

IBM InfoSphere Information Server
Version 11 Release 3

Integrationsszenarios



IBM InfoSphere Information Server
Version 11 Release 3

Integrationsszenarios



Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen und Marken“ auf Seite 31 gelesen werden.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs
IBM InfoSphere Information Server, Version 11 Release 3, Integrations Scenarios,
IBM Form SC19-4311-00,
herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2010, 2014

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:
TSC Germany
Kst. 2877
September 2014

© Copyright IBM Corporation 2010, 2014.

Inhaltsverzeichnis

| | |
|--|-----------|
| InfoSphere Information Server - Integrationsszenarios | 1 |
| Anwendungsszenario für die Datenintegration erstellen. | 5 |
| Szenario für die Modernisierung eines Data-Warehouse mit dem Schwerpunkt Datenqualität | 7 |
| Datenqualität und Überwachungsintegration | 8 |
| IBM Industry Models | 9 |
| Häufig gestellte Fragen | 10 |
| Schlüsselszenario-Eingaben | 11 |
| Schlüsselszenario-Eingabeprozesse. | 12 |
| Schlüsselszenario-Ausgaben | 17 |
| | |
| Anhang A. Eingabehilfen in den Produkten | 23 |
| | |
| Anhang B. Kontaktaufnahme mit IBM | 25 |
| | |
| Anhang C. Auf Produktdokumentation zugreifen. | 27 |
| | |
| Anhang D. Feedback zur Produktdokumentation geben | 29 |
| | |
| Bemerkungen und Marken | 31 |
| | |
| Index | 37 |

InfoSphere Information Server - Integrationsszenarios

Bei der Informationsintegration handelt es sich um eine komplexe Aktivität, die alle Bereiche eines Unternehmens betrifft. Diese Integrationsszenarios, die sich mit den allgemeineren geschäftsbezogenen Problemstellungen bei der Integration befassen, zeigen auf, wie Sie IBM® InfoSphere Information Server und die InfoSphere Foundation Tools-Komponente auf integrative Weise zusammen bereitstellen und verwenden können. Die Integrationsszenarios befassen sich mit der Datenqualität in einer Data-Warehouse-Implementierung.

Herausforderungen bei der Datenintegration

Unternehmen sind heute mit einer großen Anzahl informationsbezogener Herausforderungen konfrontiert: verschiedenartige und oft unbekannte Datenqualitätsprobleme, Meinungsverschiedenheiten hinsichtlich der Bedeutung und des Kontexts von Informationen, Verwaltung mehrerer komplexer Umsetzungen, Nutzung vorhandener Integrationsprozesse anstelle von Mehraufwand, ständig wachsende Datenmengen, kleiner werdende Verarbeitungsfenster sowie steigende Anforderungen an die Überwachung und Sicherheit, um die Einhaltung von nationalem und internationalem Recht zu gewährleisten.

Unternehmen müssen Informationen und Systeme in den Unternehmensdomänen mithilfe einer integrierten Informationsinfrastruktur optimieren und miteinander verbinden. Informationen, die keine Verbindung aufweisen, führen dazu, dass IT-Abteilungen nicht schnell auf neue Informationsanforderungen von Geschäftsbetzern und Führungskräften reagieren können. Wenn nur wenige Tools oder Ressourcen für das Aufzeichnen der Informationsmassen zur Verfügung stehen, stellt auch die Überwachung der Datenqualität und die ständige Anwendung der Geschäftsregeln eine schwierige Aufgabe für die Unternehmen dar. Infolgedessen bleiben die Informationen weiter im gesamten Unternehmen unter einer Vielzahl ungeordneter Kategorien und nicht kompatibler Beschreibungen verstreut.

Nachfolgend sind einige der wesentlichen Probleme der Datenintegration aufgeführt:

- Die Quellenmetadaten der Unternehmensanwendungen sind nicht in einem bestimmten Bereich anschaulich angeordnet, wodurch die tatsächliche Verfügbarkeit der Metadaten nicht ermittelt werden kann. Die ungeordneten Daten können auch traditionelle Quellen enthalten, die Metadaten, falls überhaupt vorhanden, oftmals nicht über eine Standard-API bereitstellen.
- Die Masterreferenzdaten, Namen und Adressen von Lieferanten und Kunden sowie Teilenummern und Beschreibungen weichen in den verschiedenen Anwendungen und in den doppelten Quellen dieser Daten voneinander ab.
- Hunderte von ETL-Jobs müssen geschrieben werden, um die Daten aus allen Quellen in die neue Zielanwendung zu versetzen.
- Vor dem Laden der Daten sind Datenumsetzungen erforderlich, damit die Daten in die neuen Umgebungsstrukturen eingefügt werden können.
- Die Fähigkeit zur Handhabung großer Mengen von Daten, die verarbeitet und termingenau abgeschlossen werden können, ist äußerst wichtig. Unternehmen benötigen die entsprechende Infrastruktur, die Unterstützung für die Ausführung der Umsetzungs- und Datenabgleichsroutinen auf Anforderung bietet.

- Es ist keine konsolidierte Ansicht der Datenqualität für das gesamte Unternehmen verfügbar.

InfoSphere Information Server-Integrationslösung

InfoSphere Information Server- und InfoSphere Foundation Tools-Komponenten wurden eigens konzipiert, um Unternehmen bei der Bewältigung der Anforderungen an die Datenintegration zu unterstützen und sie bieten eine robuste Informationsarchitektur, die eine Nutzung der vorhandenen IT-Investitionen ermöglicht. Die Lösung bietet eine bewährte Methode zur Angabe geschäftskritischer Informationen: Art, Zeit und Ort, zu der bzw. an dem sie zur Verfügung gestellt werden sollen. Des Weiteren können bei dieser Lösung die Datenverwaltungsprozesse und Governanceverfahren festgelegt und die Verwendung der Informationen kann abgeglichen werden, um sie der Geschäftsstrategie des Unternehmens anzupassen.

InfoSphere Foundation Tools-Komponenten unterstützen Ihr Unternehmen bei der Profilierung, Modellierung, Definition, Überwachung und Regelung seiner Informationen. Durch die Integration der über die InfoSphere Foundation Tools-Komponenten bereitgestellten Lösungen kann Ihr Unternehmen seine Informationsinfrastruktur ermitteln und entwerfen und mit der Erstellung vertrauenswürdiger Informationen im gesamten Unternehmen beginnen.

Die IBM InfoSphere Information Server-Plattform besteht aus mehreren Produktmodulen, die Sie, wie in Abb. 1 auf Seite 3 dargestellt, zusammen oder einzeln im Integrationsframework Ihres Unternehmens implementieren können. InfoSphere Information Server ist für die flexible Integration in vorhandene Datenintegrationsprozesse des Unternehmens konzipiert, um dem aus Erkennung, Entwurf und Governance bestehenden ständigen Zyklus zur Unterstützung von Unternehmensprojekten gerecht zu werden.

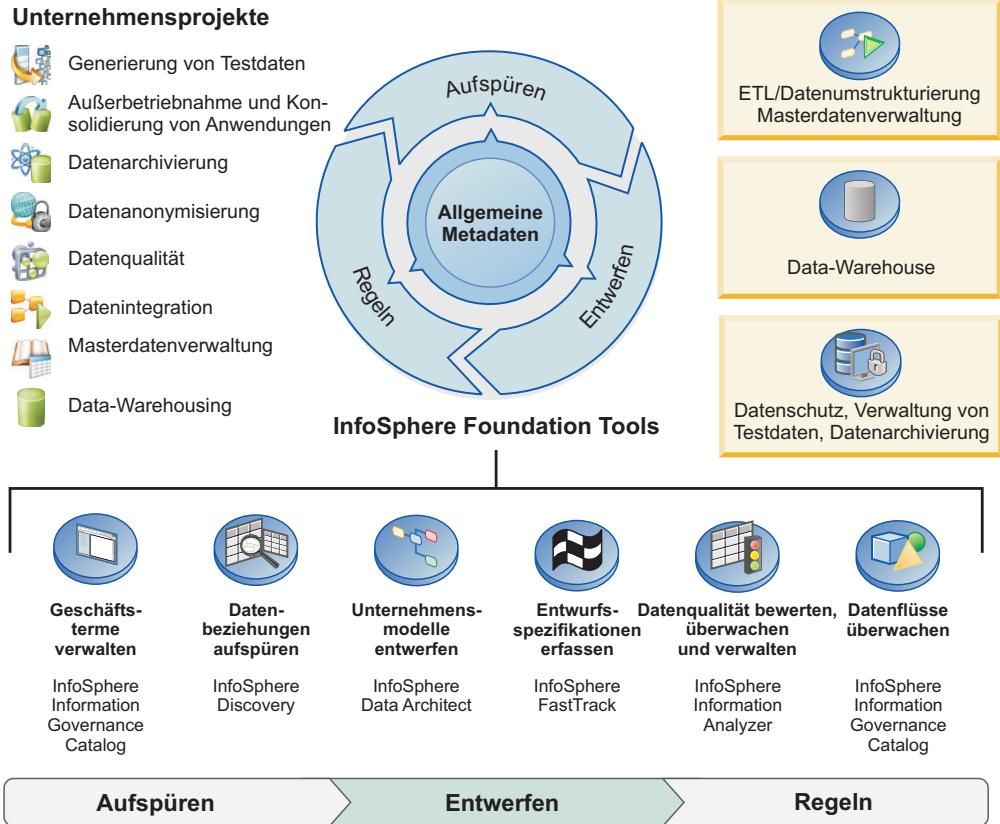


Abbildung 1. Die InfoSphere Information Server-Plattform unterstützt Ihre Datenintegrationprozesse.

Abb. 2 auf Seite 4 zeigt die Komponenten sowie die Metadaten an, die sie generieren, lesen und gemeinsam nutzen.

Der Prozess beginnt in der Regel mit der Definition von Datenmodellen. Ein Unternehmen kann Informationen aus IBM Industry Data Models importieren (verfügbar in InfoSphere Data Architect). Diese umfassen auch ein Glossar sowie ein logisches und physisches Datenmodell. Das Glossarmodell enthält Tausende von standardisierten Termen, die zum vorab stattfindenden Füllen von InfoSphere Information Governance Catalog mit Daten verwendet werden können. Unternehmen können IBM Industry Data Models modifizieren und erweitern, um die Anwendung an die jeweiligen Geschäftsanforderungen anzupassen.

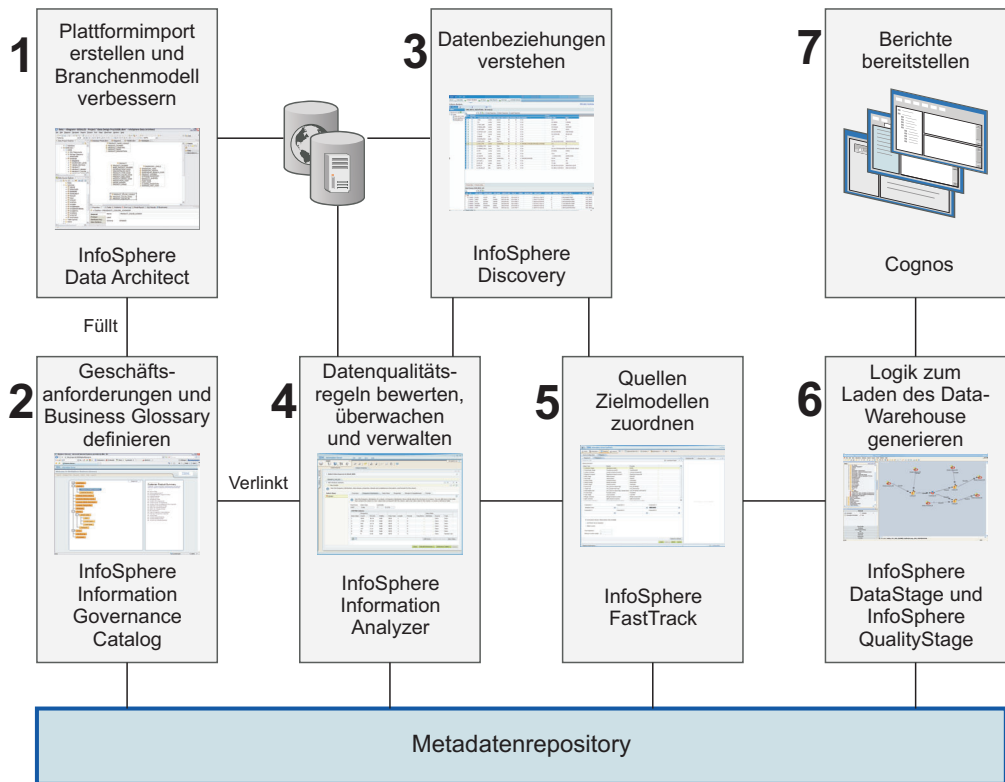


Abbildung 2. InfoSphere Information Server-Produktmodule

Nach der Definition der Datenmodelle und der Anwendung des Geschäftskontexts führt der Analytiker einen Datenerkennungsprozess für die Quellsysteme aus, um das neue Zieldatenmodell mit Daten zu füllen. Während des Erkennungsprozesses kann der Analytiker wesentliche Beziehungen, Umsetzungsregeln und Geschäftsobjekte ermitteln, mit denen das Datenmodell erweitert werden kann, wenn diese Geschäftsobjekte nicht zuvor über IBM Industry Data Models definiert wurden.

Der Analytiker kann die Arbeit anhand der ermittelten Informationen erweitern, um sich auf die Bewertung der Datenqualität zu konzentrieren und sicherzustellen, dass Unregelmäßigkeiten dokumentiert, Referenztabellen erstellt und Datenqualitätsregeln definiert werden. Er kann den Dateninhalt mit den etablierten Glossartermen verlinken, um eine entsprechende Kontext- und Datenabstammung sicherzustellen, Entwicklern Analyseergebnisse und abgeleitete Modelle bereitstellen und die Datenqualitätsregeln testen und implementieren. Wenn die Datenqualitätsregeln auf Daten aus Quellsystemen angewendet werden, können Ausnahmen der Regeln in IBM InfoSphere Data Quality Console überwacht werden.

Der Analytiker kann jetzt mit der Erstellung der Mapping-Spezifikationen beginnen, die als Eingabe bei den ETL-Jobs für die neue Anwendung verwendet werden. Er definiert mithilfe des Geschäftskontexts, der ermittelten Informationen und der Ergebnisse der Datenqualitätsbewertung die entsprechenden Umsetzungsregeln, die für die Konvertierung der Datenquellen in das für das IBM Industry Data Model-Ziel korrekte Format erforderlich sind. Während dieses Prozesses definiert der Analytiker nicht nur die entsprechenden Umsetzungsregeln für das Unternehmen, sondern gegebenenfalls auch die direkte Beziehung zwischen den Geschäfts-terminen und ihre Darstellung in physischen Strukturen. Diese Beziehungen können

anschließend in InfoSphere Information Governance Catalog für den späteren Einsatz und zum besseren Verständnis der Assetbeziehungen publiziert werden.

Die Geschäftsspezifikation dient jetzt als Langzeitdokumentation sowie auch als direkte Eingabe bei der Generierung der ETL-Jobs in IBM InfoSphere DataStage. Die definierten Geschäftsregeln werden direkt in den ETL-Job eingeschlossen, entweder als Code oder als annotierte unerledigte Tasks, die vom Entwickler abzuschließen sind. Ist der InfoSphere DataStage-Job bereit, kann der Entwickler auch durch die Verwendung von IBM InfoSphere Information Services Director entscheiden, ob der gleiche Stapelprozess wie bei einer SOA-Komponente verwendet werden soll.

Während dieses Prozesses werden Metadaten als natürliche Konsequenz der Verwendung aller InfoSphere Information Server-Module generiert und verwaltet. Die InfoSphere Information Server-Plattform nutzt während des gesamten Integrationsprozesses gemeinsam mit allen benutzerspezifischen Rollen relevante Metadaten. Aufgrund dieser eindeutigen Architektur erfordert die Verwaltung der Metadaten nur geringe manuelle Pflege. Lediglich die Metadaten anderer Anbieter erfordern Verwaltungstasks wie das Definieren der Beziehungen mit den Metadatenobjekten von InfoSphere Information Server. Administratoren und Entwickler, die eine Anzeige sowohl der Metadatenassets von InfoSphere Information Server als auch der Metadatenassets anderer Anbieter benötigen, können InfoSphere Information Governance Catalog zur Abfrage und Analyse dieser Informationen aus dem gemeinsam genutzten Repository sowie zur Berichterstellung verwenden.

Anwendungsszenario für die Datenintegration erstellen

IBM InfoSphere Information Server weist eine vereinheitlichte Suite mit Produktmodulen auf, die für die Optimierung des Erstellungsprozesses für eine Datenintegrationsanwendung konzipiert sind.

Die InfoSphere Information Server-Plattform bietet eine umfassende, integrierte Architektur, die auf einem einzelnen, gemeinsam genutzten Metadatenrepository basiert und bei der Informationen zwischen den Projektdatenintegrationstasks ohne Reibungsverluste gemeinsam genutzt werden können. Sie können die Informationsprüfung, den Zugriff und die Geschäftsverarbeitungsregeln für mehrere Projekte verwenden. Diese führen zu einem höheren Grad an Konsistenz, besserer Datensteuerung und verbesserter Effizienz. Abb. 3 auf Seite 6 verdeutlicht die folgenden Funktionen: Verstehen, Bereinigen, Umsetzen, Bereitstellen und Ausführen der vereinheitlichten Metadatenverwaltung.

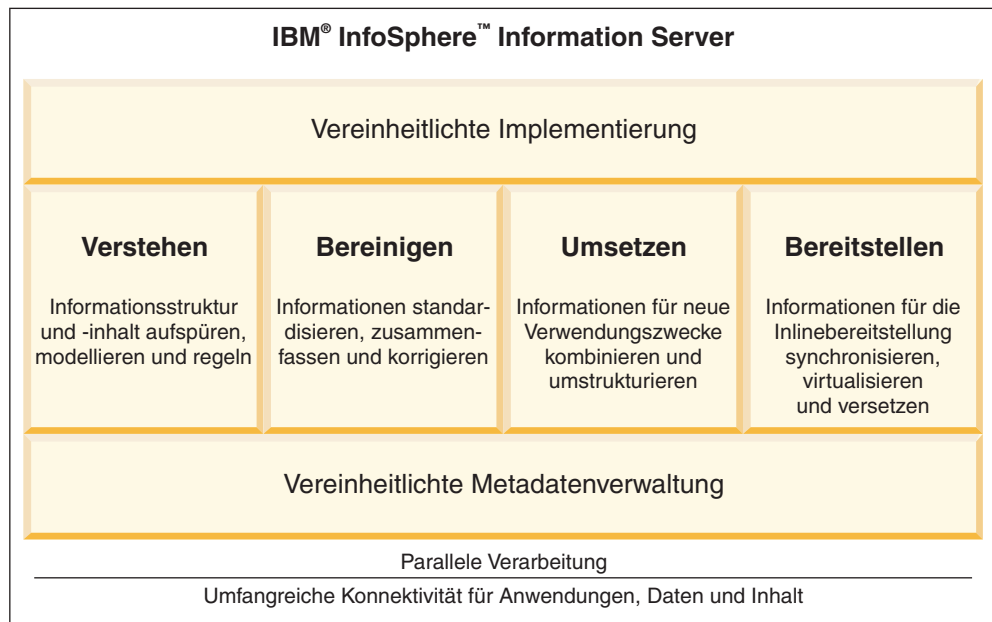


Abbildung 3. Integrationsfunktionen von InfoSphere Information Server

InfoSphere Information Server ermöglicht die Ausführung der fünf grundlegenden Integrationsfunktionen:

- *Daten verstehen.* InfoSphere Information Server unterstützt Sie beim automatischen Erkennen, Modellieren, Definieren und Regeln von Informationsinhalt und -struktur sowie beim Verstehen und Analysieren Bedeutung, der Beziehungen und der Abstammung von Informationen. Diese Funktionen dienen der besseren Veranschaulichung der Datenquellen und Beziehungen und der Definition der Geschäftsregeln, die das Risiko der Verwendung oder Vervielfältigung fehlerhafter Daten ausschließen.
- *Daten bereinigen.* InfoSphere Information Server unterstützt durch das Standardisieren, Prüfen, Abgleichen und Zusammenfassen von Daten die Qualität und Konsistenz von Informationen. Die Plattform unterstützt Sie beim Erstellen einer einzelnen, umfassenden und präzisen Informationsansicht, indem sie Datensätze innerhalb der Datenquellen oder datenquellenübergreifend abgleicht.
- *Daten in Informationen umsetzen.* InfoSphere Information Server setzt Informationen um und bereitet diese auf, um sicherzustellen, dass sie sich für neue Verwendungszwecke im geeigneten Kontext befinden. Die Anwendung stellt auch Funktionen für das Umsetzen und Versetzen großer Volumina komplexer Daten bereit, die für eigenständige ETL-Szenarios oder als Echtzeitdatenverarbeitungsengine für Anwendungen oder Prozesse verwendet werden können.
- *Korrekte Informationen zum richtigen Zeitpunkt bereitstellen.* InfoSphere Information Server ermöglicht das Virtualisieren, Synchronisieren oder Versetzen von Informationen an die Personen, Prozesse oder Anwendungen, die diese benötigen. Es unterstützt auch sensible serviceorientierte Architekturen, indem es die Implementierung von Umsetzungsregeln und deren Wiederverwendung als Services in mehreren Unternehmensanwendungen ermöglicht.
- *Vereinheitlichte Metadatenverwaltung ausführen.* InfoSphere Information Server basiert auf einer vereinheitlichten Metadateninfrastruktur, über die die an einem Datenintegrationsprojekt beteiligten Benutzerrollen Kenntnisse gemeinsam nutzen können, einschließlich Geschäfts-, Betriebs- und Technikdomänen. Diese einheitliche gesteuerte Infrastruktur reduziert die Entwicklungszeit und bietet einen

permanenten Datensatz, der die Zuverlässigkeit der Informationen erhöhen und gleichzeitig den manuellen Koordinationsaufwand eliminieren kann.

Szenario für die Modernisierung eines Data-Warehouse mit dem Schwerpunkt Datenqualität

In diesem Szenario werden die Methoden für die Nutzung von IBM InfoSphere Information Server-Komponenten zur Wahrung der Datenqualität in einer Data-Warehouse-Umgebung beschrieben.

Es gibt drei typische Anwendungsfälle, bei denen die Datenqualität im Zusammenhang mit den Data-Warehouses bewertet und überwacht wird.

Greenfield

Erstellt ein neues Data-Warehouse über Aktivitäten wie Erkennung, Terminologie, Abstammung, Datenqualität, Datenmodellierung, Mapping, Datenumsetzung und Datenbereinigung.

Modernisierung

Modifiziert ein vorhandenes Data-Warehouse mit Aktivitäten wie Terminologie, Erkennung, Impactanalyse, Datenqualität, Datenmodellierung, Mapping, Datenumsetzung und Datenbereinigung oder fügt dem Data-Warehouse über diese Aktivitäten Daten hinzu.

Governance

Verwaltet und regelt Ihr vorhandenes Data-Warehouse über Aktivitäten wie Datenqualität, Stewardship, Terminologie, Abstammung und Impactanalyse.

Bei jedem Anwendungsfall liegen eine Reihe von Aktivitäten vor, die zur Gesamtlösung beitragen. Jede Aktivität verwendet mindestens ein Produktmodul im Kontext einer umfangreicheren Methodologie oder eines umfangreicheren Prozesses, die bzw. der die Initiative bildet. Für jede Aktivität oder Phase gibt es bestimmte Prozesseingaben, bestimmte im und außerhalb des Produkts auszuführende Tasks sowie bestimmte Prozessausgaben, die in nachfolgenden Aktivitäten oder Phasen verwendet werden. Die Datenqualität stellt nur eine dieser Aktivitäten dar.

Die bei diesen Anwendungsfällen möglichen Aktivitäten stellen nicht zwingenderweise eine starre Ereignissequenz dar. Es handelt sich dabei oft um iterative Aktivitäten, die häufig gleichzeitig auftreten, wobei die Ergebnisse einer Aktivität Auswirkungen auf eine andere Aktivität haben und anschließend weitere Arbeitsschritte erfordern.

Ein Data-Warehouse enthält beispielsweise Kunden- und Accountdaten. Es wird jedoch ein umfangreicherer Einblick in die Auswirkungen gewünscht, die Verkäufe und Marketing-Aktivitäten auf die Kunden und ihr Kaufverhalten haben. Eine Gruppe von Verkaufsverwaltungsquellen soll dem vorhandenen Data-Warehouse hinzugefügt werden. Bei der Ersterkennung werden vier Verkaufsverwaltungssysteme gefunden und für die Aufnahme zugeordnet. Bei der Prüfung der Datenqualität werden jedoch beachtliche Probleme während des Prüfens der Domänen in allen vier Systemen gefunden, die anzeigen, dass viele Felder keine Daten oder Kommentare anstelle von verwendbaren Daten enthalten. Bei der Prüfung der Geschäftsterminologie wird festgestellt, dass bei der Systemverwendung eine Diskrepanz vorliegt und dass zwei andere Systeme erforderlich sind. Die Geschäftsterme durchlaufen den Erkennungsprozess, um den vorhergehenden vier Systemen Beziehungen zuzuordnen. Die Datenqualitätsprüfung stellt anschließend sicher, dass es sich dabei tatsächlich um die erforderlichen Tabellen handelt. Ableitungen von

der Datenqualitätsprüfung werden anschließend den Datenarchitekten bereitgestellt, damit sie die Modellierung der neuen Data-Warehouse-Tabellen verbessern können.

Es gibt eine Reihe allgemeiner *Problempunkte*, die Sie in diesen Anwendungsfällen angehen können. Sie umfassen Folgendes:

- *Unklare Terminologie* beim Verwalten von Data-Warehouse-Informationen. Beispiel: Wo sind die Umsatzinformationen und erfüllt der Data-Warehouse-Inhalt die Geschäftserwartungen?
- *Unbekannte Änderungsauswirkungen*, die vorhandene Prozesse und das Geschäft stören können.
- *Unbekannte Abstammung von Informationen*, die sich negativ auf die Datenzuverlässigkeit auswirkt. Beispiel: Verwenden Sie die richtigen Datenquellen für die angeforderten Informationen?
- *Unbekannte Datenqualität* ist eine der Hauptursachen für das fehlende Vertrauen von Geschäftsbenutzern in ihre Daten.
- *Unbekanntes Stewardship*, wobei nicht klar ist, wer Kenntnisse über die Daten aufweist, wer die Aufrechterhaltung der Qualität sicherstellt und wer den Datenzugriff regelt.

In diesem Szenario soll ein Data-Warehouse erweitert oder *modernisiert* werden. Dabei kann es sich um ein IBM Data-Warehouse oder ein Data-Warehouse eines anderen etablierten Herstellers wie Teradata oder Oracle handeln. Dieses Data-Warehouse enthält bereits verschiedene Finanzdaten, die eine effektive Berichterstattung ermöglichen, doch es müssen jetzt Kundendaten hinzugefügt werden, um weitere Analyseinformationen bereitstellen zu können. Wie bei den meisten Unternehmen entwickelt sich dieses Data-Warehouse zu einem bedeutenden Bereich, in dem eine Kombination aus Finanz-, Kunden- und Verkaufsinformationen zu Analysezwecken gespeichert wird.

Datenqualität und Überwachungsintegration

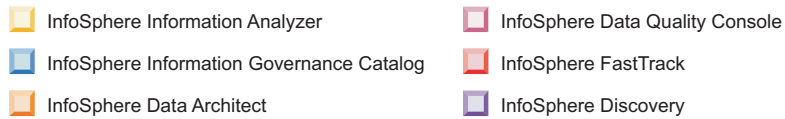
Eine Strategie für die Bewertung und Überwachung der Datenqualität behandelt die im Zusammenhang mit der Informationsqualität und -integrität auftretenden Probleme. Die Datenqualitätsprozeduren müssen darüber hinaus diesen in einem Data-Warehouse auftretenden Problemen gerecht werden.

IBM InfoSphere Information Governance Catalog, IBM InfoSphere Discovery, IBM InfoSphere Data Architect, IBM InfoSphere Information Analyzer, IBM InfoSphere FastTrack und IBM InfoSphere Data Quality Console verwenden vorhandene Informationsassets, um ein Data-Warehouse mithilfe der Informationsintegration auf der Grundlage einer Reihe von Business-Intelligence-Anforderungen oder Spezifikationen mit Daten zu füllen, die auf einem vorhandenen branchenspezifischem Modell oder einem entsprechenden Standard basieren können.

Anwendungsfall für Data-Warehouse

InfoSphere Information Analyzer kann die Probleme aufdecken, die im Zusammenhang mit der Informationsqualität und -integrität sowie der Erstellung von Datenqualitätsprozeduren oder -regeln in einer Mehrbenutzerumgebung zur Überwachung der Datenqualität über einen Zeitraum auftreten. Dieses Szenario ist eigenständig oder im Rahmen einer größeren Initiative verwendbar, beispielsweise in Data-Warehousing mit integrierter Datenqualität.

Schlüsselprodukte



Workflow

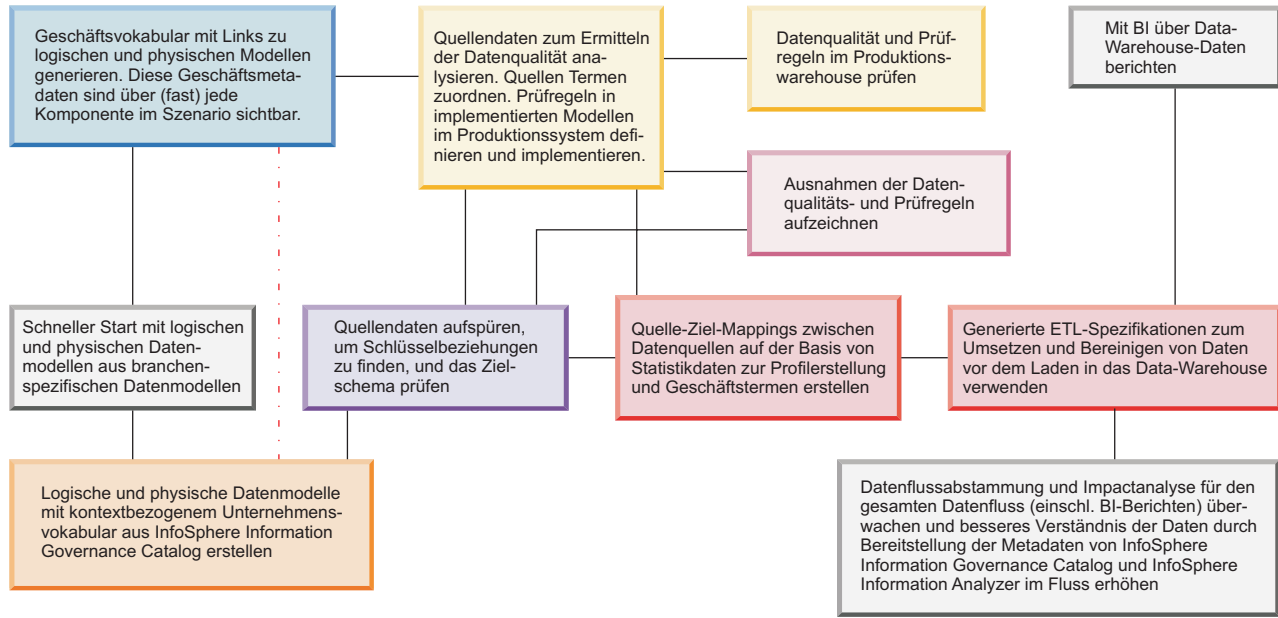


Abbildung 4. Integrationsworkflow für das Data-Warehouse

IBM Industry Models

Damit Unternehmen noch schneller Ergebnisse erzielen können, hat IBM das Wissen aus seiner langjährigen Erfahrung mit Informationsprojekten in verschiedenen Branchen jetzt in die IBM Industry Models gepackt.

Diese Modelle bieten ein umfassendes Geschäftsdatenmodell sowie Berichtsschablonen, die die wesentlichen Leistungsindikatoren, Messkriterien und Konformitätsfaktoren für die einzelnen Branchen zusammenfassen. IBM stellt diese branchenspezifischen Modelle für sechs Branchen bereit: das Bankwesen, die Finanzmärkte, das Gesundheitswesen, das Versicherungswesen, den Einzelhandel und das Vertriebswesen sowie die Telekommunikationsbranche.

Diese Modelle dienen der maßgeblichen Beschleunigung von Migrations- und Integrationsprojekten und bieten eine bewährte branchenspezifische Vorlage als Referenz für alle Projektteilnehmer. Quellensystemdaten können direkt in InfoSphere Information Server geladen werden, wodurch Zieldatenstrukturen und vordefinierte Geschäftsglossare bereitgestellt werden, mit denen die Entwicklung beschleunigt werden kann. Darüber hinaus stellen Business-Intelligence-Schablonen Schablonen für Berichte und Datenkuben in Cognos 8 Business Intelligence bereit. Mit der Verwendung von IBM Industry Models können Unternehmen ihre Projekte deutlich beschleunigen und Risiken vermindern sowie konventionelle unternehmensspezifische Probleme, die bei der Integration von Informationen üblicherweise auftreten, durch die Bereitstellung eines bewährten, neutralen Basismodells beseitigen.

Häufig gestellte Fragen

Wie bei jeder Initiative gibt es auch bei einer Data-Warehouse-Initiative eine Reihe häufig gestellter Fragen bezüglich der Datenqualität.

Diese Fragen beziehen sich auf folgende Aspekte:

Vergleich der Geschäfts- und Metadatendefinition mit der Datenrealität

Die Kenntnisse der Geschäftsbenutzer oder Fachleute können sich von denen der IT-Ressourcen (Architekten, Modellersteller und Entwickler) wesentlich unterscheiden. Auch kann das Verständnis der allgemeinen Geschäftsterminologie bei den verschiedenen Geschäftsgruppen sehr unterschiedlich sein.

- Besteht eine konkrete Vorstellung von der Geschäftsterminologie?
- Besteht eine konkrete Vorstellung von der Beziehung zwischen der Geschäftsterminologie und der Qualität der tatsächlichen Quelldaten, die geprüft oder zertifiziert werden muss?
- Bestehen Diskrepanzen zwischen der Geschäftsterminologie und den Metadaten (Quellenmetadaten oder Ziel-Warehouse-Modell)?
- Werden die tatsächlichen Metadaten und Daten zum Erkennen und Prüfen der Geschäftssemantik und Datenqualität verwendet?

Auswirkung von Änderungen

Beim Modernisieren eines vorhandenen Systems sind Kenntnisse der Auswirkungen der verschiedenen geplanten Änderungen unabdingbar.

- Was versteht man unter den vorgeschalteten Systemen (die dem zu ändernden System Daten zuführen) und den nachgeschalteten Systemen (Anwendungen oder Systeme, von denen die Informationen gelesen werden), die von der Änderung betroffen sind?
- Wer sind die Stewards?
- Liegen Prozesse oder Anwendungen wie Business-Intelligence-Berichte vor, die möglicherweise nicht fehlerfrei ausgeführt werden und modifiziert werden müssen?

Datenfokus

Die einem Data-Warehouse hinzugefügten Informationen stammen in der Regel aus mehreren unterschiedlichen Datenquellen oder aus Datenquellen, die auf viele Tabellen aufgeteilt sind.

- Liegt das Augenmerk auf Kernsystemen oder -tabellen, bestimmten Datentypen oder -klassen oder bestimmten Attributen oder Domänen?
- Werden Systeme, Quellen oder Entitäten als "Datensatzquelle" betrachtet?
- Wie wird mit der quellenübergreifenden Konsistenz oder mit Konflikten umgegangen oder wie werden diese gelöst?

Prüfung und Informationsbereitstellung

Initiativen lassen oftmals wenig Zeit für vorherige Überlegungen zur Datenqualität. Werden die Daten jedoch bereitgestellt und weisen sie nicht die gewünschte Datenqualität auf, führt dies zu hohen Korrektur- und Neuerstellungskosten.

- Wann wird die Datenqualität analysiert? Erfolgt die Analyse der Datenqualität erst nach der anfänglichen Erkennung und Prüfung oder während des gesamten Datenintegrationslebenszyklus?
- Welche Messkriterien sind für die Prüfung der Datenqualität entscheidend?

- Wie werden die Informationsergebnisse zur Datenqualität bereitgestellt?

Schlüsselszenario-Eingaben

Ist das Data-Warehouse neu oder wurde es modifiziert, um neue Geschäftsobjekte (beispielsweise Kundendaten) oder Datenquellen (beispielsweise Kundeninformationen aus dem Bestellsystem) hinzuzufügen, gelten für die Erweiterung des Glossars zur Integration neuer Terminologie, für die Erweiterung des Modells zur Aufnahme aller neuen Entitäten und Attribute und für die Erweiterung der Metadaten zur Aufnahme der neuen Quellen mit den zugehörigen Analyseinformationen (zu Beziehungen und Datenqualität) bestimmte Voraussetzungen.

Eines der Grundprinzipien für die Datenintegration ist die Möglichkeit, Informationen auszufüllen und als Metadaten zu jedem Prozess zu speichern.

Modelle

IBM bietet so genannte *3NF*-Datenmodelle (3. Normalform) für die Integration von großen Detaildatengruppen sowie vordefinierte analytische Modelle, so genannte TDW-Berichte (Tivoli Data Warehouse) oder Geschäftslösungsvorlagen (Business Solution Templates), die aus Maßeinheiten und Bemaßungen für Zusammenfassungen und für die Berichterstellung bestehen, um Ihnen die ersten Schritte zu erleichtern. IBM entwickelt auch Metadatenmodelle mit Wörterverzeichnissen für Geschäftsterme, die zum Erstellen von Glossaren und umfangreichen Projekten verwendet werden können. Alle IBM Modelle sind für die Beschleunigung der Entwurfserstellung und aus Abstammungsgründen einander zugeordnet. Die Modelle können sowohl mit InfoSphere Data Architect als auch mit CA ERwin Data Modeler geliefert werden.

Durch die vorhandene MetaBroker-Technologie können die physischen Modelle aus InfoSphere Data Architect als Metadateninhalt in das Metadatenrepository von IBM InfoSphere Information Server geladen werden.

Terminologie

Geschäftsterme sind bei der Beschreibung der Datentypen, mit denen Sie arbeiten, sowie in der in Ihrem Unternehmen üblichen Sprache maßgeblich. Diese Art der Terminologiedefinition kann nicht nur Terme zu den Zielsystemen umfassen, sondern auch Schlüsselinformationen, die das Unternehmen steuern, beispielsweise wesentliche Leistungsindikatoren oder erwartete Vorteile. Dieses Data-Warehouse kann beispielsweise das Rentabilitäts- oder Umsatzprotokoll enthalten. Die Terminologie steuert auch die Zusammenarbeit. Daher ist zu beachten, dass die einheitliche Sprache die Basis aller Verbindungen ist.

Nach dem Import der Informationen werden diese im Rahmen der Zusammenarbeit mit dem Unternehmen gemeinsam genutzt. Sie können Metadaten verwalten, um Standardgeschäftsterme und -beschreibungen des Unternehmens zu erfassen, die die Sprache der Geschäftsbutzer widerspiegeln. Unternehmen institutionalisieren die formelle Datenverwaltung oder die Datengovernance kann diese Terme publizieren, um sicherzustellen, dass alle Geschäftsbutzer konsistente Kenntnisse der verfügbaren Unternehmensinformationen besitzen, die auf Standardgeschäftsdefinitionen basieren.

IBM InfoSphere Governance Catalog dient als Basis für die Erstellung geschäftsorientierter Semantik, einschließlich Kategorien und Termen.

Wenn Sie andere Tools für den Import von Assets in Business Glossary verwenden, können Sie InfoSphere Governance Catalog verwenden, um ein Asset einem Term zuzuweisen. InfoSphere Governance Catalog wird in der Regel zum Zuweisen einer großen Anzahl Assets zu Termen verwendet. Da der Glossarinhalt im Metadatenrepository von InfoSphere Information Server gespeichert wird, können Sie über den Glossarinhalt mit den anderen Komponenten der InfoSphere Information Server-Suite interagieren.

Schlüsselszenario-Eingabeprozesse

Die folgenden Themen beschreiben die Schlüsselszenario-Eingabeprozesse.

Quellenmetadaten erkennen

Für dieses Szenario sollten Sie mit den Anforderungen des Unternehmens sowie der Terminologie vertraut sein, die zum Beschreiben der Geschäftsanforderungen verwendet wird. Sie verstehen Begriffe wie Kunde oder Standort und können diese unternehmensweit nutzen. Außerdem kennen Sie die Struktur Ihres Ziel-Data-Warehouse. Es ist notwendig, Ihre Kenntnisse bezüglich der Struktur und des Inhalts der tatsächlichen ankommenden Daten zu erweitern oder zu verbessern.

IBM InfoSphere Discovery

IBM InfoSphere Discovery wird zum Ermitteln der Umsetzungsregeln verwendet, die auf ein Quellensystem angewendet wurden, um ein Ziel, beispielsweise ein Data-Warehouse oder einen Speicher für operative Daten mit Daten zu füllen. Nach ihrer korrekten Definition stellen diese Geschäftsobjekte und Umsetzungsregeln die wesentliche Eingabe für informationsgerichtete Projekte wie die Datenintegration, die IBM InfoSphere-Masterdatenverwaltung und die Archivierung dar.

InfoSphere Discovery analysiert die Datenwerte und Muster aus mindestens einer Quelle, um diese verdeckten Korrelationen zu erfassen, und veranschaulicht sie in der Anzeige. InfoSphere Discovery wendet heuristische und ausgereifte Algorithmen an, um den vollständigen Bereich der Datenanalysetechniken zu nutzen: Überschneidung von Daten aus einer oder mehreren Quellen und Beziehungsanalyse, erweiterte Erkennung von Abgleichsschlüsseln, logische Umsetzungserkennung usw. Das Programm umfasst den größten Bereich von Geschäftsdatenquellen: relationale Datenbanken, hierarchische Datenbanken und alle strukturierten Datenquellen im Textdateiformat.

InfoSphere Discovery durch Suchen nach Datenwerten und Mustern zum Veranschaulichen von Datenbeziehungen verwenden

In diesem Szenario werden drei verteilte Datenquellen miteinander kombiniert, um zugehörige Informationen zu Kundennamen, Adressen und Steuerkennungen zu ermitteln.

Führen Sie mit InfoSphere Discovery Folgendes aus:

1. Erstellen Sie ein Quellendatenerkennungsprojekt. In Quellendatenerkennungsprojekten prüfen Sie Datenwertüberschneidungen zwischen Tabellen, die sich in einem oder mehreren Datasets befinden. Darüber hinaus können Sie ein vereinheitlichtes Schema erstellen und primäre Fremdschlüssel für das vereinheitlichte Schema erkennen. Die Ergebnisse beinhalten eine Zusammenfassung aller Tabellen und Spalten in den Datasets, die Anzahl exklusiver Spalten, den Prozentsatz der Wertüberschneidungen, die Anzahl Tabellen und Spalten mit überschneidenden Daten sowie weitere ausführliche Statistikdaten einschließlich der Ansichten der sich tatsächlichen überschneidenden Daten selbst.

2. Importieren Sie eine Tabellengruppe aus jeder Datenquelle.
3. Erstellen Sie Datasets. Ein Dataset ist eine Objektgruppe von Datenbanktabellen und Textdateien, die verarbeitet, analysiert oder zugeordnet werden sollen. Es kann eine beliebige Anzahl Datenbanktabellen und begrenzte oder positionsgebundene Textdateien aus einer beliebigen Anzahl ODBC-Verbindungen enthalten.
4. Führen Sie die Datenerkennung aus. Bestimmte Datenbanktabellen enthalten korrekte Datentypdefinitionen in Metadaten. Andere Datenbanktabellen enthalten jedoch falsche oder unvollständige Metadaten und Textdateien, die keine Datentypen definieren.

In diesem Schritt berechnet InfoSphere Discovery die Statistikdaten zu den in den Datasets enthaltenen Daten und zeigt sie zusammen mit den Informationen zu den Metadaten an. InfoSphere Discovery prüft darüber hinaus auch alle VARCHAR-Zeichenfolgen, um festzustellen, ob sie Datums-/Zeitwerte oder numerische Werte enthalten, und weist ihnen den geeigneten Datentyp zu.

Nach dem Ausführen der Datenerkennung müssen Sie die Spaltenanalysedaten prüfen. Prüfen Sie die Datentypen und nehmen Sie alle erforderlichen Änderungen vor, beispielsweise die Anpassung einer Spaltenlänge oder die Korrektur eines in den Metadaten definierten falschen Datentyps. Markieren Sie die für Ihr Projekt maßgeblichen Spalten als kritische Datenelemente. Nach jeder Modifizierung der Tabellen in einem Dataset müssen Sie die Spaltenanalyse erneut ausführen.

5. Erkennen Sie primäre Fremdschlüssel. InfoSphere Discovery importiert automatisch die Beziehungen der primären Fremdschlüssel, wenn diese in den Metadaten einer Tabelle definiert sind. Sind die Beziehungen nicht definiert, sucht InfoSphere Discovery Spaltenübereinstimmungen durch Prüfen der tatsächlichen Daten. Spaltenübereinstimmungen mit den höchsten Trefferraten und der höchsten Selektivität werden automatisch als primäre Fremdschlüssel zugeordnet.

Prüfen Sie nach dem Ausführen der Task zur Erkennung der primären Fremdschlüssel die Genauigkeit der Ergebnisse. Wenn Sie erkannte Spaltenübereinstimmungen oder primäre Fremdschlüssel definieren oder modifizieren, führen Sie die Funktion zur Erkennung der primären Fremdschlüssel erneut aus.
6. Erkennen Sie Datenobjekte. Ein Datenobjekt bietet eine konzeptionelle Möglichkeit zum Anzeigen von Tabellenbeziehungen in einem Dataset. Ein Datenobjekt stellt eine Gruppe von Tabellen dar, die über primäre Fremdschlüssel miteinander in Beziehung stehen. Ein Dataset kann viele Datenobjekte enthalten und jedes Datenobjekt kann wiederum aus einer beliebigen Anzahl Tabellen bestehen. Eine Tabelle kann sowohl ein übergeordnetes als auch untergeordnetes Element darstellen und kann daher in mehreren Datenobjekten vorliegen. In der Regel beziehen sich diese Datenobjekte auf Schlüsselkonzepte im Glossar oder auf die neuen Entitäten, die Sie im Data-Warehouse hinzufügen oder modernisieren.
7. Erkennen Sie Überschneidungen. Der Schritt zum Erkennen von Überschneidungen stellt ein klares Bild der sich überschneidenden Daten in Ihren Quellen bereit. Prüfen Sie die Spaltendaten, um sicherzustellen, dass die erkannten Überschneidungen sinnvoll und gültig sind. Löschen Sie alle offensichtlichen Abweichungen in den Wertüberschneidungsdetails. Wenn bei einer bestimmten Überschneidung Zweifel an den Daten bestehen, zeigen Sie die tatsächlichen Daten mithilfe der Spaltenzusammenfassung und der Spaltenüberschneidungen an. Definieren Sie neue Überschneidungen und ermitteln Sie die Statistikdaten mithilfe von InfoSphere Discovery.

Markieren Sie die für Ihr Projekt maßgeblichen Spalten als kritische Datenelemente.

Ist die Erkennung der Überschneidungen abgeschlossen, können Sie mit der Definition eines vereinheitlichten Schemas beginnen. Wenden Sie diese Methode an, wenn die aktuellen Quelldatenmodellstrukturen wiederverwendet werden sollen und können, wenn eine schnelle Prototyperstellung erforderlich ist oder wenn keine bestimmten Anforderungen hinsichtlich des Zieldatenmodells bestehen, die einen sehr eindeutigen Entwurf erfordern. Selbst in letzterem Fall kann es vorteilhaft sein, ein vereinheitlichtes Schema als Arbeitsbasis für standardisierte und abgestimmte Daten aus den Datenquellen zu erstellen, bevor mit dem Konsolidieren und dem Laden von Daten in das Data-Warehouse begonnen wird.

8. Erstellen Sie eine Zieltabelle für ein vereinheitlichtes Schema. Definieren Sie das Zieltabellenschema. Füllen Sie die Tabelle mit der erforderlichen Anzahl Zielspalten. Sie können Datentypen modifizieren und die Spaltenreihenfolge ändern.
9. Definieren Sie das Quellenmapping. Ordnen Sie Tabellenspalten Zielspalten zu und erstellen Sie für jede Tabellenspalte so viele Filter, bis jede Zielspalte die geeigneten Daten aus den relevanten Tabellen enthält. Rufen Sie gegebenenfalls die Registerkarte für das Zieltabellenschema erneut auf, um die Zielspaltenmerkmale zu optimieren.
10. Führen Sie eine Analyse vereinheitlichter Schemata aus. Konsolidieren Sie Daten aus verschiedenen Quellen in der Ergebniszieltabelle. Mit den in dieser Anzeige enthaltenen Statistikdaten- und Datenvoranzeigen werden Übereinstimmungsbedingungen definiert, die im nächsten Schritt (Abgleichen und Zusammenfassen) verwendet werden.
Prüfen Sie die resultierenden Statistikdaten. Es werden Statistikdaten zu den Ergebnisspalten in der Zieltabelle als auch zu jeder einzelnen in der Zieltabelle verwendeten Tabellenspalte angezeigt.
Grenzen Sie die Zieltabelle gegebenenfalls ein und rufen Sie dazu das Fenster für das Zieltabellenschema erneut auf. Fügen Sie dort Zieltabellenspalten hinzu oder entfernen Sie sie. Sie können auch das Quellenmapping ändern.
11. Führen Sie die Abgleichs- und Zusammenfassungsanalyse aus. Definieren Sie Übereinstimmungsbedingungen für jede Zielspalte. Definieren Sie auch Konflikterkennungsregeln, um zu ermitteln, ob die Werte in den einzelnen Gruppen als Konflikte betrachtet werden oder nicht. Fassen Sie die Ergebnisse zusammen und definieren Sie Konfliktlösungsregeln, die den besten Wert aus den Alternativen für doppelte Datensätze auswählen. Das vereinheitlichte Schema ist jetzt fertig gestellt.
12. Exportieren Sie Daten aus InfoSphere Discovery in das Metadatenrepository. Sie können jetzt die Ergebnisse aus Ihrem vereinheitlichten Schema unter Verwendung des Import-/Exportmanagers in DBM MetaBroker in das Metadatenrepository exportieren, um InfoSphere Information Analyzer die Ergebnisse zur Verfügung zu stellen.

Datenqualität bewerten

Für eine effektive Nutzung des Data-Warehouse reicht es nicht aus, die Zusammenhänge zwischen den Tabellen und Datenspalten zu verstehen. Sie müssen die Daten bewerten, damit sie nachgeschalteten Prozessen effektiv bereitgestellt werden können.

Bei dieser Bewertung werden Sie die Inhalte gründlich analysieren, um alle in den Informationen enthaltenen Unregelmäßigkeiten zu ermitteln. Bei dieser Qualitätsbewertung können Sie die Fehler in den Daten mit Anmerkungen versehen, damit

in nachgeschalteten Prozessen die entsprechenden Maßnahmen ergriffen werden. Diese Anmerkungen werden auf der Metadatenstufe gemeinsam genutzt. Wenn Sie Datenprobleme finden und diese mit Anmerkungen versehen, können Sie Datenmonitore einsetzen, um in den Daten weiter nach diesen Mustern zu suchen. Dieser Prozess gewährleistet die Erhaltung der Datenqualität beim Versetzen der Daten von den unveränderten Quellsystemen auf das zukünftige Ziel.

Die Datenerkennung bietet einen Ausgangspunkt, von dem aus die Erfüllung der Kriterien für die Datenqualität sichergestellt und sowohl beim ersten Laden des Data-Warehouse mit diesen neuen Informationen als auch bei den fortlaufenden Arbeiten im Data-Warehouse berücksichtigt werden können. Dennoch wird empfohlen, die Datenqualität genau zu bewerten.

Dabei sind zwei Schwerpunkte gesetzt:

- Zunächst prüfen Sie Informationen zu Unregelmäßigkeiten, Schlüsselreferenztabellen für gültige Werte, Abstimmungen von Quellen mit der Geschäftsterminologie und Anforderungen an die Bedingungen für die Datenprüfung und kommunizieren diese Informationen.
- Anschließend entwerfen und testen Sie Regeln für die Datenprüfung, Benchmarks und Maßnahmen, die auf den Data-Warehouse-Ladeprozess und nach dem Laden auf das Data-Warehouse selbst angewendet werden.

Qualität der Quelldaten analysieren

Dies ist der erste Schritt bei der Bewertung, Prüfung und Kommunikation von Informationen zur Qualität von Kerndaten im Hinblick auf die Quellen, die dem Data-Warehouse zugeführt werden.

Datenqualität mithilfe von IBM InfoSphere Information Analyzer bewerten und sensible Daten definieren

Das Schema aus der InfoSphere Discovery-Analyse kann von InfoSphere Information Analyzer verwendet werden, um die Implementierung zu beschleunigen und die Genauigkeit der Datenqualitätsregeln für die Analyse, Überwachung und Verwaltung der Qualität im Laufe der Zeit zu verbessern.

Bei diesem Szenario stellen die Ergebnisse des Erkennungsprozesses die anfänglichen Datenquelleninformationen mit Metadaten zu Kundennamen, Adressen und Steuerkennungen bereit.

Führen Sie mit InfoSphere Information Analyzer Folgendes aus:

1. Erstellen Sie ein InfoSphere Information Analyzer-Projekt oder öffnen Sie ein vorhandenes Projekt.
2. Fügen Sie Datenquellen hinzu, für die InfoSphere Discovery eine Analyse im Information Analyzer-Projekt durchgeführt hat.
3. Importieren Sie das InfoSphere Discovery-Schema aus dem Metadatenrepository aus dem vorhergehenden Szenario (Quellenmetadaten erkennen) in InfoSphere Information Analyzer als Ziel für die Datenqualitätsregeln.
4. Prüfen Sie die Zusammenfassung der Analyseergebnisse aus InfoSphere Discovery im Rahmen der Spaltenanalysetask. Ignorieren Sie Felder, deren Verwendung im Data-Warehouse nicht relevant ist. Führen Sie die Spaltenanalyse für Felder aus, bei denen die Datenqualität für die Bewertung entscheidend ist. Wenn die Prüfung sensibler Daten wichtig ist, aktivieren Sie bei der Ausführung die Option für die erweiterte Datenklassifizierung.

5. Prüfen Sie Datenklassen auf die korrekte Datenklassifizierung. Beim Anzeigen und Prüfen von Spaltenanalyseergebnissen können Sie die vom System abgeleitete Datenklasse für die ausgewählte Spalte akzeptieren oder überschreiben. Diese Datenklassen stellen Schlüsselkomponenten bei der Steuerung weiterer Analysen dar.
6. Analysieren Sie Detailanalyseergebnisse auf der Basis von Datenklassifizierungen im Hinblick auf bestimmte Bedingungen und geeignete Feldanmerkungen.
 - **Kennungen** - Auf Duplikate und ungültige Formate prüfen
 - **Anzeiger** - Auf ungültige und inkonsistente Markierungswerte oder Datenabweichungen prüfen. Es werden Referenztabellen gültiger Werte generiert.
 - **Codes** - Auf ungültige Codewerte und Standardcodewerte prüfen. Es werden Referenztabellen gültiger Werte generiert.
 - **Mengen** - Auf Daten prüfen, die außerhalb der gültigen Bereiche liegen. Es werden Referenztabellen gültiger Bereiche generiert.
 - **Datumsangaben** - Auf Daten prüfen, die außerhalb der gültigen Bereiche liegen. Es werden Referenztabellen gültiger Bereiche generiert.
 - **Text** - Gibt Textfelder an, die für das Data-Warehouse erforderlich sind. Sucht nach problematischen Formaten. Verwendet Datenregeln, um nach bekannten Bedingungen zu suchen.
 - **Sensible Daten** - Sucht nach ungültigen Bedingungen.
 - **Sonstige Bedingungen** - Verwendet Datenregeln zur Erweiterung der Analyse auf andere unternehmensdefinierte Bedingungen wie gültige Wertekombinationen, Datenreihenfolge, Berechnungsausdrücke, komplexe Datenformatierung usw.
7. Weisen Sie während der Analyse den in IBM InfoSphere Governance Catalog etablierten, erkannten Geschäftstermen Spalten zu. Diese datenorientierte Ansicht unterstützt Sie bei der Prüfung und stellt sicher, dass die korrekten Daten mit den richtigen Geschäftskonzepten verlinkt werden.
8. Teilen Sie Detailanalyseergebnisse mit und prüfen Sie sie. Dabei kann es sich um einen iterativen Überarbeitungszyklus mit Fachleuten handeln. Beispiele für in diesem Schritt gefundene Bedingungen:
 - Diskrepanz bei den Daten (die gefundenen Daten weichen beispielsweise von den für das Data-Warehouse erforderlichen Daten ab oder es liegen keine Daten vor, die eine bestimmte Anforderung erfüllen. Zur Behebung der Diskrepanzen müssen möglicherweise weitere Quellen hinzugezogen werden und die Durchführung eines weiteren Zyklus durch IBM InfoSphere Discovery kann erforderlich sein.
 - Kenntnislücken (Beispiel: kein Benutzer weiß, was ein bestimmter Code bedeutet oder was als gültig betrachtet wird)
 - Probleme in Daten (Beispiel: In den Daten liegen Probleme vor, die behandelt und dokumentiert werden müssen, oder Regeln müssen umgesetzt werden)
9. Publizieren Sie Analyseergebnisse. Sie können ein Analyseergebnis anzeigen oder es im Metadatenrepository publizieren. Diese Publizierung aktualisiert das Metadatenrepository mit den zusätzlichen Details, die über die Bewertung der Datenqualität ermittelt wurden. Diese Publizierung kann von Entwicklern, die mit IBM InfoSphere DataStage oder IBM InfoSphere QualityStage arbeiten, direkt verwendet werden, um die korrekten Ladeprozesse aus den angegebenen Datenquellen in das Zielschema aufzunehmen, das über IBM InfoSphere Discovery oder das Data-Warehouse definiert wurde.

10. Entwerfen Sie neue Prüfregeln und Messkriterien zur Datenqualität oder verwenden Sie vorhandene Prüfregeln und Messkriterien wieder, die auf den Ergebnissen der Datenanalyse basieren. Für die fortlaufende Prüfung werden nicht alle Felder ausgewählt.
11. Verwenden Sie die geplanten Quelldaten, um neue Datenregeln und Messkriterien zu testen und zu prüfen. Diese Regelprüfungen orientieren sich an den Anforderungen des Data-Warehouse. Die Ergebnisse stellen einen Teil des iterativen Prüfungszyklus mit Fachleuten dar und werden zur Erstellung erster Benchmarks verwendet. Die Regelprüfungen werden auf die im obigen Schritt zum Mitteilen und Prüfen von Detailanalyseergebnissen aufgeführten Elemente angewendet oder beheben Probleme mit diesen Elementen.
12. Publizieren oder implementieren Sie Datenregeln und Messkriterien. Diese Datenregeln und Messkriterien stellen die Grundlage für die fortlaufende Überwachung der Datenqualität dar.

InfoSphere Information Analyzer nimmt eine entscheidende Rolle im Integrationsprozess ein. Die Profilerstellungsergebnisse aus InfoSphere Information Analyzer können direkt in der Anwendung IBM InfoSphere FastTrack aufgerufen werden, in der die Spezifikationen definiert und dokumentiert sind. Dadurch werden umfassende und präzise Mapping-Spezifikationen ermöglicht. Diese Spezifikationen werden daraufhin als Eingabeanforderungen für ETL- und Bereinigungsjobs in InfoSphere DataStage und IBM InfoSphere QualityStage verwendet, die die zu entwickelnde Geschäftsanwendung unterstützen. Je mehr Informationen der Analytiker zu den wirklichen Datenstrukturen und -inhalten hat, desto genauer sind die Anforderungen für nachgeschaltete Entwickler.

Zugehörige Informationen:

Regelkonzepte und -methoden

Schlüsselszenario-Ausgaben

Die Beziehungsanalyse mit IBM InfoSphere Discovery und die Datenqualitätsanalyse mit IBM InfoSphere Information Analyzer erzeugt eine Reihe von Ausgaben, mit denen die Modernisierung des Data-Warehouse vereinfacht wird. Dies umfasst auch das bessere Mapping und die Integration in die neuen Datenstrukturen, Schlüsselreferenztabellen für die Prüfung von Ladevorgängen und Qualitätskontrollen für die laufende Überwachung der Datenqualität.

Nachfolgend sind die vier Hauptausgaben aufgeführt, die vom Prozess der Datenqualitätsbewertung bereitgestellt werden: gemeinsam genutzte Metadaten einschließlich Analyseergebnisse und Referenztabellen, gemeinsam genutzte Metadaten einschließlich Verlinkungen zur Geschäftsterminologie und Datenprüfregeln für die laufende Qualitätsüberwachung.

Gemeinsam genutzte Metadaten zum Veranschaulichen des Umfangs und der Auswirkung von Änderungen verwenden

IBM InfoSphere Metadata Workbench stellt eine Impactanalyse für alle Systeme im Bereich der Data-Warehousing-Initiative bereit. Beteiligte vorgeschaltete und nachgeschaltete Systeme sowie die Routinen, mit denen die Verbindungen zwischen den Systemen hergestellt werden, können problemlos ermittelt werden. Die Impactanalyse reduziert das Risiko fehlender erforderlicher Änderungen und unterbrochener Datenflüsse.

Die Modernisierung eines Data-Warehouse kann zu Änderungen der vorhandenen Data-Warehouse-Struktur führen, einschließlich ihrer aktualisierten Tabellen, Fakten oder neuen Datenbeziehungen. Sind die Änderungen kumulativ, d. h., waren sie zuvor nicht vorhanden, haben sie üblicherweise nur minimale nachgeschaltete

Auswirkungen. Wenn die Änderungen die Data-Warehouse-Struktur modifizieren, ist die Impactanalyse sowohl auf Geschäfts- als auch auf Integrationsebene entscheidend.

Nachgeschaltete Kunden erhalten anhand der Abstammung der Geschäftsdaten ein besseres Verständnis der Auswirkungen von Änderungen auf Datamarts und Berichte, denen Daten aus dem Data-Warehouse zugeführt werden, insbesondere, wenn sich Terminologie oder Dateninhalte ändern.

Die Datenabstammung verdeutlicht den Entwicklern die Auswirkungen und die für die Datenintegrationsjobs erforderlichen Änderungen, die nachgeschalteten Datamarts, Systemen oder Anwendungen Daten zuführen.

Gemeinsam genutzte Metadaten bei Mapping und Entwicklung verwenden

Sie besitzen umfassende Kenntnisse der in Ihrer Umgebung enthaltenen Informationen. Sie kennen die Form der Quellen und Ziele, die Größe sowie die Geschäftsterme. Sie haben solide Kenntnisse der Bewertung und haben Ihre Datenbeziehungen ermittelt.

Als Nächstes müssen Sie entscheiden, wie Sie die neuen oder aktualisierten Datenquellen dem Data-Warehouse-Ziel zuordnen möchten. Setzen Sie das gerade erlangte Wissen ein, um dem Unternehmen die richtigen Informationen bereitzustellen. Hierbei werden die Quelldaten verwendet und es wird sichergestellt, dass sie der Form und Größe des Ziels entsprechen.

Auf die Analyseergebnisse aus InfoSphere Discovery und InfoSphere Information Analyzer kann direkt über die Anwendung InfoSphere FastTrack zugegriffen werden, in der die Spezifikationen definiert und dokumentiert sind, sowie in InfoSphere DataStage, InfoSphere QualityStage und InfoSphere Business Catalog. Dadurch eine umfassende und präzise Entwicklungsarbeit ermöglicht.

Die Mapping-Spezifikationen in InfoSphere FastTrack können die Datenquellen direkt integrieren, die als Quelleninformationen für das Füllen des Data-Warehouse mit Daten angegeben sind. Das vereinheitlichte Schema (falls dies verwendet wird) oder ein aus den Data-Warehouse-Modellen generiertes physisches Modell können als Zielinformationen integriert werden. Sie können die Umsetzungsregeln ermitteln, die für die korrekte Abstimmung der Daten aus der Quelle mit dem Ziel zu verwenden sind, indem Sie die Analyseinformationen einschließlich der Anmerkungen und angegebenen Werte prüfen. Annotierte Referenztabellen können für Suchprozesse gekennzeichnet werden.

Wenn Sie Lücken zwischen den Quellen und dem Ziel oder fehlende Informationen finden, müssen Sie möglicherweise die folgenden Aktionen ausführen:

- Differenzen in der Terminologie prüfen, die den Quellen- und Zieldaten zugeordnet ist
- Zusätzliche Datenquellen anfordern
- Ersterkennungszyklus wieder aufrufen
- Weitere Informationen im Hinblick auf die Datenqualitätsanforderungen prüfen
- Datenqualitätsbewertungszyklus wieder aufrufen

Diese Spezifikationen werden daraufhin als Eingabeanforderungen für ETL-, Bereinigungs- und Datenumstrukturierungsjobs in InfoSphere DataStage und InfoSphere QualityStage verwendet, die die zu entwickelnden Data-Warehouse-Modifikationen

unterstützen. Je mehr Informationen Sie zum tatsächlichen Dateninhalt haben, desto genauer sind die Anforderungen für nachgeschaltete Entwickler.

Gemeinsam genutzte Metadaten in Business Glossary verwenden

Da die neuen Quellen und Informationen dem Data-Warehouse zugeführt werden, müssen Sie sich möglicherweise auch Gedanken über die Verwendung dieser Informationen machen.

Folgende Fragen können von Bedeutung sein:

- Was geschieht in dieser Umgebung, wenn Sie ein System hinzufügen und anschließend Informationen zum Unternehmen schnellstmöglich ändern oder bereitstellen müssen?
- Was ist, wenn die Führungskräfte die Quellen aller Daten zu finanziellen Risiken oder das Datum ihrer letzten Aktualisierung wissen möchten?

Bei diesem Prozess ist Berichterstellung und Datengovernance erforderlich, um alle Änderungen zu verwalten, alle Berichte bereitzustellen und allen Benutzern dieselben Informationen zugänglich machen zu können. Über die allgemeinen, gemeinsam genutzten Metadaten, in denen die Geschäftsterminologie mit tatsächlichen Datenressourcen (sowohl Datenquellen als auch das Data-Warehouse) verlinkt ist, können Informationen im Hinblick auf diese Fragentypen schnell kommuniziert werden.

Sie können gemeinsam genutzte Informationen finden, die Terme mit Datenquellen zu Analyseinformationen verlinken, Ihre neuen Data-Warehouse-Komponenten verwalten und Informationen für die Benutzer bieten, die mit den neuen Inhalten in Data-Warehouse arbeiten, ungeachtet dessen, ob Sie IBM InfoSphere Governance Catalog verwenden, um die zwischen der Quelle und dem Data-Warehouse bestehende Abstammung der Geschäftsdaten zu verstehen oder die Details zur Metadatenumgebung zu prüfen.

Fortlaufende Datenqualität überwachen

Wenn Sie die Datenqualitätsregeln und Messkriterien über die Bewertung der Datenqualität entwickelt und implementiert haben, können Sie die Informationsqualität im Data-Warehouse im Laufe der Zeit prüfen und aufzeichnen. Die Bereitstellung und Verfügbarkeit von Informationen stellen während dieser Stage Schlüsselfaktoren für Entscheidungen dar. Unterschiedliche Bereitstellungsoptionen stehen für die Aufzeichnung und Überwachung der Datenqualität zur Auswahl.

Datenqualität mithilfe von IBM InfoSphere Information Analyzer überwachen

In diesem Szenario richten Sie die Aktivitäten für die Datenqualitätsüberwachung ein.

1. Geben Sie Ihre bevorzugte Methode für die Bereitstellung und Überwachung der Datenqualität bezüglich dieser angegebenen Regeln und Messkriterien an. Die Optionen schließen die Benutzerschnittstelle, die Berichte und die Publikation von Ergebnissen über Befehlszeilenschnittstellen oder Anwendungsschnittstellen (API) ein. Die API kann zum Implementieren Ihrer unternehmensspezifischen Berichterstellungslösungen einschließlich des eigentlichen Data-Warehouse verwendet werden.
2. Geben Sie den geeigneten Ausführungszeitplan für die angegebenen Regeln und Messkriterien sowie die zugeordneten Berichte an. Prozesse, die über die Schnittstelle oder über Berichte gesteuert werden, können über den Job-Schedu-

ler von IBM InfoSphere Information Analyzer oder über die Berichtskonsolle von IBM InfoSphere Information Server terminiert werden. Wenn Prozesse über Befehlszeilenschnittstellen oder über APIs gesteuert werden, müssen Sie sie extern umwandeln oder terminieren.

3. Sofern eine bestimmte Überwachungsstufe über die Benutzerschnittstelle und die Berichte von InfoSphere Information Analyzer ausgeführt wird, können Sie zusammengefasste Regel- und Messergebnisse prüfen, die anhand eingerichteter Benchmarks (Grenzwerte) gemessen wurden.
4. Werden Varianzen angezeigt, rufen Sie die Details zu den entsprechenden Regeln oder Messkriterien auf. Diese Ergebnisse umfassen zusammengefasste Statistikdaten zu allen Ausführungen der Regel oder des Messkriteriums, über die Sie Ausnahmedetails aufrufen können.
5. Stellen Sie Fachleuten Datenexporte oder Berichte bereit. Neue und nicht erwartete Bedingungen treten in den Daten auf. Diese Bedingungen müssen geprüft und korrigiert werden. Möglicherweise ist eine Änderung der Quelldaten oder Prozesse erforderlich. Informationen im Data-Warehouse müssen möglicherweise korrigiert werden oder die Regeln müssen gegebenenfalls anschließend aktualisiert und modifiziert werden.
6. Kehren Sie gegebenenfalls zum Prozess der Datenqualitätsbewertung zurück, um die Quelldaten erneut zu analysieren oder um Regeln und Messkriterien entsprechend den geänderten Bedingungen zu modifizieren.

Datenqualität mithilfe von IBM InfoSphere Data Quality Console überwachen

Sie können Ausnahmen, die von InfoSphere Information Analyzer oder InfoSphere Discovery generiert wurden, mithilfe von Data Quality Console überwachen. Zu den Ausnahmen zählen alle Entitäten, die weitere Informationen oder Untersuchungen erfordern, beispielsweise ein Datensatz, der eine Datenqualitätsregel in InfoSphere Information Analyzer nicht erfüllt, oder mehrdeutige Übereinstimmungen aus InfoSphere Discovery.

Wenn Sie Ausnahmen mit Data Quality Console überwachen, können Sie mithilfe von Suchkriterien Informationssätze zu Ausnahmen finden. Ein Geschäftssteward ist beispielsweise für die Datenqualität implementierter Datenressourcen aus der Finanzabteilung verantwortlich. Der Geschäftssteward kann die Suchkriterien verwenden, um nur Informationen zu Ausnahmen anzuzeigen, die sich auf Datenbanken, Tabellen oder andere Ressourcen der Finanzabteilung beziehen. Die Ausnahmeininformationen können im Verlauf der Zeit überwacht und zum Verfolgen und Verbessern der Datenqualität verwendet werden.

Geschäftserkenntnisse aus qualitativ hochwertigen Daten gewinnen

Die Zusätze zum Data-Warehouse dienen dazu, umfassendere Kenntnisse und weitere Informationen zum Unternehmen bereitzustellen. Werden den Finanzdaten im Data-Warehouse Kundeninformationen hinzugefügt, kann dies Einblicke in das Kauf- oder Konsumverhalten von Kunden, in Verkaufsaktivitäten mit hohem Wert oder Gewinn usw. gewähren. Die Qualität dieser Informationen hängt jedoch von der Qualität der dem Data-Warehouse bereitgestellten Daten ab. Auch wenn die Datenqualität der Quellen als hoch eingestuft wird, können alle folgenden Situationen nachgeschaltete Probleme verursachen: Fehlen konsistenter Terminologie oder klar strukturierter Modelle, Probleme bei der Datenabstimmung, inkonsistente Werte oder falsch zugeordnete Beziehungen. Eine Hervorhebung der resultierenden Datenabstimmung reicht nicht aus, um eine effektive Nutzung und Informationsqualität zu erzielen. Wird ein entsprechendes Augenmerk auf die Datenerkennung,

die Bewertung der Datenqualität sowie die aktuelle Qualitätsüberwachung gerichtet, kann dadurch die Zuverlässigkeit der resultierenden Informationen und der Analysen beachtlich gesteigert werden.

Zugehörige Konzepte:

Szenario: Ausnahmen überwachen

Zugehörige Informationen:

Regelkonzepte und -methoden

Anhang A. Eingabehilfen in den Produkten

Sie erhalten Informationen zum Status der Eingabehilfen in IBM Produkten.

Die Produktmodule und Benutzerschnittstellen von IBM InfoSphere Information Server sind nicht uneingeschränkt für behindertengerechte Bedienung geeignet.

Informationen zum Status der Eingabehilfen in IBM Produkten finden Sie unter http://www.ibm.com/able/product_accessibility/index.html.

Dokumentation im behindertengerechten Format

Dokumentation im behindertengerechten Format für steht im IBM Knowledge Center zur Verfügung. Im IBM Knowledge Center wird zur Darstellung der Dokumentation das Format XHTML 1.0 verwendet, das mit den meisten Web-Browsern geöffnet werden kann. Da das IBM Knowledge Center XHTML verwendet, können Sie in Ihrem Browser Anzeigevorgaben festlegen. Darüber hinaus ist der Einsatz von Sprachausgabeprogrammen und anderen Unterstützungseinrichtungen für den Zugriff auf die Dokumentation möglich.

Die im IBM Knowledge Center zur Verfügung stehende Dokumentation wird auch in Form von PDF-Dateien bereitgestellt, die nicht uneingeschränkt für behindertengerechte Bedienung geeignet sind.

IBM und Eingabehilfen

Weitere Informationen zum Engagement von IBM hinsichtlich der Eingabehilfen finden Sie im IBM Human Ability and Accessibility Center.

Anhang B. Kontaktaufnahme mit IBM

Sie können sich an IBM wenden, um Unterstützung, Informationen zu Software-Services, Produktinformationen sowie allgemeine Informationen zu erhalten. Darüber hinaus können Sie Feedback zu den Produkten und zur Dokumentation an IBM senden.

In der folgenden Tabelle sind Ressourcen für die Kundenunterstützung, für Software-Services, für Schulungen sowie für Produkt- und Lösungsinformationen aufgeführt.

Tabelle 1. IBM Ressourcen

| Ressource | Beschreibung und Position |
|-----------------------------|--|
| IBM Support-Portal | Sie können die Unterstützungsinformationen je nach Bedarf anpassen, indem Sie die Produkte und Themen, die für Sie von Interesse sind, unter www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server auswählen. |
| Software-Services | Informationen zu Software-, IT- und Unternehmensberatungsservices erhalten Sie auf der Site 'Lösungen' unter www.ibm.com/businesssolutions/de . |
| Meine IBM | Auf der Site 'Meine IBM' unter www.ibm.com/account/de/de/ können Sie ein Konto einrichten und so Links auf IBM Websites und Informationen Ihren speziellen Anforderungen an die technische Unterstützung entsprechend verwalten. |
| Schulung und Zertifizierung | Informationen zu technischen Schulungs- und Ausbildungsservices, mit deren Hilfe Einzelpersonen sowie Mitarbeiter von Unternehmen und öffentlichen Organisationen IT-Kenntnisse erwerben, optimieren und auf dem neuesten Stand halten können, finden Sie unter http://www.ibm.com/training . |
| IBM Ansprechpartner | Einen IBM Ansprechpartner, bei dem Sie Informationen zu Lösungen erhalten, finden Sie unter www.ibm.com/connect/ibm/us/en/ bzw. www.ibm.com/contact/de/de/ . |

Anhang C. Auf Produktdokumentation zugreifen

Die Dokumentation wird in einer Vielzahl von Formaten bereitgestellt: online im IBM Knowledge Center, optional in einem lokal installierten Information Center sowie in Form von Handbüchern im PDF-Format. Sie können direkt über die Produktclientschnittstelle auf die Onlinehilfe oder die lokal installierte Hilfe zugreifen.

IBM Knowledge Center ist die beste Methode, um aktuelle Informationen zu InfoSphere Information Server zu suchen. Das IBM Knowledge Center enthält Hilfe für die meisten Produktschnittstellen sowie die gesamte Dokumentation für alle Produktmodule der Suite. Sie können das IBM Knowledge Center über das installierte Produkt oder über einen Web-Browser öffnen.

Auf das IBM Knowledge Center zugreifen

Es gibt verschiedene Möglichkeiten, auf die Onlinedokumentation zuzugreifen:

- Klicken Sie auf den Link **Hilfe** rechts oben in der Clientschnittstelle.
- Drücken Sie die Taste F1. Mit der Taste F1 wird normalerweise das Thema aufgerufen, das eine Beschreibung des aktuellen Kontexts der Clientschnittstelle enthält.

Anmerkung: In Web-Clients kann die Taste F1 nicht verwendet werden.

- Geben Sie die Adresse in einem Web-Browser ein, beispielsweise, wenn Sie nicht am Produkt angemeldet sind.

Geben Sie die folgende Adresse ein, um auf alle Versionen der Dokumentation zu InfoSphere Information Server zuzugreifen:

<http://www.ibm.com/support/knowledgecenter/SSZJPZ/>

Wenn Sie auf ein bestimmtes Thema zugreifen wollen, geben Sie die Versionsnummer zusammen mit der Produkt-ID, dem Namen des Dokumentations-Plugins und dem Themenpfad in der URL an. Die URL für Version 11.3 dieses Themas lautet beispielsweise wie folgt, wobei das Symbol \Rightarrow eine Zeilenfortsetzung angibt:

http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html

Tipp:

Für das Knowledge Center gibt es auch eine Kurz-URL:

<http://ibm.biz/knowctr>

Zur Angabe einer Kurz-URL zu einer bestimmten Produktseite oder Version oder zu einem bestimmten Thema geben Sie zwischen der Kurz-URL und der Produkt-ID ein Hashzeichen (#) an. Die Kurz-URL für die gesamte Dokumentation zu InfoSphere Information Server lautet beispielsweise wie folgt:

<http://ibm.biz/knowctr#SSZJPZ/>

Die folgende URL ist die leicht verkürzte URL zum obigen Thema (das Symbol \Rightarrow gibt eine Zeilenfortsetzung an):

http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html

Hilfelinks so ändern, dass sie auf lokal installierte Dokumentation verweisen

Das IBM Knowledge Center enthält die aktuelle Version der Dokumentation. Sie können jedoch eine lokale Version der Dokumentation in Form eines Information Center installieren und Ihre Hilfelinks so konfigurieren, dass sie auf dieses Information Center verweisen. Ein lokales Information Center ist sinnvoll, wenn Ihr Unternehmen keinen Zugriff auf das Internet bereitstellt.

Befolgen Sie die Installationsanweisungen im Installationspaket für das Information Center, um das Information Center auf einem Computer Ihrer Wahl zu installieren. Nach der Installation und dem Start des Information Center können Sie die von der Taste F1 der Produkthilfe und den Hilfelinks verwendete Speicherposition für die Dokumentation mithilfe des Befehls **iisAdmin** auf der Serviceschicht ändern (das Symbol \Rightarrow gibt eine Zeilenfortsetzung an):

Windows

```
IS-Installationspfad\ASBServer\bin\iisAdmin.bat -set -key  $\Rightarrow$   
com.ibm.iis.infocenter.url -value http://<Host>:<Port>/help/topic/
```

AIX Linux

```
IS-Installationspfad/ASBServer/bin/iisAdmin.sh -set -key  $\Rightarrow$   
com.ibm.iis.infocenter.url -value http://<Host>:<Port>/help/topic/
```

Dabei ist <Host> der Name des Computers, auf dem das Information Center installiert ist, und <Port> ist die Portnummer für das Information Center. Die Standardportnummer lautet 8888. Für einen Computer mit dem Namen `server1.example.com`, der den Standardport verwendet, lautet der URL-Wert beispielsweise `http://server1.example.com:8888/help/topic/`.

PDF- und Hardcopy-Dokumentation abrufen

- Die PDF-Versionen der Handbücher sind online verfügbar und können über <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1> aufgerufen werden.
- Sie können IBM Veröffentlichungen auch im Hardcopy-Format online oder über den zuständigen IBM Ansprechpartner bestellen. Wenn Sie Veröffentlichungen online bestellen möchten, rufen Sie das IBM Publications Center unter <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss> auf.

Anhang D. Feedback zur Produktdokumentation geben

Sie können hilfreiches Feedback zur IBM Dokumentation geben.

Ihr Feedback hilft IBM, Informationen von hoher Qualität bereitzustellen. Sie haben die folgenden Möglichkeiten, Kommentare zu senden:

- Wenn Sie einen Kommentar zu einem Thema im IBM Knowledge Center abgeben wollen, das sich auf der IBM Website befindet, melden Sie sich an und klicken Sie auf die Schaltfläche **Kommentar hinzufügen** am Ende des Themas. Auf diese Weise eingegebene Kommentare werden für alle Benutzer angezeigt.
- Wenn Sie einen Kommentar zu einem Thema im IBM Knowledge Center an IBM senden wollen, der nicht für alle Benutzer angezeigt werden soll, melden Sie sich an und klicken auf den Link **Feedback** unten im IBM Knowledge Center.
- Senden Sie Ihre Kommentare mithilfe des Onlineformulars für Leserkommentare unter www.ibm.com/software/awdtools/rcf/.
- Senden Sie Ihre Kommentare per E-Mail an comments@us.ibm.com. Geben Sie dabei den Namen des Produkts, die Versionsnummer des Produkts sowie den Namen und die Teilenummer der Informationen (falls vorhanden) an. Wenn sich Ihr Kommentar auf einen bestimmten Text bezieht, geben Sie die Position des Texts an (z. B. eine Überschrift, eine Tabellenummer oder eine Seitenzahl).

Bemerkungen und Marken

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden. IBM stellt dieses Material möglicherweise auch in anderen Sprachen zur Verfügung. Für den Zugriff auf das Material in einer anderen Sprache ist eine Kopie des Produkts oder der Produktversion in der jeweiligen Sprache erforderlich.

Bemerkungen

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

IBM Director of Licensing
IBM Europe, Middle East & Africa
Tour Descartes
2, avenue Gambetta
92066 Paris La Defense
France

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des

vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 USA

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des in diesem Dokument beschriebenen Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können davon abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung dient nur zu Planungszwecken. Die in dieser Veröffentlichung enthaltenen Informationen können geändert werden, bevor die beschriebenen Produkte verfügbar sind.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

COPYRIGHTLIZENZ:

Diese Veröffentlichung enthält Beispielanwendungsprogramme, die in Quellsprache geschrieben sind und Programmier Techniken in verschiedenen Betriebsumgebungen veranschaulichen. Sie dürfen diese Beispielprogramme kostenlos kopieren, ändern und verteilen, wenn dies zu dem Zweck geschieht, Anwendungsprogramme zu entwickeln, zu verwenden, zu vermarkten oder zu verteilen, die mit der Anwendungsprogrammierschnittstelle für die Betriebsumgebung konform sind, für die diese Beispielprogramme geschrieben werden. Diese Beispiele wurden nicht

unter allen denkbaren Bedingungen getestet. Daher kann IBM die Zuverlässigkeit, Wartungsfreundlichkeit oder Funktion dieser Programme weder zusagen noch gewährleisten. Die Beispielprogramme werden ohne Wartung (auf "as-is"-Basis) und ohne jegliche Gewährleistung zur Verfügung gestellt. IBM übernimmt keine Haftung für Schäden, die durch die Verwendung der Beispielprogramme entstehen.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

© (Name Ihrer Firma) (Jahr). Teile des vorliegenden Codes wurden aus Beispielprogrammen der IBM Corporation abgeleitet. © Copyright IBM Corp. _Jahr/Jahre angeben_. Alle Rechte vorbehalten.

Hinweise zur Datenschutzrichtlinie

IBM Softwareprodukte, einschließlich Software as a Service-Lösungen ("Softwareangebote"), können Cookies oder andere Technologien verwenden, um Informationen zur Produktnutzung zu erfassen, die Endbenutzererfahrung zu verbessern und Interaktionen mit dem Endbenutzer anzupassen oder zu anderen Zwecken. In vielen Fällen werden von den Softwareangeboten keine personenbezogenen Daten erfasst. Einige der IBM Softwareangebote können Sie jedoch bei der Erfassung personenbezogener Daten unterstützen. Wenn dieses Softwareangebot Cookies zur Erfassung personenbezogener Daten verwendet, sind nachfolgend nähere Informationen über die Verwendung von Cookies durch dieses Angebot zu finden.

Abhängig von den implementierten Konfigurationen kann dieses Softwareangebot Sitzungscookies oder persistente Cookies verwenden. Wird ein Produkt oder eine Komponente nicht aufgelistet, verwendet dieses Produkt bzw. diese Komponente keine Cookies.

Tabelle 2. Verwendung von Cookies durch Produkte und Komponenten von InfoSphere Information Server/InfoSphere Information Server

| Produktmodul | Komponente oder Feature | Typ des verwendeten Cookies | Erfasste Daten | Zweck der Daten | Inaktivierung des Cookies |
|---|--|---|--|--|-------------------------------|
| Beliebig (Bestandteil der InfoSphere Information Server-Installation) | InfoSphere Information Server-Webkonsole | <ul style="list-style-type: none"> • Sitzung • Persistent | Benutzername | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung | Kann nicht inaktiviert werden |
| Beliebig (Bestandteil der InfoSphere Information Server-Installation) | InfoSphere Metadata Asset Manager | <ul style="list-style-type: none"> • Sitzung • Persistent | Keine personenbezogenen Daten | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung • Besserer Bedienungskomfort • SSO-Konfiguration | Kann nicht inaktiviert werden |
| InfoSphere DataStage | Stage 'Big Data File' | <ul style="list-style-type: none"> • Sitzung • Persistent | <ul style="list-style-type: none"> • Benutzername • Digitale Signatur • Sitzungs-ID | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung • SSO-Konfiguration | Kann nicht inaktiviert werden |
| InfoSphere DataStage | Stage 'XML' | Sitzung | Interne IDs | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung | Kann nicht inaktiviert werden |

Tabella 2. Verwendung von Cookies durch Produkte und Komponenten von InfoSphere Information Server/InfoSphere Information Server (Forts.)

| Produktmodul | Komponente oder Feature | Typ des verwendeten Cookies | Erfasste Daten | Zweck der Daten | Inaktivierung des Cookies |
|--|---|---|--|--|-------------------------------|
| InfoSphere DataStage | IBM InfoSphere DataStage and QualityStage Operations Console | Sitzung | Keine personbezogenen Daten | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung | Kann nicht inaktiviert werden |
| InfoSphere Data Click | InfoSphere Information Server-Webkonsole | <ul style="list-style-type: none"> • Sitzung • Persistent | Benutzername | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung | Kann nicht inaktiviert werden |
| InfoSphere Data Quality Console | | Sitzung | Keine personbezogenen Daten | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung • SSO-Konfiguration | Kann nicht inaktiviert werden |
| InfoSphere QualityStage Standardization Rules Designer | InfoSphere Information Server-Webkonsole | <ul style="list-style-type: none"> • Sitzung • Persistent | Benutzername | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung | Kann nicht inaktiviert werden |
| InfoSphere Information Governance Catalog | | <ul style="list-style-type: none"> • Sitzung • Persistent | <ul style="list-style-type: none"> • Benutzername • Interne IDs • Status der Baumstruktur | <ul style="list-style-type: none"> • Sitzungsmanagement • Authentifizierung • SSO-Konfiguration | Kann nicht inaktiviert werden |
| InfoSphere Information Analyzer | Stage 'Data Rules' im InfoSphere DataStage and QualityStage Designer-Client | Sitzung | Sitzungs-ID | Sitzungsmanagement | Kann nicht inaktiviert werden |

Wenn die für dieses Softwareangebot genutzten Konfigurationen Sie als Kunde in die Lage versetzen, personenbezogene Daten von Endbenutzern über Cookies und andere Technologien zu erfassen, müssen Sie sich zu allen gesetzlichen Bestimmungen in Bezug auf eine solche Datenerfassung, einschließlich aller Mitteilungspflichten und Zustimmungsanforderungen, rechtlich beraten lassen.

Weitere Informationen zur Nutzung verschiedener Technologien, einschließlich Cookies, für diese Zwecke finden Sie in der "IBM Online-Datenschutzerklärung, Schwerpunkte" unter <http://www.ibm.com/privacy>, in der "IBM Online-Datenschutzerklärung" unter <http://www.ibm.com/privacy/details> im Abschnitt "Cookies, Web-Beacons und sonstige Technologien" und in "IBM Software Products and Software-as-a-Service Privacy Privacy Statement" unter <http://www.ibm.com/software/info/product-privacy>.

Marken

IBM, the IBM Logo und [ibm.com](http://www.ibm.com) sind Marken oder eingetragene Marken der International Business Machines Corporation. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie im Web unter <http://www.ibm.com/legal/copytrade.shtml>.

Die folgenden Namen sind Marken oder eingetragene Marken anderer Unternehmen:

Adobe ist eine eingetragene Marke der Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel und Itanium sind Marken oder eingetragene Marken der Intel Corporation oder ihrer Tochtergesellschaften in den USA und/oder anderen Ländern.

Linux ist eine Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows und Windows NT sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java[™] und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.

United States Postal Service ist Inhaber der folgenden Marken: CASS, CASS Certified, DPV, LACS^{Link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS und United States Postal Service. Die IBM Corporation ist ein nicht ausschließlicher Lizenznehmer für DPV und LACS^{Link}.

Weitere Unternehmens-, Produkt- oder Servicenamen können Marken anderer Hersteller sein.

Index

A

Auswirkungen von Data-Warehousing-
Änderungen 17

B

Bemerkungen 31

D

Data-Warehouse-Prozess
Datenqualität bewerten, Szenario 14
Quellenmetadatenprozess erken-
nen 12
Data-Warehouse-Szenario
Datenqualität 7
Datenqualität überwachen 19
Fragen 10
gemeinsam genutzte Metadaten bei
Mapping und Entwicklung 18
gemeinsam genutzte Metadaten in
Business Glossary 19
Schlüsselszenario-Eingaben 11
Data-Warehousing-Bereich 17
Datenintegrationsanwendung 5
Datenqualität
Data-Warehouse-Szenario 7
Datenqualität bewerten 14
Data-Warehouse-Prozess 14
Datenqualität überwachen 19
Data-Warehouse-Szenario 19
Datenqualität und Überwachungsintegra-
tion
Data-Warehouse 8

G

Gemeinsam genutzte Metadaten bei Map-
ping und Entwicklung 18
Data-Warehouse-Szenario 18
Gemeinsam genutzte Metadaten in Busi-
ness Glossary 19
Data-Warehouse-Szenario 19

H

Häufig gestellte Fragen
Data-Warehouse-Szenario 10

I

IBM Industry Models 9
Integration
Data-Warehouse 8
Integrationszenarios, Übersicht 1

K

Kundenunterstützung
kontaktieren 25

M

Marken
Liste 31

P

Produktdokumentation
Zugriff auf 27
Produkteingabehilfen
Eingabehilfen 23

Q

Quellenmetadatenprozess erkennen
Data-Warehouse-Prozess 12

S

Schlüsselszenario-Eingaben 11
Data-Warehouse-Szenario 11
Schlüsselszenario-Eingabeprozesse 12
Software-Services
kontaktieren 25

U

Unterstützung
Kundenunterstützung 25



SC43-1113-00

