abfragen.
- **Analytics: Mining**  
Die in die DB2 Datenbank integrierten Intelligent Miner Funktionen unterstützen den Business Analyst mit „Easy Mining“ Algorithmen bei Design und Erzeugung von Mining Modellen für strukturierte Daten, sowie bei Echtzeit-Scoring und Visualisierung der Mining Ergebnisse. Neu in V9.5 ist die Integration von Text-Mining-Fähigkeiten, die es ermöglichen, auch unstrukturierte Daten bei klassischen Analysen zu integrieren



- **Analytics: Alphablox (ABX)**  
Alphablox ist eine Plattform zur Entwicklung von analytischen J2EE-Web-Applikationen sowie J2EE-basierten, analytischen Komponenten, die in anderen Applikationen eingebettet werden können. Mit V9.5 erhielt AlphaBlox eine neu gestaltete, grafische Entwicklungsoberfläche, die es auch nicht-Java-Experten ermöglicht, mit Alphablox zu entwickeln.
- **Integrated Solutions Console**  
Integrierte, Web-basierte Konsole zur Verwaltung, Ausführung und Überwachung von Datenintegrationsprozessen, ABX-Frontend-Applikationen sowie Cubing Services.

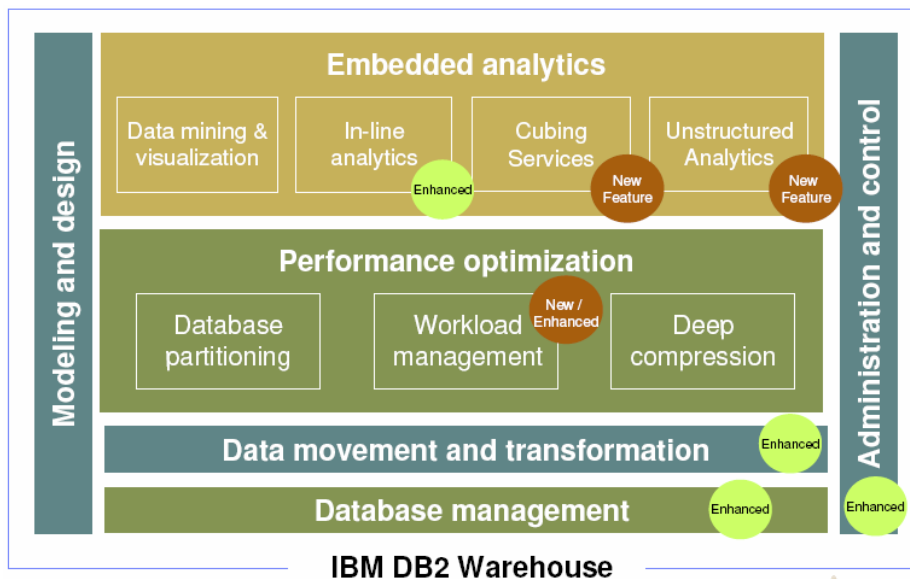


Abbildung 1: Überblick IW und der Erweiterungen in V9.5

Während der InfoSphere Information Server (IS) vor allem auf heterogene, serviceorientierte Informationsintegration für operative und dispositive Anwendungen fokussiert ist, bedient InfoSphere Warehouse vor allem Datenintegration (DI) und Datenmanagement rund um DB2-basierte Data Warehouses und Data Marts.

Besondere Stärken zeigt das Paket beim Aufbau von dimensionalen Data Marts für analytische Anwendungen, da es über umfangreiche Funktionen im Bereich Modellierung, Performance-Optimierung und Metadatenaustausch verfügt. Mit Hilfe von Alphablox können analytischen Applikationen entwickelt werden, die sich leicht in Portale und Java-Anwendungen einbetten lassen. Für weitergehende Bedürfnisse in diesem Bereich ermöglicht eine mitgelieferte Testversion von IBM Cognos 8 BI (Cognos 8 BI Starter Edition) ein Prüfen dieser Plattform auf Eignung zur Erfüllung der Anforderungen.

InfoSphere Warehouse lässt sich durch weitere Angebote aus dem IBM Information Management Portfolio sinnvoll ergänzen. Hier ist zunächst das Balanced-Warehousing-Angebot der IBM zu nennen. Mit dem Balanced Warehouse bietet IBM eine vorinstallierte Software/Hardware-Kombination (Appliance), die für den Einsatz spezifischer Data-Warehouse-Anforderungen optimiert ist. Hier



unterscheidet IBM verschiedene Größenklassen. Darüber hinaus bietet IBM die Möglichkeit, bestimmte Softwarekomponenten aus dem Information Server Portfolio auch unabhängig als Foundation Tools in Kombination mit InfoSphere Warehouse einzusetzen. So lässt sich das IW-Angebot in den Bereichen Metadaten Management (Metadata Workbench und Business Glossary), Datenqualität (Information Analyzer) und DI-Konzeption (FastTrack) verstärken. Eine weitere mögliche Ergänzung findet sich bei den IBM Industry Data Models sowie der IBM Cognos Performance Management Blueprints.

Herzstück des InfoSphere Warehouse ist die hybride DB2 Datenbank, die relationales, dimensionales und semistrukturiertes (XML) Datenmanagement unterstützt. Besondere Stärken für das Warehousing hat DB2 durch seine Shared-Nothing Architektur, die besonders gut zur Unterstützung von Scale-Out Maßnahmen auf Basis eines Clusters, bestehend aus mehreren Servern, geeignet ist. Hierfür bedient sich DB2, wie viele der DBMS-Mitbewerber auch, einer Datenbank-Partitionierung. Daten werden mittels eines Hash-Algorithmus auf alle Knoten im Cluster gleichmäßig verteilt, um so eine größtmögliche Parallelisierung von Anfragen zu ermöglichen. Dies beschränkt sich nicht nur auf lesende Abfragen. Datenintegrationsprozesse, die mit dem Datenintegrationswerkzeug SQW entwickelt werden, bedienen sich der Parallelisierbarkeit von schreibenden Operationen in DB2, um möglichst performant und skalierbar ablaufen zu können. Hierin besteht ein signifikanter Unterschied zu der technischen Architektur vom Information Server. Während der IS einem ETL (Extract, Transform, Load)-Prinzip mit einer eigenen Engine folgt, bezeichnet man die IW-Lösung als ELT (Extract, Load, Transform). Quelldaten werden dabei zunächst in die Datenbank geladen und Transformationen mit Hilfe der zugrunde liegenden DBMS-Funktionalität vor allem SQL-basiert umgesetzt.

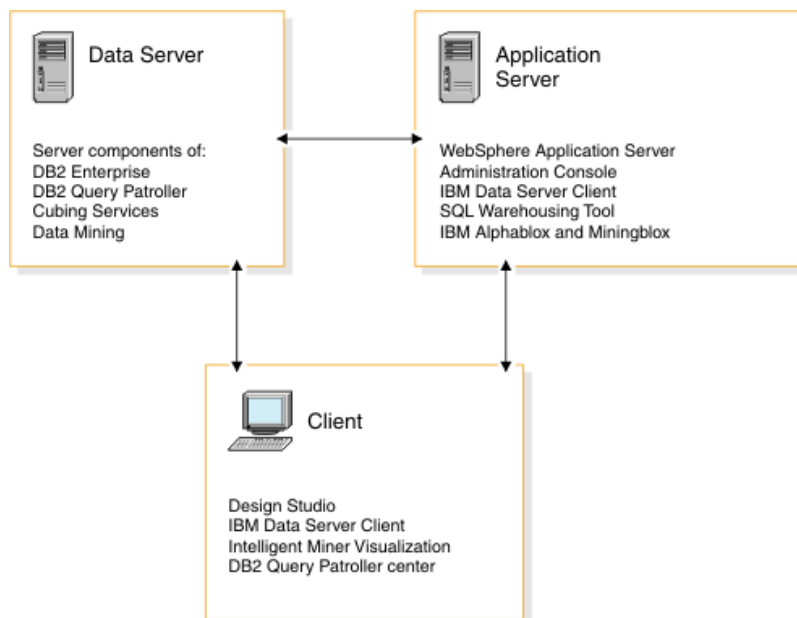


Abbildung 2: Entwicklungs- und Ablaufumgebungen des IW

Die Ablaufumgebung der mit SQW entwickelten Daten- und Kontrollflüsse ist der WebSphere Application Server (WAS). Eine eingeschränkte Lizenz von WAS wird daher mit IW verpackt. Die Auslieferung an den WAS erfolgt mittels Paketen, die technisch auf sogenannten Enterprise Graphs (EPG)

basieren. Diese werden ähnlich einem klassischen Java Programm unter Kontrolle der IW-Administrationskonsole in der WAS-Umgebung ausgeführt. EPGs beschreiben den Ablauf verschiedener Code-Einheiten sowie den Typ dieser Einheiten (JDBC, DB2® SQL SCRIPT, Java™ Methode, Kommando) und setzen so Daten-, Kontroll und Miningflüsse in der Ablaufumgebung um. Die Browserbasierte Administrationskonsole ermöglicht dabei eine aktive Überwachung laufender Prozesse und liefert Laufzeitstatistiken sowie etwaige Fehlerprotokolle. Es werden außerdem diverse Funktionen rund um Administration von weiteren IW-Komponenten wie Alphablox-Applikationen und Cubing Services angeboten.

Betrachtet man das allgemeine Metadaten-Management, so lässt IW das Konzept eines übergreifenden Metadaten Repositories vermissen. Metadaten werden teils als XML-Dateien im Dateisystem (Datenintegration), teils als Bestandteil des DB2-Katalogs (Datenmodelle, Cubes) abgelegt. Ein dediziertes Metadatenaustauschkonzept ist lediglich im Bereich der Datenmodellierung vorgesehen. Es sind Konnektoren verfügbar, um Daten einerseits aus gängigen Modellierungswerkzeugen zu importieren, andererseits erstellte dimensionale Modelle an gängige BI-Werkzeuge weiter zu reichen. Impact-Analysen auf vorhandene Flüsse sind innerhalb des Design Studios möglich. Bei Bedarf lässt sich IW hier erweitern. Zur Unterstützung einer Versionierung können entsprechend dem Eclipse-Standard Quellcodeverwaltungssysteme in das Design Center eingebunden werden, beispielsweise aus dem Hause IBM mit Rational ClearCase oder auf Basis der Open-Source-Lösung CVS (Concurrent Versions System).

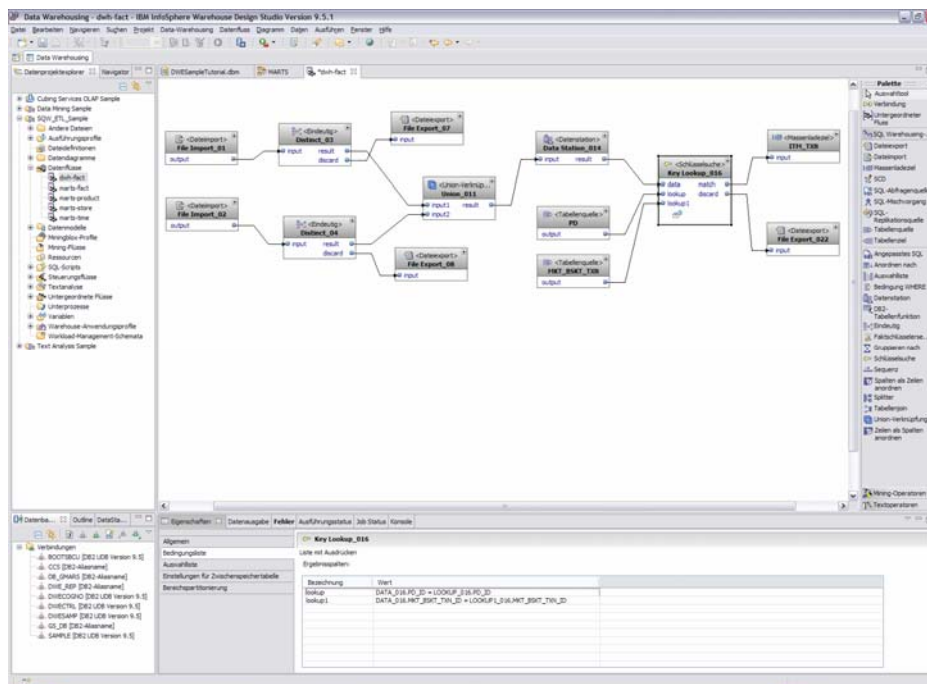


Abbildung 3: Datenfluss des SQL Warehousing Tools im Design Center

Funktionen zur Entwicklung von Datenintegrationsprozessen sind zahlreich und im marktüblichen Umfang vorhanden. Konstrukte zur Modularisierung sowie zur Verschachtelung und Parallelisierung von Abläufen wurden in der Version 9.5 erweitert. Auch die Datenbankbindung wurde verbessert (Nutzung der Komprimierung, Rollin/Rollout von Range-Partitions) und Komponenten des

Data Studios integriert. Data Studio ist das bei IBM strategisch gesetzte Werkzeug für Datenbankmanagement und wird sukzessive DB2 Control Center ablösen. Weitere Neuigkeiten in V9.5 umfassen einen Debugger für Kontrollflüsse sowie die Möglichkeit bei wiederholten Paketfreigaben nur die veränderten Teile des Pakets freigeben zu müssen (delta deployment).

Als Datenquellen werden die Mitglieder der DB2 Familie, gängige relationale Datenbanken, XML und sequentielle Dateien unterstützt. Gehen die Konnektivitätsanforderungen darüber hinaus stehen mit der InfoSphere Federation Server Familie umfangreiche Erweiterungen zur Verfügung, die auch einen Durchgriff auf Host-basierte Datenspeicher ermöglichen. Spielen Change Data Capture (CDC) oder Replikationsanforderungen bei der Anbindung von Quellen eine Rolle, kann auch hier auf umfangreiche Möglichkeiten im IBM Connectivity Software Angebot zurückgegriffen werden. Dabei werden verschiedene Technologien und ebenfalls heterogene Datenquellen (Host- wie Server-basiert) unterstützt.

Das Design Studio bietet über Daten- und Kontrollflüsse hinaus auch die Entwicklung von Mining-Flüssen. DB2 Intelligent Miner unterstützt Modellierung, Visualisierung und Scoring. Es werden Mining-Methoden für Segmentierung (2 Varianten), Assoziationen, Sequentielle Muster, Entscheidungsbaum und Regression (3 Varianten) angeboten. Das in V9.5 neu eingeführte Text Mining ermöglicht aus unstrukturierten Textquellen wichtige Zusammenhänge zu erkennen und zur Integration in klassischen Analysen zu extrahieren. Die IBM-Lösung zeichnet sich durch die Integration dieser Text-Mining-Technologie in die eigene Datenintegrationsplattform aus. Dedizierte Datenqualitätsbausteine bietet die IW Plattform allerdings nicht. Es können aber Funktionen aus dem Data-Mining-Baukasten herangezogen werden, um ein Data Profiling durchzuführen. Für weitergehende Funktionalität verweist IBM auf die Information Server Plattform, allerdings kann der Information Analyzer als Foundation Tool auch einzeln im Kontext von IW lizenziert und verwendet werden.

Die DB2 Datenbank-Engine verfügt über alle wichtigen Eigenschaften, um große Datenmengen und komplexe Abfragen stemmen zu können. Die bereits aus früheren Versionen bekannten Mechanismen zur physischen Aufteilung von Daten, DPF (Data Partitioning Feature) und MDC (Multidimensional Clustering) wurden mit DB2 V9 um das bereits seit vielen Jahren aus dem DB2 for z/OS bekannte Range Partitioning ergänzt. Die Varianten können auch sinnvoll in Kombination angewendet werden. Gleichfalls neu in DB2 V9.5 ist die Komprimierung, die sich technologisch ebenfalls an die auf der Host-Plattform bewährte Variante anlehnt.

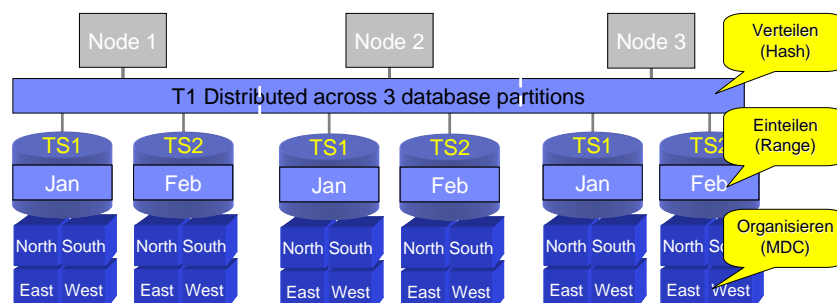


Abbildung 4: Möglichkeiten der Datenverteilung in DB2 UDB 9

Besondere Eigenschaften zeigt DB2 im Umgang mit dimensional modellierten Daten, eine Modellierungsform, die sich vor allem für aufgabenspezifische Data Marts und OLAP-Anwendungen eignet. Die Fähigkeit, vorberechnete Abfrageergebnisse (Materialized Query Tables) anzulegen durch die Datenbankengine transparent zu verwenden und synchron als auch asynchron verwalten zu lassen teilt DB2 mit anderen DBMS-Mitbewerbern am Markt. Nicht so gängig ist aber die Möglichkeit, den DB2-eigenen Abfrage Optimierer (Query-Optimizer) mit erweiterten Metadaten zu Dimensionen, Hierarchien und Kennzahlen zu bestücken. Somit kann eine noch bessere Verwendung von Aggregat-MQTs zur Unterstützung typischer OLAP-Abfragen sichergestellt werden. Einzig Oracle teilt diese Fähigkeit mit DB2 auf dem Markt der relationalen Datenbanken. Die Lösungen werden damit mit Eigenschaften bestückt, die bislang ausschließlich den multidimensionalen Datenbanken (MDBMS) zugesprochen wurden.

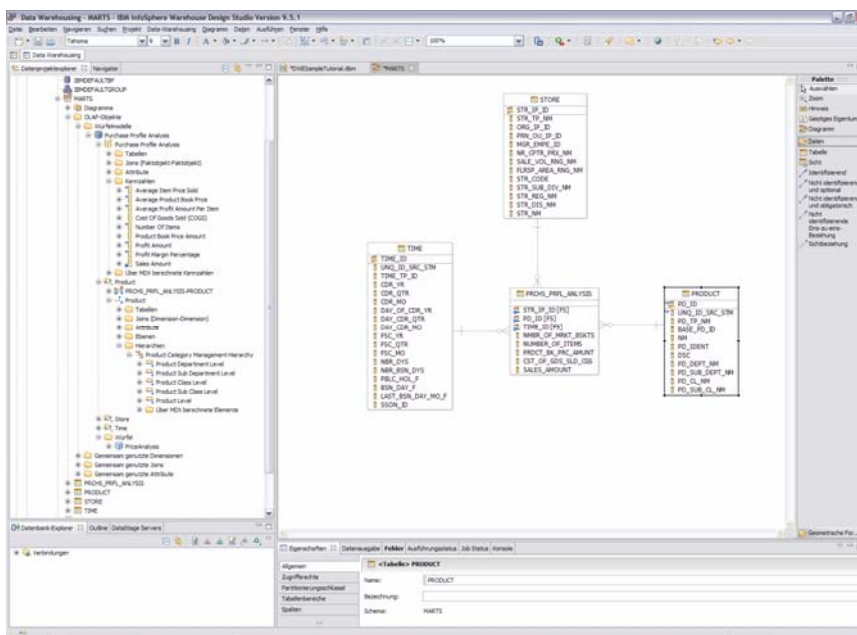


Abbildung 5: Cube Modellierung in Design Center

Noch ein Schritt weiter geht IBM in Version 9.5 mit den Cubing Services. Da viele OLAP-Werkzeuge lediglich die für OLAP optimierte Abfragesprache MDX unterstützen, war es bislang nicht möglich, aus diesen BI-Umgebungen heraus auf relationale DB2 Daten zuzugreifen. Vielmehr war in diesen Fällen ein Replizieren der Daten in ein MDBMS notwendig, was Datenlatenz und Verwaltungskosten erhöhte. Cubing Services bietet nun eine MDX-API auf relationalen DB2 Daten, wie auch einen zuschaltbaren In-Memory-Cache, der es ermöglicht Ergebnisse aus früheren Abfragen für folgende Drilloperationen oder ähnliche Abfragen wieder zu verwenden. Cubing Services unterstützt auch XMLA und seit V9.5.1 auch OLE DB für OLAP, das einen Direktzugriff auf DB2 aus gängigen Desktop Werkzeugen heraus wie beispielsweise Microsoft Excel ermöglicht. Die Quelldaten in DB2 können, müssen aber nicht als dimensionales Star-/Snowflake-Schema modelliert sein. Normalisierte ERM-Schema werden ebenfalls unterstützt.

Eine weitere wichtige Neuerung in V9.5 ist das neue Workload Management (WLM). Bislang war DB2 für Warehousing-Umgebungen mit einem „Türwächter“ namens Query Patroller bestückt. Hier war

es zwar möglich, eingehende Abfragen nach definierten Regeln hinsichtlich des Ressourcenverbrauchs zu überprüfen und unter Umständen die Ausführung zu verweigern. Das neue WLM ergänzt diese Funktionalität um eine laufende Überwachung von aktuell aktiven Abfragen. Zeigt sich eine Query dabei als zu aufwändig, kann sie in eine andere Prioritätsklasse umgeleitet, gedrosselt oder gar abgebrochen werden. Insgesamt ist WLM mit feineren Steuerungsmechanismen ausgestattet. Im Vergleich zu Query Patroller fehlt allerdings derzeit noch die historische Sicht auf abgewickelte Abfragen, weshalb in manchen Fällen noch beide Werkzeuge zum Einsatz kommen. Mittelfristig soll WLM den QP allerdings vollständig ablösen.

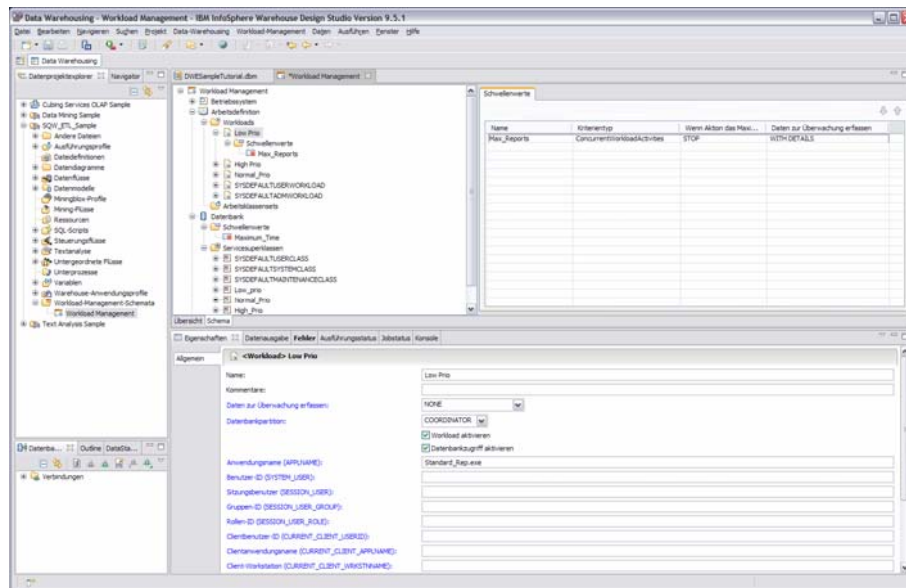


Abbildung 6: Workload Management in DB2 V9.5

## IBM InfoSphere Information Server 8.1

Die Datenintegrationsplattform IBM InfoSphere Information Server, vormals IBM WebSphere Data Integration Suite, erhielt mit der Version 8 nicht nur einen neuen Namen, sondern wurde vollkommen überarbeitet. Insbesondere war die neue Architektur eine Reaktion auf die mangelnde Integration der einzelnen Produkte untereinander. Seit 2008 kennzeichnet „InfoSphere“ als neuer Brand die Integrationsbemühungen innerhalb des Information Management Software Portfolios. Mit dem aktuellen Release 8.1 des InfoSphere Information Server (IS) setzt IBM nun die Integration fort, vor allem im Bereich des Metadatenmanagement.

Eine der Neuerungen in der aktuellen Version 8.1 sind Ergänzungen in der Plattformunterstützung für 64-bit Architekturen. Auch wird der IS seit 2007 als Appliance unter dem Namen InfoSphere Information Server Blade angeboten um einen schnellen Einstieg in die Datenintegration zu ermöglichen und einfacher bei wachsenden Datenvolumen durch Hinzukaufen einer weiteren Information Server Box skalieren zu können.

Ein Architekturmerkmal des IS sind die zentralen Plattformdienste, die jedem Produkt zur Verfügung stehen. Hierzu gehören ein integriertes Metadatenmanagement, parallele Verarbeitung, Logging-Mechanismen, Implementierung der Zugriffssicherheit sowie Erweiterungen der Konnektivität (Adapter für Standardanwendungssoftware). Die einzelnen Produkte nehmen unterschiedliche Aufgaben innerhalb der Datenintegration wahr. So wird mit dem InfoSphere Information Server eine Plattform mit Services bereitgestellt, in die sich folgende Produkte als Module integrieren.

- **InfoSphere DataStage**  
Herzstück des IS ist diese ETL-Komponente, die sowohl die Konnektivität als auch die Funktionen zur Datenintegration bereitstellt. Die Entwicklung der Datenintegrationsprozesse erfolgt im Windows-basierten InfoSphere DataStage and QualityStage Designer. Ablaufumgebung für die Engine ist eine Server-Umgebung oder ein Host mit z/Linux. Für eine native Host-Ablaufumgebung ist alternativ die InfoSphere DataStage MVS Edition erhältlich.
- **InfoSphere QualityStage**  
Dieses Produkt erweitert die Stages (Transformationsobjekte) der InfoSphere DataStage um Datenqualitätsfunktionen.
- **InfoSphere Information Analyzer**  
Mit diesem Werkzeug können Profile der inhaltlichen Datenqualität erstellt und Datenregeln überwacht werden.
- **InfoSphere Metadata Server**  
Seit Version 8.0 kann das zentrale Metadaten Repository in die relationalen Datenbanksysteme IBM DB2 UDB, Oracle oder MS SQL Server gespeichert werden. Eine DB2 restricted use Lizenz wird mit dem IS ausgeliefert. Zur Auswertung der Metadaten kann der InfoSphere DataStage und QualityStage Designer oder die InfoSphere Metadata Workbench sowie die InfoSphere Business Glossary Module zur Anzeige von Geschäftsbegriffen und deren Beziehungen zu Datenobjekten genutzt werden.



- **InfoSphere Information Services Director**  
Die Bereitstellung von DataStage-, QualityStage oder Federation Server-Objekten als Services erfolgt durch den Information Services Director. Dieses Werkzeug erzeugt alle Informationen um Ergebnisse aus der Entwicklung als Web Services oder auch als J2EE konforme EJB (Enterprise Java Beans) zu registrieren und im Applikationsserver zu installieren (der WebSphere Application Server wird mit dem IS ausgeliefert).
- **Connectivity Software**  
Unter diesem Begriff subsumieren sich Werkzeuge für den erweiterten Echtzeit-Zugriff auf Datenquellen, wie beispielsweise der InfoSphere Federation Server (implementiert Echtzeit-Zugriff), InfoSphere Replication Server, InfoSphere Change Data Capture.



Abbildung 7: IBM InfoSphere Information Server Plattformarchitektur

InfoSphere DataStage selbst ist ein klassisches ETL Werkzeug, in dem die Transformationen vorwiegend in einer eigenen Engine durchgeführt werden. Alternativ können Transformationen aber auch an externe Werkzeuge delegiert werden. So können bspw. Datenintegrationsprozesse aus dem IBM SQL Warehousing Tool oder Cognos Data Manager direkt in einen DataStage-Prozess integriert werden. Eine dritte Variante bietet die Verlagerung von Transformationsschritten nach dem „push-down“-Prinzip auf die Quell- oder Zielsysteme unter Verwendung von nativem SQL.

Diese Thematik greift IBM im neuen Release gesondert durch die InfoSphere DataStage Balanced Optimization auf. Diese Option ermöglicht es dem Entwickler den Datenintegrationsprozess semi- oder automatisch zu optimieren ohne sich um komplizierte SQL-Syntax sorgen zu müssen. Dabei entscheidet das System nach Erstellung der Transformationsstrecke im Designer, welche Teilstrecken in SQL übersetzt und auf das Quell- oder Zielsystem verlagert werden können. Aktuell werden die Datenbanken IBM DB2 sowie Teradata unterstützt.

Um wachsenden Anforderungen zu begegnen unterstützt der InfoSphere Information Server sowohl Scale-up wie auch Scale-Out Szenarien. Einzelne Prozesse können mittels Datenpartitionierung sehr granular über mehrere Prozessoren/Rechner hinweg oder durch Pipelining in einem Datenfluss



parallelisiert werden. Damit ist der InfoSphere Information Server für die Verarbeitung großer Datenmengen geeignet.

Alle Produktmodule des IS legen ihre Metadaten direkt ohne lokale Zwischenspeicherung im zentralen Repository, dem Metadata Server, ab. Die Metadaten der einzelnen Produktmodule werden in einem integrierten, dem CWM Standard folgenden Modell gespeichert, welches nahezu alle notwendigen Metadaten für das Data Warehousing beinhaltet.

Das Metadatenmodell ist um Objekte und Attribute erweiterbar und bietet somit die notwendige Flexibilität um bspw. berichts- oder analysespezifische Metadaten von Business-Intelligence-Werkzeugen zu integrieren. Dies ermöglicht Abhängigkeiten von den einzelnen Datenquellen bis zum Bericht zu dokumentieren. So können Auswirkungen etwaiger Änderungen in der Quelle oder im Ziel transparent für abhängige Objekte nachvollzogen und analysiert werden. Diesbezüglich unterstützt der InfoSphere DataStage & QualityStage Designer den Entwickler mit detaillierten Auswertungsmöglichkeiten innerhalb der Plattform. Für die Verwaltung sowie zur grafischen und textuellen Auswertung werkzeugübergreifender Metadaten stellt die Web-basierte Metadata Workbench entsprechende Funktionen zur Verfügung.

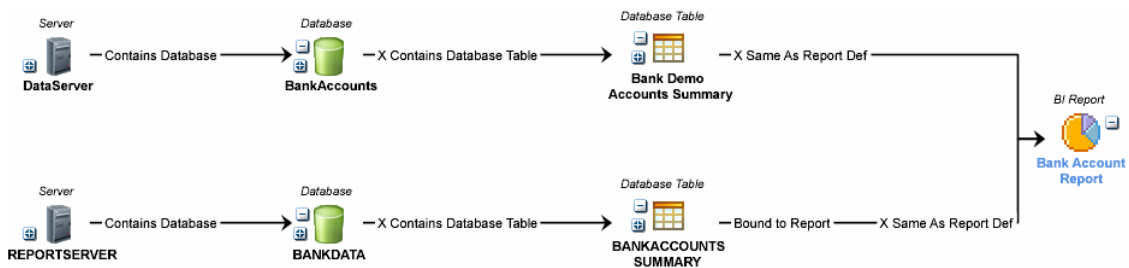


Abbildung 8: Werkzeugübergreifende Abhängigkeitsanalyse von der Quelle bis zu einem Bericht aus der Metadata Workbench.

Zum Aufbau eines unternehmensweiten Verständnisses des Geschäftes setzt IBM auf das Foundation Tool Business Glossary. Das Business Glossary basiert auf dem Metadata Server und bietet durch eine Web-basierte Oberfläche Funktionen zur Navigation in und Definition von ergänzenden Metadaten pro Objekt an. Business Glossary Anywhere kann zudem auch aus anderen Anwendungen heraus (IBM oder nicht-IBM) genutzt werden. Begriffe können aus verschiedenen Anwendungen markiert und per Kontextmenü die Metadaten angezeigt werden.

Der Austausch der Metadaten mit zahlreichen Anwenderwerkzeugen (u.a Business Objects, Cognos oder MicroStrategy) ist über MetaBroker oder MetaBridges möglich. Speziell Metadaten des von der IBM übernommenen Business Intelligence Anbieters Cognos sind so schon erfolgreich integriert worden.

Eine Versionierung der Metadaten direkt im Repository ist derzeit noch nicht vorgesehen. IBM empfiehlt die Nutzung eines externen Werkzeuges zur Quellcode-Verwaltung, welches derzeit noch manuell angebunden werden muss. Die Integration des hierfür von der IBM strategisch gesetzten Rational Clearcase steht mit auf der Roadmap für weitere Entwicklungen des InfoSphere Information Server. Aktuell bietet IBM mit dem InfoSphere Information Server Manager allerdings die Möglichkeit

Exportpakete zum Austausch zwischen verschiedenen Installationen zu schnüren. Dazu können Objekte nach ihren Abhängigkeiten durchsucht und in ein Paket gepackt werden. Diese Exportpakete werden in IBM-eigenen Dateiformaten im Dateisystem versioniert gespeichert. Ein Vergleich zwischen den verschiedenen Versionen eines Paketes ist in der grafischen Benutzeroberfläche möglich.



Abbildung 9: Drei Sichten auf das Business Glossary. Im Hintergrund zu sehen ist die Hierarchie des Objektes „Customer“, rechts das Ergebnis aus einer Suche im Business Glossary aus der man per Mausklick auf die Detailansicht zu einem Objekt wechseln kann, links unten sind Detailinformationen zu einem Geschäftsbegriff zu sehen mit Beziehungen zur verwendeten Datenbank-Tabelle.

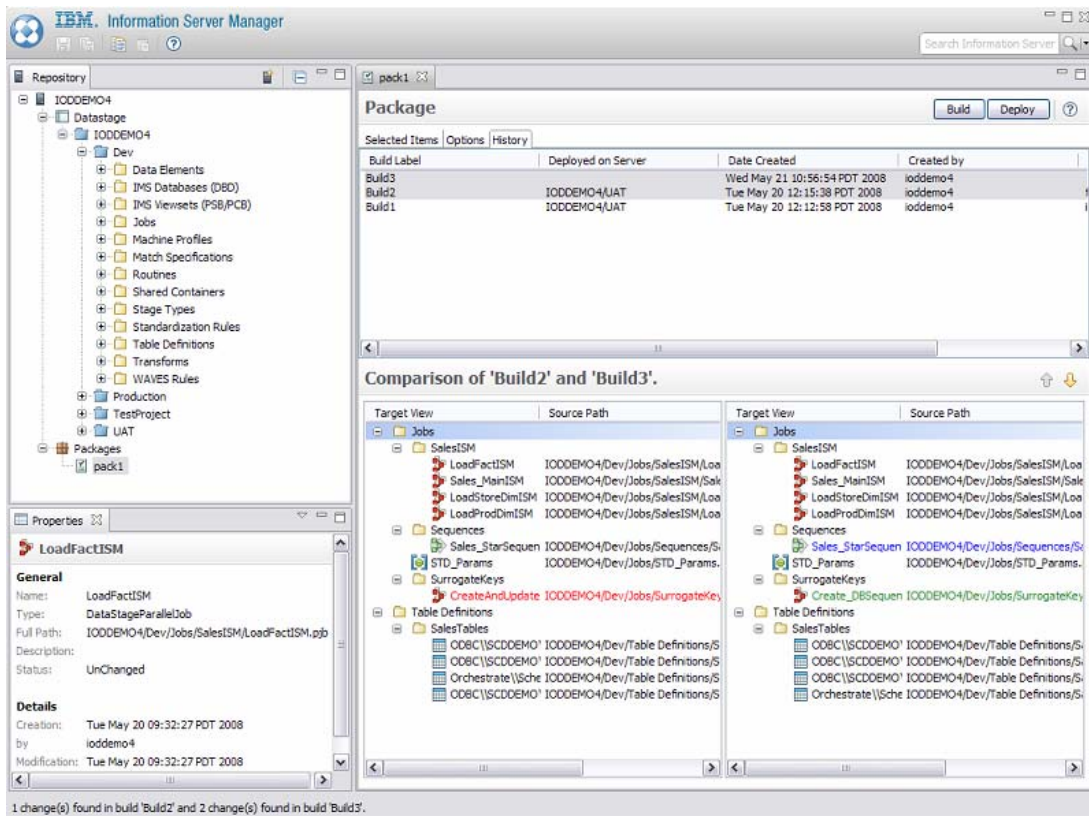


Abbildung 10: Versionierung mit dem InfoSphere Information Server Manager

Für die grafische Entwicklung der Datenintegrations- und Datenqualitätsprozessen wird ein gemeinsamer Designer für DataStage und QualityStage genutzt. Für die Entwicklung der Datenflüsse steht eine Fülle von Stages (vordefinierte Logiken/Transformationen) bereit, die den Großteil der klassischen Datenintegrationsaufgaben abdecken. Weitere Funktionen können durch eigene Stages, Transforms und Routinen ergänzt werden, die in C++ oder Java entwickelt werden können.

Einmal entwickelte Logik kann gekapselt und mehrfach in unterschiedlichen Datenflüssen (Jobs) wieder verwendet werden. Grundsätzlich werden Variablen in Jobs und logische Templates (shared container) zur Laufzeit mit Parametern befüllt. So ist es bspw. möglich Daten aus unterschiedlichen Quellen mit demselben Job/Template zu extrahieren, in dem die Variablen für das Quellsystem zur Laufzeit durch die entsprechenden Parameter geändert werden. Solche Parameter werden entweder auf Job-Ebene oder in Parameter Sets festgelegt, die in DataStage-Objekten im Repository gespeichert werden. Für die Systemkonfiguration liefern in diesem Fall eigene Konfigurationsdateien die nötigen Informationen, u.a den Grad der Parallelität.

Die Kontrollflüsse werden ebenfalls in der Designumgebung gestaltet. Dazu stehen ausreichende Konstrukte (wie Schleifen) oder Funktionen (z.B. Sende Mail) zur Verfügung. Ein verwaltetes Wiederanlaufverfahren für Datenintegrationsprozesse ist nicht implementiert, kann jedoch leicht mit Bordmitteln umgesetzt werden. Dies obliegt allerdings dem Entwickler.

In Sachen Entwicklungskomfort hebt sich IBM besonders in zwei Bereichen hervor. Zum einem werden neun Sprachen unterstützt (u.a. Deutsch, Französisch, Spanisch und Italienisch für den europäischen Raum), zum anderen durch den Job Viewer. Dieser ermöglicht die Vorschau auf einen Job durch das Erscheinen eines kleinen Vorschaubildes, welches durch einen Mausklick vergrößert

werden kann, wenn der Entwickler in der Job-Bibliothek mit dem Mauszeiger über die einzelnen Jobnamen fährt.

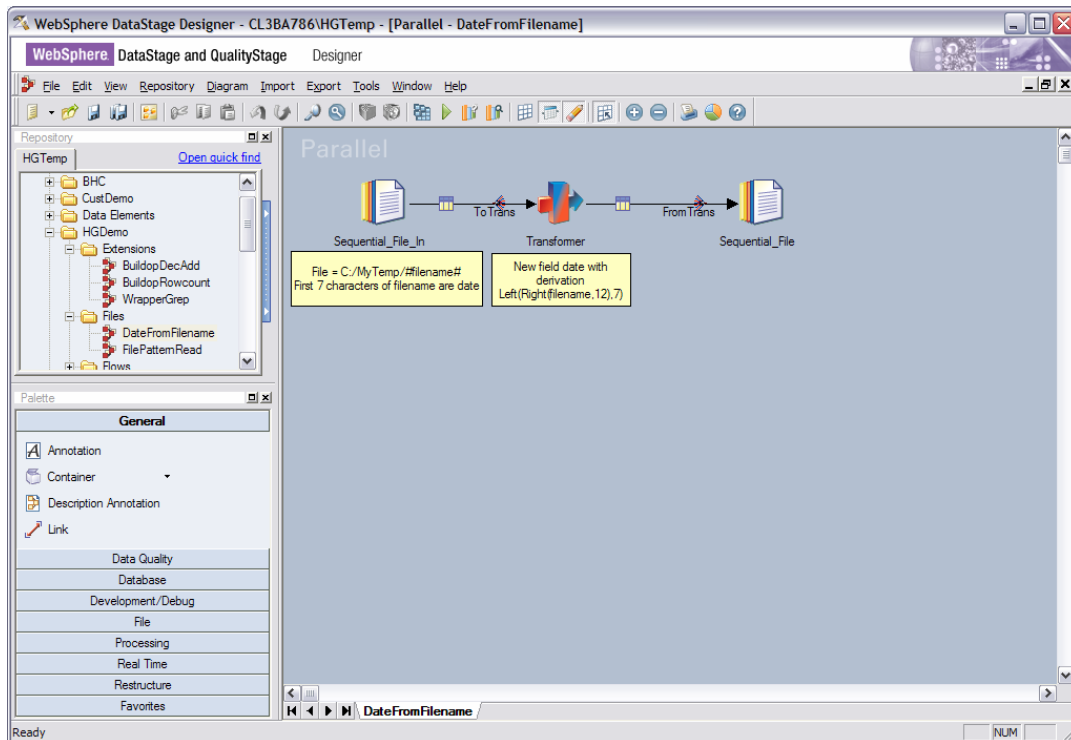


Abbildung 11: InfoSphere DataStage & QualityStage Designer

Zur Einbindung des Know-how aus den Fachbereichen in eher technisch-orientierte Datenintegrationsprozesse bietet IBM das Foundation Tool InfoSphere Information Server FastTrack an. Damit sind Fachbereiche in der Lage einfache Mappings zwischen Datenquelle und Datenziel zu erstellen. Die resultierenden Transformationsregeln können als von FastTrack generierte DataStage Jobs direkt in die Datenflüsse des IS integriert und dort bei Bedarf weiter bearbeitet werden.

Zur Integration von Daten bringt der IS zwei Varianten mit. Einerseits die im InfoSphere DataStage vorhandenen Schnittstellen zu Datenbanksystemen, Dateien und anderen Datenhaltungen, andererseits die durch den InfoSphere Federation Server möglichen Zugriffe auf strukturierte wie auch unstrukturierte Daten. Weitere Systeme können über die IBM Connectivity Software bereitgestellt werden. Damit deckt IBM eine sehr breite Palette an Systemen ab. Für die Anbindung der gängigen Standardanwendungssoftware (z.B. SAP ERP) können sogenannte InfoSphere DataStage Packs implementiert werden. In Sachen Echtzeit-Unterstützung bietet IBM das umfangreichste Lösungsangebot im Testfeld. Besondere Stärken liegen hier im Umfang der unterstützten Systeme für die Extraktion geänderter Daten, dem sogenannte Changed Data Capture (CDC), speziell im Bereich Replikation. Breite Unterstützung ist auch bei der Bereitstellung und Anbindung von Web-Services durch den InfoSphere Information Server Services Director und bei der Föderation durch den InfoSphere Federation Server gegeben.

Vor allem Datenqualitätsfunktionen fanden in den letzten Jahren Einzug in die Datenintegrationsplattformen. Das Data-Profiling-Werkzeug InfoSphere Analyzer umfasst Funktionen für Spalten, Tabellen und Beziehungsanalysen. Für die Analysen profitiert der Data Profiler aufgrund der serviceorientierten Architektur des InfoSphere Information Server von den zentralen Plattformdiensten und Produktmodulen, so z.B. von einer umfangreichen Konnektivität oder Fähigkeiten zur parallelen Verarbeitung von Daten.

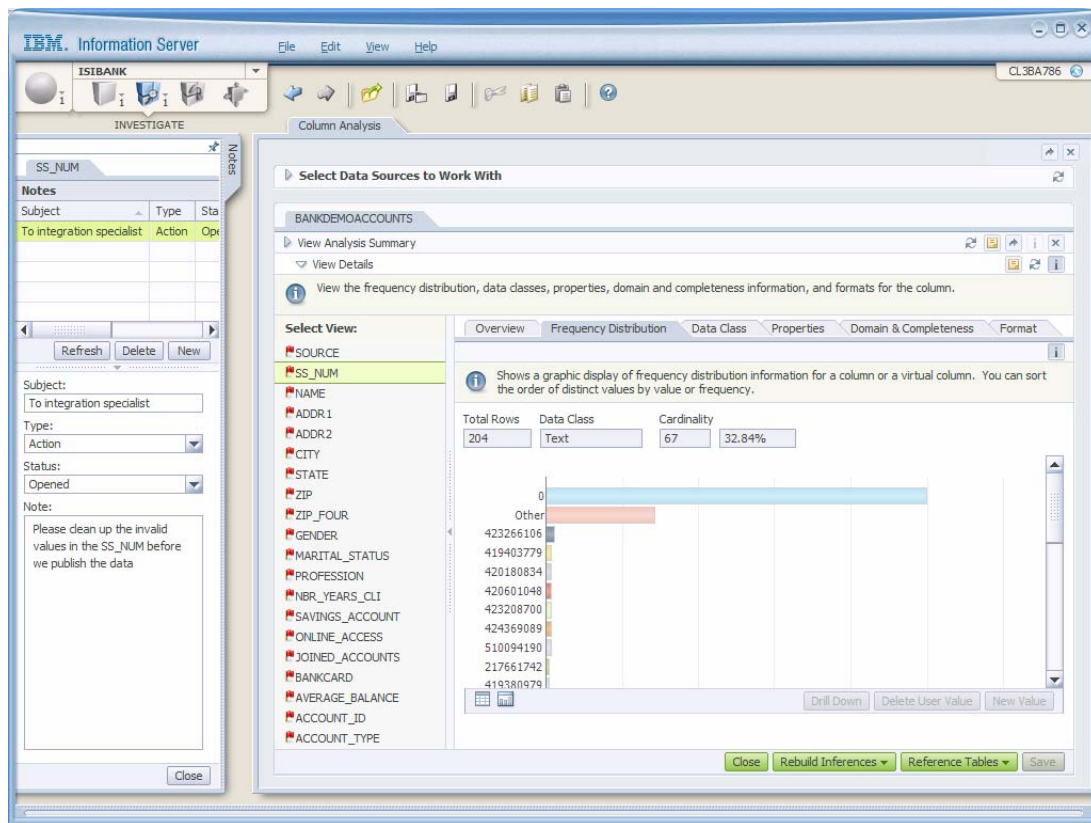


Abbildung 12: Ergebnis einer Spaltenanalyse im InfoSphere Information Analyzer

Für die Datenanalyse werden die Daten aus den Quellsystemen in ein eigenes Arbeits-Repository geladen. Die Analyseergebnisse können danach in das zentrale Repository des InfoSphere Information Server überführt werden und stehen in einem integrierten Modell zur weiteren Verarbeitung zur Verfügung. In der neuen Version des Information Analyzer wurde der Funktionsumfang dahingehend erweitert, dass vor allem der Datenqualitätszyklus durch Funktionen zum Aufbau und Auswertung von Regeln und Metriken breiter unterstützt wird.

Im Bereich Data Cleansing werden mit dem architekturell vollständig integrierten QualityStage die geforderten Funktionen angefangen beim Parsing, Standardisierung über Matching, Deduplizierung, Konsolidierung bis hin zur Anreicherung unterstützt. Für Adressdaten stehen dabei Standards aus über 200 Ländern zur Verfügung. Zudem werden weiterführende Analysen im Bereich Identitätsmanagement in separaten Modulen zur Verfügung gestellt, wie das Erkennen von Personenbeziehungen (Entity Analytics Solutions) oder der länderübergreifende Namensabgleich (Global Name

Recognition). Weitere Funktionen bietet die Integration mit dem IBM InfoSphere MDM Server (Master Data Management).

Während die Integration der Werkzeuge in den Bereichen Entwicklungsunterstützung und Metadatenmanagement bereits gut gelungen ist, bleibt bei der Systemverwaltung noch Verbesserungspotenzial. Das betrifft vor allem die Administration durch unterschiedliche Werkzeuge für die jeweiligen Benutzerrollen. Wesentliche Neuerungen präsentiert die Version 8.1 im Bereich der Prozessüberwachung und -analyse. So können zusätzliche Daten zum Ablauf optional zu den operativen Metadaten gesammelt und grafisch sowie tabellarisch analysiert werden. Bislang werden die operativen Metadaten in XML-Dateien gespeichert, können aber über entsprechende Import-Mechanismen in das relationale Repository überführt werden. Dadurch werden dann Performanceanalysen über die Zeit hinweg möglich. Log-Dateien können nun über die Web-basierte Administrationsoberfläche des IS analysiert und konfiguriert werden. Administratoren sind somit in der Lage Log-Dateien auf die nötigen Informationen zu begrenzen und die Übersichtlichkeit damit zu steigern.



## Bewertung

Die untersuchten Produkte wurden in der BARC Software Evaluation mit anderen Werkzeugen ihrer Kategorie anhand des BARC-Kriterienkataloges für Data Warehousing und Datenintegration in über 150 Kriterien verglichen und bewertet. Die Bewertung basiert auf einer Skala von 1 (minimale Unterstützung) bis hin zu 10 Punkten (Referenz). Die Ergebnisse weiterer Anbieter wurden für diesen Studienauszug anonymisiert.

Für den Vergleich der Werkzeuge werden folgende Kategorien genutzt:

- **Architektur**  
Neben der Breite an unterstützten Plattformen und Betriebssystemen wird vor allem die Architektur hinsichtlich der Möglichkeiten für Performanceoptimierungen, Skalierbarkeit sowie ihrer Wartbarkeit bewertet.
- **Modelle und Metadaten**  
Die Speicherung, Vollständigkeit und Integration der einzelnen Produktkomponenten im Metadatenmodell, dessen Erweiterbarkeit, Austausch von Metadaten sowie Analyse- und Auswertungsmöglichkeiten definieren die Kriterien für die Bewertung. Mit eingeflossen sind ebenfalls vordefinierte Modelle und Funktionen für die Datenmodellierung.
- **Entwicklungsunterstützung**  
Die Kapselung von Logik, Wiederverwendbarkeit von Objekten und Unterstützung bei der Fehlersuche spielen hier eine ebenso große Rolle wie Mehrbenutzerunterstützung, Versionierung oder den gebotenen Entwicklungskomfort.
- **Konnektivität**  
Die Qualität und der Umfang der Schnittstellen sowie die Unterstützung von near-time Integrationsanforderungen bestimmen die Bewertung in dieser Kategorie. Die Betrachtung reicht angefangen bei einfachen Textdateien, über relationale wie multidimensionale Datenspeicher, Anbindung von Standardanwendungssoftware bis hin zu komplexen Host-Dateien. Unter dem Aspekt near-time Anforderungen werden Technologien für das Changed Data Capture, Replikation, Föderation (EII), Messaging-Systeme (EAI) oder Web-Services mit berücksichtigt.
- **Datentransformation**  
Gegenstand der Bewertung ist die Abdeckung der Anforderungen an Transformationsfunktionen, Verarbeitungslogik, Funktionen zur Datenanreicherung (bspw. Schlüsselgenerierung) sowie die Unterstützung bei der Abbildung von Historienanforderungen (bspw. slowly changing dimensions). Mechanismen für Sicherstellung von Performance und Skalierbarkeit werden ergänzend zur Bewertung in der Kategorie Architektur hier auf funktionaler Ebene betrachtet.
- **Datenmanagement-Applikationen**  
Dieser Bereich fokussiert einerseits Lösungen zur Prüfung und Sicherstellung der Datenqualität. Kriterien bewerten im Detail Funktionen für die Datenanalyse (Data Profiling), Datenbereinigung (Data Cleansing), Datenanreicherung (Data Enrichment) wie die Möglichkeiten zur



Überwachung, Steuerung und Auswertung der Datenqualität. Desweiteren werden Funktionen zur Umsetzung von Data Mining in der Data-Warehouse-Plattform beleuchtet.

- **Datenspeicherung (nur bei Data-Warehouse-Plattformen)**  
Hier werden die Fähigkeiten zur Verwaltung von relational und dimensional modellierten Daten betrachtet, sowie die Integration dieser Ansätze miteinander. Performance und Skalierbarkeit werden aus Sicht sowohl lesender als auch schreibender Operationen, beispielsweise die Vorberechnung von Abfrageergebnisse oder das in-memory Caching, bewertet.
- **Systemverwaltung**  
Aus Sicht der Systemverwaltung sind sowohl die Aspekte der Datenintegration als auch die der Datenspeicherung zu betrachten. Dazu gehört die Sicherstellung eines reibungslosen Betriebes auf Basis störungsfreier Infrastruktur, wie auch die Überwachung laufender Prozesse und das Handling etwaiger Fehlersituationen. Schlussendlich spielt auch die Verwaltung der Systemicherheit hier eine große Rolle, also Funktionen zur Datenverschlüsselung, Benutzer- und Rechteverwaltung.

## IBM InfoSphere Warehouse

Anbieter	IBM InfoSphere Warehouse V9.5.1	A	B	C	D	Durchschnitt	Top-Werte
<b>Bewertungskategorien</b>							
Architektur	8	5	4	7	5	6	8
Modelle und Metadaten	7	7	4	5	5	6	7
Entwicklungsunterstützung	7	6	7	7	-	7	7
Konnektivität	6	5	3	6	2	4	6
Datentransformation	9	6	6	8	-	7	9
Datenmanagement-Applikationen	5	9	4	6	4	6	9
Datenspeicherung	9	5	7	8	9	8	9
Systemverwaltung	6	5	5	6	3	5	6

Tabelle 1: Aggregierte Bewertung von IBM InfoSphere Warehouse im Vergleich zu vier anderen leistungsfähigen Data-Warehouse-Plattformen auf Basis des BARC-Kriterienkataloges für Data Warehousing und Datenintegration. (1 = schlecht, 10 = beste Bewertung)

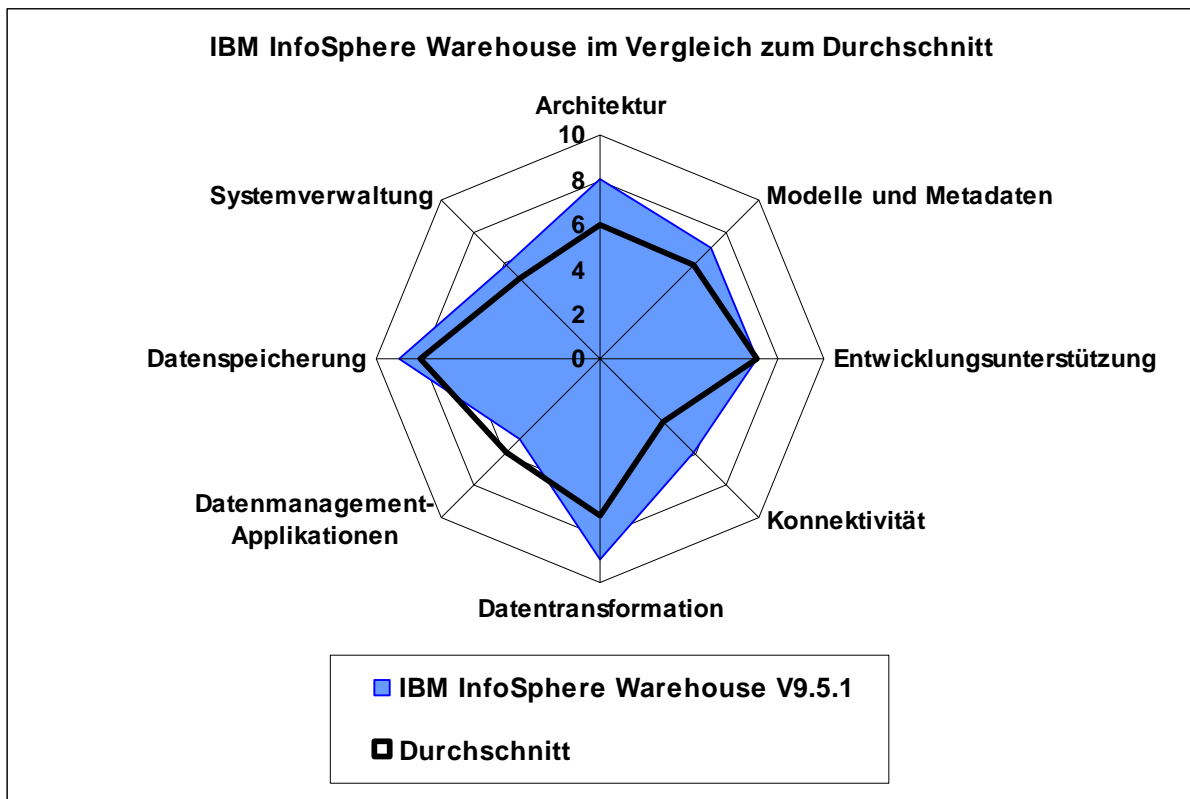


Abbildung 13: IBM InfoSphere Warehouse im Vergleich zur durchschnittlich erreichten Punktzahl pro Kategorie.

IBM InfoSphere Warehouse (IW) liegt mit Ausnahme von zwei Kategorien über dem Durchschnitt und erreicht in sieben von acht Bewertungsbereichen Top-Werte. Lediglich im Bereich der Datenmanagement-Applikationen fällt IW auffällig zurück, was im Wesentlichen an der fehlenden Integration des Datenqualitätsangebotes von IBM in InfoSphere Warehouse liegt. Diese Lücke kann durch andere Produkte aus dem Portfolio des Anbieters kompensiert werden. So stellen der InfoSphere Information Server als Gesamtpaket und der InfoSphere Information Analyzer als Foundation Tool entsprechende Funktionen zur Verfügung. An der Kategorie-Bewertung nicht direkt zu erkennen ist, dass die Data-Mining-Komponente des IW eine gute Bewertung erreicht.

Verbesserungswürdig erscheint uns im IW die Systemverwaltung. Die vollzogenen Integrationsanstrengungen in Richtung Administrationskonsole haben sich zwar ausgezahlt. Nach wie vor ist der Administrator aber konfrontiert mit Werkzeugen aus verschiedenen Umgebungen (IW, WebSphere Application Server und DB2). Auch im Bereich DB2-Verwaltung gibt es im Vergleich zu Vorgängerversionen deutliche Verbesserungen, insbesondere das neue Workload Management ist hier zu nennen. Als verwaltungsfrei kann sich DB2 jedoch, wie alle etablierten Datenbanken, nicht bezeichnen. Der zunehmende Wettbewerbsdruck durch die neu erschienenen Data-Warehouse-Appliances, die sich einfachere Verwaltbarkeit auf die Fahnen schreiben, wird nach unserer Einschätzung für noch weitere Entwicklungsschübe sorgen.

Im Bereich Modelle und Metadaten erreicht IW mit die höchste Bewertung, was auf die gute Unterstützung von Datenmodellierung für analytische Umgebungen zurückzuführen ist. Das Metadatenmanagement an sich weist trotz der Möglichkeit der Nutzung des Metadata Servers über die Foundation Tools noch Verbesserungspotential auf. Hier wäre eine noch bessere Integration innerhalb des eigenen Software-Portfolios wünschenswert.

Stärken sind für den IW und die DB2 im Datenmanagement in Data-Warehousing-Umgebungen auszumachen. Die Technologieunterstützung lässt weitgehend freie Wahl bei der Gestaltung der eigenen Datenarchitektur und bevorzugt weder zentralistische noch dezentralistische Datenstrategien. So können Kunden mit dieser Plattform die Strategie wählen, die am besten zur Informationsstrategie des Unternehmens passt, ohne zwangsläufig auf weitere Datenbanktechnologien wie etwa MDBMS ausweichen zu müssen. Das kann sich positiv auf die Gesamtkosten auswirken.

Zusammenfassend sehen wir die herausragenden Merkmale der InfoSphere Warehouse Lösung in folgenden Bereichen:

- DB2 als skalierfähiges DBMS für Data Warehousing
- DB2 Workload Management zur Steuerung von verschiedenartigen Workloads
- OLAP Optimization mit Cubing Services und Cube Views zur performanten Unterstützung von marktgängigen BI-Applikationen inklusive OLAP
- In die Entwicklungsumgebung sehr gut integrierte Modellierungsfunktionen, die sowohl ERM als auch dimensionale Datenmodelle unterstützen
- Einbindung von Data und Text Mining in die Datenintegrationsplattform

## IBM InfoSphere Information Server

Anbieter	IBM InfoSphere Information Server 8.1	A	B	C	D	Durchschnitt	Top-Werte
<b>Bewertungskategorien</b>							
Architektur	8	8	7	9	7	8	9
Modelle und Metadaten	7	5	7	5	6	6	7
Entwicklungsunterstützung	8	8	6	9	9	8	9
Konnektivität	9	7	5	8	7	7	9
Datentransformation	9	8	6	9	8	8	9
Datenmanagement-Applikationen	8	8	9	6	7	8	9
Systemverwaltung	7	8	6	8	7	7	8

Tabelle 2: Aggregierte Bewertung von IBM InfoSphere Information Server im Vergleich zu vier anderen leistungsfähigen Datenintegrationsplattformen auf Basis des BARC-Kriterienkataloges für Data Warehousing und Datenintegration. (1 = schlecht, 10 = beste Bewertung)

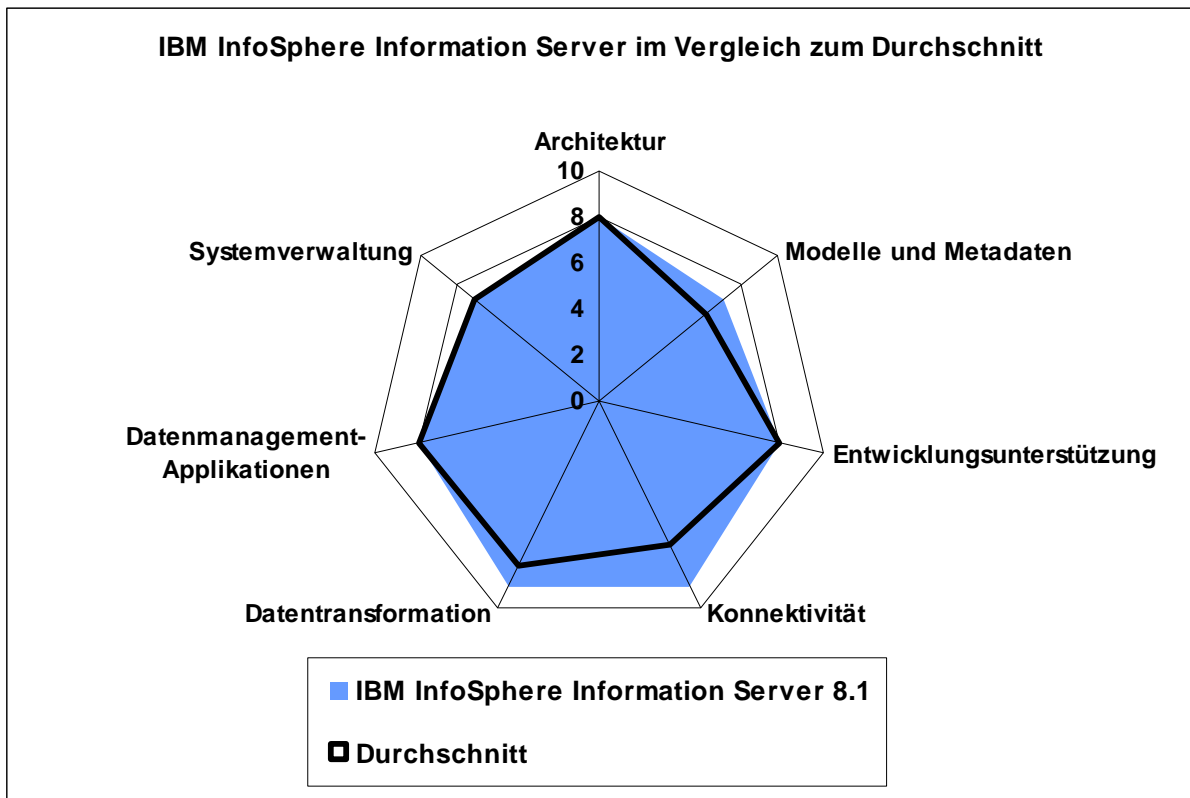


Abbildung 14: IBM InfoSphere Information Server im Vergleich zur durchschnittlich erreichten Punktzahl pro Bewertungskategorie.

IBM InfoSphere Information Server (IS) liegt in drei Kategorien über dem Durchschnitt und erreicht in diesen Bereichen eine Top-Bewertung. Im Bereich Modelle und Metadaten tragen vor allem die große Offenheit und damit einhergehende Integrationsfähigkeit in heterogene Lösungsansätze zu dieser guten Bewertung bei. Schwächen in den Modellierungsfunktionen werden durch das leicht zu integrierende Produkt IBM Rational Data Architect behoben.

Eine Top-Bewertung erreicht der IS im Bereich Konnektivität. Hier verfügt IBM insgesamt über das umfangreichste Angebot unter den betrachteten Herstellern. Der InfoSphere Server an sich bietet schon gute Möglichkeiten, die durch die Anbindung weiterer IBM-Lösungen, wie beispielsweise InfoSphere Federation Server, InfoSphere Replication Server und InfoSphere Change Data Capture, zusätzlich erweitert werden können.

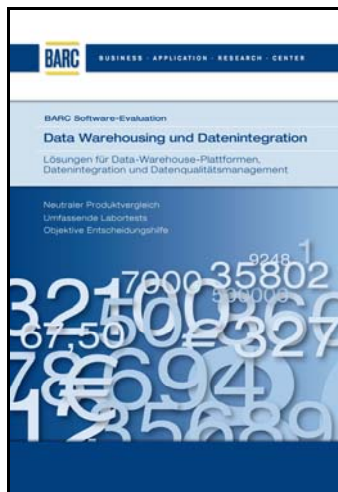
Mittlere Bewertungen erreicht der IS in den Bewertungskategorien Entwicklung und Systemverwaltung. Im Bereich Entwicklung sehen wir vor allem Verbesserungsbedarf im Debugging und der noch fehlenden Anbindung von Quellcodeverwaltungslösungen, die aber für die Folgeversion angedacht sind. Die momentan am besten bewertete Lösung legt hier ein ausgeklügeltes Sandbox-Prinzip, ein starkes Versionierungskonzept und eine sehr flexible Testunterstützung vor. Die Systemverwaltung, der schwächste Bereich des IS, bedarf nach unserer Einschätzung noch weiterer Integrationsarbeiten. Hier zeigt sich die Historie der getrennten Werkzeugkomponenten noch deutlich. Operationale Metadaten (Laufzeitstatistiken) werden derzeit in XML-Dateien gesammelt, können aber zur Auswertung automatisiert in eine Datenbank übertragen werden. Die mitgelieferten Auswertungsmechanismen zeigen noch Verbesserungspotential auf. IBM hat angekündigt, dass hier künftig die zugekaufte Cognos-Plattform genutzt werden kann und somit diese Verbesserungen liefern soll. Abschließend soll das lediglich programmatisch umsetzbare Wiederanlaufverfahren erwähnt werden. Das Referenzprodukt in diesem Bereich liefert hier eine für den Entwickler deutlich komfortablere Unterstützung.

Zusammenfassend sehen wir die herausragenden Merkmale von InfoSphere Information Server in folgenden Bereichen:

- Datenintegrationsplattform, die sich durch sehr gute Konnektivität und umfangreiche near-time Eigenschaften nicht nur für das Data Warehousing eignet.
- Sehr gut integrierte, funktional starke Datenqualitätskomponenten.
- Offenes Metadatenmanagement, das die Bedürfnisse von Fachanwendern mit berücksichtigt und sich mittels Business Glossary Anywhere auch aus anderen Umgebungen heraus verwenden lässt.
- InfoSphere DataStage Balanced Optimization ermöglicht eine gemischte Verteilung der Transformationslast auf ETL-Engine, Quell- oder Zielsystem.
- Skalierfähige Architektur, die gemeinsam mit nur wenigen Mitbewerbern eine Plattform für sehr hochvolumige Anforderungen liefert.

IBM besitzt mit den Lösungen InfoSphere Warehouse und Information Server zwei umfangreiche und leistungsstarke Angebote am Markt für Data Warehousing- und Datenintegrationssoftware. Der IS unterstützt als Plattform die unternehmensweite Datenintegration vom operativen Umfeld bis hin zum Enterprise Data Warehouse. Das IW setzt sinnvoll am Data Warehouse auf und unterstützt das analytische Datenmanagement bis hin zu den analytischen Applikationen.

## BARC Software-Evaluation Data Warehousing und Datenintegration



### Data Warehousing und Datenintegration

**1 Jahr Online-Zugriff auf Produktbewertungen,  
Hintergrundanalysen und Marktforschungser-  
gebnisse**

Fortlaufende Erweiterung und Aktualisierung

<http://www.barc.de/dw>

Preis: 2.500 Euro

Die Software-Evaluation Data Warehousing und Datenintegration vergleicht und bewertet die führenden Werkzeuge für Data Warehousing und Datenintegration im deutschen Markt. Grundlage der Bewertungen ist ein umfassender Kriterienkatalog. Dieser Kriterienkatalog basiert auf den Erfahrungen der BARC-Mitarbeiter aus über zehn Jahren Softwaretests sowie Softwareauswahl- und Beratungsprojekten.

Da BARC weder Software vertreibt noch implementiert und Anbieter auch keine Gebühren für eine Aufnahme in die Studie zahlen müssen, ist gewährleistet, dass Bewertungen und Empfehlungen nicht von versteckten Interessen beeinflusst werden.

Kunden der Studie erhalten 12 Monate Zugriff auf das Informationsangebot, bestehend aus Produktbewertungen, Strategiepapieren, Marktanalysen sowie Neuigkeiten aus dem DW-Markt. Dies schließt alle in dem Zeitraum durchgeführten Aktualisierungen und Ergänzungen zur Studie mit ein.





Copyright © BARC GmbH 2009. Alle Rechte vorbehalten.

Business Application Research Center (BARC)

Steinbachtal 2b  
97082 Würzburg

Tel.: +49 931 8806510

E-Mail: [info@barc.de](mailto:info@barc.de)

[www.barc.de](http://www.barc.de)