



Informatica® Data Quality
10.1.1 HotFix 2

Kurzanleitung

Diese Software und die Dokumentation werden nur im Rahmen eines eigenen Lizenzvertrags zur Verfügung gestellt, der Beschränkungen für die Verwendung und Weitergabe enthält. Ohne ausdrückliche schriftliche Genehmigung der Informatica LLC darf kein Teil dieses Dokuments zu irgendeinem Zweck vervielfältigt oder übertragen werden, unabhängig davon, auf welche Art und Weise oder mit welchen Mitteln (elektronisch, mechanisch, durch Fotokopieren, Aufzeichnen usw.) dies geschieht.

Den RECHTEN DER REGIERUNG DER VEREINIGTEN STAATEN unterliegende Programme, Software, Datenbanken und zugehörige Dokumentation und technische Daten, die an Kunden der Regierung der Vereinigten Staaten geliefert werden, sind "kommerzielle Computersoftware" oder "kommerzielle technische Daten" gemäß der anwendbaren Beschaffungsverordnung der Vereinigten Staaten (Federal Acquisition Regulation – FAR) und der ergänzenden Bestimmungen der spezifischen Behörde. Damit unterliegen die Nutzung, das Kopieren, die Offenlegung, das Modifizieren und die Anpassung den im anwendbaren Regierungsvertrag gemachten Einschränkungen und Lizenzbedingungen und, soweit im Rahmen der Bedingungen des Regierungsvertrags und der in FAR 52.227-19 aufgeführten Rechte anwendbar, der Lizenz für die kommerzielle Computersoftware.

Informatica, das Informatica-Logo, PowerCenter und PowerExchange sind Marken oder eingetragene Marken der Informatica LLC in den Vereinigten Staaten von Amerika und zahlreichen anderen Ländern der Welt. Eine aktuelle Liste der Informatica-Marken ist im Internet auf <https://www.informatica.com/trademarks.html> verfügbar. Alle weiteren Produkt- und Firmennamen sind möglicherweise Markennamen oder Warenzeichen der jeweiligen Eigentümer.

Teile dieser Software und/oder Dokumentationen unterliegen dem Urheberrecht Dritter. Die erforderlichen Hinweise auf Drittanbieter sind im Lieferumfang des Produkts enthalten.

Weitere Informationen über die Patente finden Sie unter <https://www.informatica.com/legal/patents.html>.

HAFTUNGSAUSSCHLUSS: Informatica LLC stellt diese Dokumentation „wie besehen“ bereit, ohne ausdrückliche oder stillschweigende Gewährleistung, einschließlich, jedoch nicht beschränkt auf die Gewährleistungen der Nichtverletzung der Rechte von Dritten, der Handelsüblichkeit oder Eignung für einen bestimmten Zweck. Informatica LLC garantiert nicht die Fehlerfreiheit dieser Software oder Dokumentation. Die in dieser Software oder Dokumentation bereitgestellten Informationen können technische Ungenauigkeiten oder Druckfehler enthalten. Die in dieser Software und in dieser Dokumentation enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden.

HINWEISE

Dieses Informatica-Produkt (die „Software“) umfasst bestimmte Treiber (die „DataDirect-Treiber“) von DataDirect Technologies, einem Betreiber von Progress Software Corporation („DataDirect“), die folgenden Bedingungen und Bestimmungen unterliegen:

1. DIE DATADIRECT-TREIBER WERDEN „WIE GESEHEN“ OHNE JEGLICHE GEWÄHRLEISTUNG, WEDER AUSDRÜCKLICH NOCH STILLSCHWEIGEND, BEREITGESTELLT, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDEN GEWÄHRLEISTUNGEN DER HANDELSÜBLICHKEIT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK UND DER NICHTVERLETZUNG VON RECHTEN DRITTER.
2. IN KEINEM FALL SIND DATADIRECT ODER DRITTANBIETER DEM ENDBENUTZER GEGENÜBER HAFTBAR FÜR UNMITTELBARE, MITTELBARE, KONKRETE, NEBEN-, FOLGE- ODER ANDERE SCHÄDEN, DIE SICH AUS DER VERWENDUNG DER ODBC-TREIBER ERGEBEN, UNABHÄNGIG DAVON, OB SIE IM VORAUS ÜBER DIE MÖGLICHKEIT SOLCHER SCHÄDEN INFORMIERT WORDEN SIND ODER NICHT. DIESE BESCHRÄNKUNGEN GELTEN FÜR ALLE KLAGEGEGENSTÄNDE, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF VERTRAGSBRUCH, GEWÄHRLEISTUNGSBRUCH, FAHRLÄSSIGKEIT, KAUSALHAFTUNG, TÄUSCHUNG UND ANDERE UNERLAUBTE HANDLUNGEN.

Die in dieser Dokumentation enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden. Wenn Sie Probleme in dieser Dokumentation finden, melden Sie sie uns unter infa_documentation@Informatica.com.

Informatica-Produkte unterliegen einer Gewährleistung gemäß den Geschäftsbedingungen der Vereinbarungen, unter denen sie bereitgestellt werden. INFORMATICA STELLT DIE INFORMATIONEN IN DIESEM DOKUMENT OHNE MÄNGELGEWÄHR UND OHNE AUSDRÜCKLICHE ODER STILLSCHWEIGENDE GEWÄHRLEISTUNG JEDLICHER ART ZUR VERFÜGUNG. DIES GILT EINSCHLIESSLICH FÜR GEWÄHRLEISTUNGEN DER MARKTGÄNGIGKEIT, DER EIGNUNG FÜR EINEN BESTIMMTEN ZWECK UND GEWÄHRLEISTUNGEN ODER ZUSICHERUNGEN ÜBER DIE NICHTVERLETZUNG VON RECHTEN DRITTER.

Publikationsdatum: 2018-06-09

Inhalt

Einleitung	7
Informatica-Ressourcen.	7
Informatica-Netzwerk.	7
Informatica-Wissensdatenbank.	7
Informatica-Dokumentation.	7
Informatica-Produktverfügbarkeitsmatrizen.	8
Informatica Velocity.	8
Informatica Marketplace.	8
Globaler Kundensupport von Informatica.	8
 Kapitel 1: Erste Schritte - Übersicht.....	 9
Informatica-Domänenübersicht.	9
Verfügbarkeit von Funktionen.	11
Einführung in Informatica Analyst.	11
Informatica Developer - Übersicht.	12
Informatica Developer -Begrüßungsseite.	13
Spickzettel.	13
Datenqualität und Profiling.	14
Die Tutorial-Story.	15
Die Tutorial-Struktur.	15
Informatica Analyst-Tutorial.	15
Informatica Developer Tool.	16
Voraussetzungen für das Tutorial.	17
 Teil I: Erste Schritte mit Informatica Analyst.....	 18
 Kapitel 2: Lektion 1. Einrichten von Informatica Analyst.....	 19
Einrichten von Informatica Analyst - Übersicht.	19
Aufgabe 1. Anmelden bei Informatica Analyst.	20
Aufgabe 2. Erstellen eines Projekts.	20
Aufgabe 3. Erstellen eines Ordners.	21
Einrichten von Informatica Analyst - Zusammenfassung.	21
 Kapitel 3: Lektion 2. Erstellen von Datenobjekten.....	 22
Erstellen von Datenobjekten - Übersicht.	22
Aufgabe 1. Erstellen der Einfachdatei-Datenobjekte.	23
Aufgabe 2. Anzeigen der Datenobjekteigenschaften.	23
Erstellen von Datenobjekten - Zusammenfassung.	24

Kapitel 4: Lektion 3. Erstellen von Standardprofilen.	25
Erstellen von Standardprofilen – Übersicht.	25
Aufgabe 1. Erstellen und Ausführen eines Standardprofils.	26
Aufgabe 2. Anzeigen der Profilergebnisse in der Zusammenfassungsansicht.	27
Erstellen von Standardprofilen – Zusammenfassung.	27
 Kapitel 5: Lektion 4. Erstellen von Kundenprofilen.	 29
Erstellen von benutzerdefinierten Profilen - Übersicht.	29
Aufgabe 1. Erstellen eines benutzerspezifischen Profils.	30
Aufgabe 2. Ausführen des Profils.	31
Aufgabe 3. Drilldown der Profilergebnisse.	31
Erstellen von benutzerdefinierten Profilen - Zusammenfassung.	32
 Kapitel 6: Lektion 5. Erstellen von Expressionsregeln.	 33
Erstellen von Ausdrucksregeln - Übersicht.	33
Aufgabe 1. Erstellen von Expressionsregeln und Ausführen des Profils.	34
Aufgabe 2. Anzeigen der Expressionsregel-Ausgabe.	34
Aufgabe 3. Bearbeiten der Expressionsregeln.	35
Erstellen von Expressionsregeln - Zusammenfassung.	36
 Kapitel 7: Lektion 6. Erstellen und Ausführen von Scorecards.	 37
Erstellen und Ausführen von Scorecards - Übersicht.	37
Aufgabe 1. Erstellen einer Scorecard aus den Profilergebnissen.	38
Aufgabe 2. Ausführen der Scorecard.	39
Aufgabe 3. Anzeigen der Scorecard.	39
Aufgabe 4. Bearbeiten der Scorecard.	40
Aufgabe 5. Konfigurieren von Schwellenwerten.	41
Aufgabe 6. Anzeigen von Score-Trenddiagrammen.	41
Erstellen und Ausführen von Scorecards - Zusammenfassung.	41
 Kapitel 8: Lektion 7. Erstellen von Referenztabellen aus Profilspalten	 43
Erstellen von Referenztabellen aus Profilspalten - Übersicht.	43
Aufgabe 1. Erstellen einer Referenztable aus Profilspalten.	44
Aufgabe 2. Bearbeiten der Referenztable.	45
Erstellen von Referenztabellen aus Profilspalten - Zusammenfassung.	45
 Kapitel 9: Lektion 8. Erstellen von Referenztabellen.	 46
Erstellen von Referenztabellen - Übersicht.	46
Aufgabe 1. Erstellen einer Referenztable.	47
Erstellen von Referenztabellen - Zusammenfassung.	47

Teil II: Inbetriebnahme des Informatica Developer.....	48
Kapitel 10: Lektion 1. Einrichten von Informatica Developer.	49
Einrichten von Informatica Developer - Übersicht.	49
Aufgabe 1 Informatica Developer starten.	50
Aufgabe 2 Eine Domäne hinzufügen.	50
Aufgabe 3 Ein Model Repository hinzufügen.	51
Aufgabe 4 Ein Projekt erstellen.	51
Aufgabe 5 Einen Ordner erstellen.	51
Aufgabe 6 Standard-Data Integration Service auswählen..	52
Einrichten von Informatica Developer - Zusammenfassung.	52
Kapitel 11: Lektion 2: Importieren von physischen Datenobjekten..	53
Importieren von physischen Datenobjekten - Übersicht.	53
Aufgabe 1 Importieren des Flatfile-Datenobjekts Boston_Customers.	54
Aufgabe 2 Import des Einfachdatei-Datenobjekts LA_Customers.	60
Aufgabe 3. Importieren des Flatfile-Datenobjekts All_Customers.	61
Importieren von physischen Datenobjekten - Zusammenfassung.	61
Kapitel 12: Lektion 3. Ausführen eines Profils auf Quelldaten.	62
Profiling von Daten - Übersicht.	62
Aufgabe 1. Durchführen einer Join-Analyse für zwei Datenquellen.	63
Aufgabe 2. Anzeigen der Ergebnisse einer Join-Analyse.	64
Aufgabe 3. Ausführen eines Profils für eine Datenquelle.	65
Aufgabe 4. Anzeigen der Spalten-Profiling-Ergebnisse.	65
Profiling von Daten - Zusammenfassung.	66
Kapitel 13: Lektion 4. Parsen von Daten.	67
Parsen von Daten - Übersicht.	67
Aufgabe 1. Erstellen eines Target-Datenobjekts.	68
Schritt 1. Erstellen eines Datenobjekts LA_Customers_tgt.	68
Schritt 2. Konfigurieren der Lesen- und Schreiben-Optionen.	69
Schritt 3. Hinzufügen von Spalten zum Datenobjekt.	69
Aufgabe 2. Erstellen eines Mappings zum Parsen von Daten.	70
Schritt 1. Erstellen eines Mappings.	70
Schritt 2. Hinzufügen von Datenobjekten zum Mapping.	70
Schritt 3. Hinzufügen einer Parser-Umwandlung zum Mapping.	71
Schritt 4. Konfigurieren der Parser-Umwandlung.	71
Aufgabe 3. Ausführen eines Profils für die Parser-Umwandlung.	72
Aufgabe 4. Ausführen des Mappings.	72
Aufgabe 5. Anzeigen der Mapping-Ausgabe.	72
Parsen von Daten - Zusammenfassung.	73

Kapitel 14: Lektion 5. Standardisieren von Daten 74

Standardisieren von Daten - Übersicht.	74
Aufgabe 1. Erstellen eines Target-Datenobjekts.	75
Schritt 1. Erstellen eines Datenobjekts All_Customers_Stdz_tgt.	75
Schritt 2. Konfigurieren der Lesen- und Schreiben-Optionen.	76
Aufgabe 2 Erstellen eines Mappings zum Standardisieren von Daten.	76
Schritt 1. Erstellen eines Mappings.	77
Schritt 2. Hinzufügen von Datenobjekten zum Mapping.	77
Schritt 3. Hinzufügen einer Standardisierer-Umwandlung zum Mapping.	77
Schritt 4. Konfigurieren der Standardisierer-Umwandlung.	78
Aufgabe 3. Ausführen des Mappings.	79
Aufgabe 4. Anzeigen der Mapping-Ausgabe.	79
Standardisieren von Daten - Zusammenfassung.	79

Kapitel 15: Lektion 6. Validieren von Adressdaten. 80

Validieren von Adressdaten - Übersicht.	80
Aufgabe 1 Erstellen eines Target-Datenobjekts	81
Schritt 1. Erstellen des Datenobjekts All_Customers_av_tgt.	82
Schritt 2. Konfigurieren der Lesen- und Schreiben-Optionen.	82
Schritt 3. Hinzufügen von Ports zum Datenobjekt.	83
Aufgabe 2 Erstellen eines Mappings zum Validieren von Adressen.	83
Schritt 1. Erstellen eines Mappings.	83
Schritt 2. Hinzufügen von Datenobjekten zum Mapping.	84
Schritt 3. Hinzufügen einer Adressvalidierer-Umwandlung zum Mapping.	84
Aufgabe 3. Konfigurieren der Adressvalidierungsumwandlung.	84
Schritt 1. Festlegen des Standardlands für Adressvalidierung.	85
Schritt 2. Konfigurieren der Eingabeports der Adressvalidierer-Umwandlung.	85
Schritt 3. Konfigurieren der Ausgabe-Ports der Adressvalidierer-Umwandlung.	86
Schritt 4. Verbinden unbenutzter Datenquellenports mit dem Daten-Target.	87
Aufgabe 4. Ausführen des Mappings.	87
Aufgabe 5. Anzeigen der Mapping-Ausgabe.	88
Validieren von Adressdaten - Zusammenfassung.	90

Anhang A: Häufig gestellte Fragen (FAQ)..... 91

Informatica Analyst - Häufig gestellte Fragen (FAQ).	91
Informatica Developer - Häufig gestellte Fragen (FAQ).	91

Index. 93

Einleitung

Die *Kurzanleitung* für die Datenqualität wurde für Datenqualitäts-Entwickler und Analytiker geschrieben. Sie enthält Tutorials, um Einsteigern die Anwendung von Informatica Developer und Informatica Analyst zu erleichtern. Diese Anleitung geht davon aus, dass Ihnen die Datenqualitätskonzepte, die Konzepte von Flatfiles und relationalen Datenbanken sowie die Datenbank-Engines in Ihrer Umgebung bekannt sind.

Informatica-Ressourcen

Informatica-Netzwerk

Im Informatica-Netzwerk finden Sie den globalen Kundensupport von Informatica, die Informatica-Wissensdatenbank und andere Produktressourcen. Für den Zugriff auf das Informatica-Netzwerk besuchen Sie <https://network.informatica.com>.

Als Mitglied können Sie:

- zentral auf alle Ihre Informatica-Ressourcen zugreifen.
- Durchsuchen Sie die Wissensdatenbank nach Produktressourcen, einschließlich Dokumentation, häufig gestellter Fragen und bewährter Methoden.
- Zeigen Sie Informationen zur Produktverfügbarkeit an.
- Ihre Support-Fälle prüfen.
- Ihr lokales Informatica-Netzwerk für Benutzergruppen suchen und mit anderen Benutzern zusammenarbeiten.

Informatica-Wissensdatenbank

Verwenden Sie die Informatica-Wissensdatenbank, um das Informatica-Netzwerk nach Produktressourcen, wie z. B. Dokumentation, Ratgeberartikeln, bewährten Methoden und PAMs, zu durchsuchen.

Für den Zugriff auf die Wissensdatenbank besuchen Sie <https://kb.informatica.com>. Wenn Sie Fragen, Kommentare oder Ideen zur Wissensdatenbank haben, wenden Sie sich per E-Mail an das Team der Informatica-Wissensdatenbank unter KB_Feedback@informatica.com.

Informatica-Dokumentation

Navigieren Sie zur Informatica-Wissensdatenbank unter https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx, um die aktuelle Dokumentation für Ihr Produkt abzurufen.

Wenn Sie Fragen, Kommentare oder Ideen zu dieser Dokumentation haben, wenden Sie sich per E-Mail an das Informatica-Dokumentationsteam unter infa_documentation@informatica.com.

Informatica-Produktverfügbarkeitsmatrizen

Produktverfügbarkeitsmatrizen (PAMs) geben die Versionen der Betriebssysteme, Datenbanken und anderen Typen von Datenquellen und Zielen an, die in einer Produktversion unterstützt werden. Als Mitglied des Informatica-Netzwerks können Sie unter <https://network.informatica.com/community/informatica-network/product-availability-matrices> auf PAMs zugreifen.

Informatica Velocity

Bei Informatica Velocity handelt es sich um eine Sammlung von Tipps und bewährten Methoden, die von den professionellen Informatica-Diensten entwickelt wurden. Informatica Velocity basiert auf der Praxiserfahrung aus Hunderten von Datenmanagementprojekten und umfasst das kollektive Wissen unserer Berater, die mit Unternehmen aus der ganzen Welt an der Planung, Entwicklung, Bereitstellung und Wartung erfolgreicher Datenmanagementlösungen gearbeitet haben.

Als Mitglied des Informatica-Netzwerks können Sie unter <http://velocity.informatica.com> auf Informatica Velocity-Ressourcen zugreifen.

Wenn Sie Fragen, Anregungen oder Ideen zu Informatica Velocity haben, wenden Sie sich an die professionellen Informatica-Dienste unter ips@informatica.com.

Informatica Marketplace

Informatica Marketplace ist ein Forum, das Lösungen zur Erweiterung und Verbesserung Ihrer Informatica-Implementierungen bereitstellt. Indem Sie die zahlreichen Lösungen von Informatica-Entwicklern und -Partnern nutzen, können Sie Ihre Produktivität steigern und die Implementierungsdauer Ihrer Projekte verkürzen. Zugriff auf den Informatica Marketplace erhalten Sie unter <https://marketplace.informatica.com>.

Globaler Kundensupport von Informatica

Sie können sich telefonisch oder über den Online-Support mit einem globalen Support-Center im Informatica-Netzwerk in Verbindung setzen.

Die Telefonnummer des globalen Kundensupports von Informatica vor Ort finden Sie auf der Informatica-Website unter folgender Verknüpfung:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

Als Mitglied des Informatica-Netzwerks können Sie den Online-Support unter <http://network.informatica.com> verwenden.

KAPITEL 1

Erste Schritte - Übersicht

Dieses Kapitel umfasst die folgenden Themen:

- [Informatica-Domänenübersicht, 9](#)
- [Einführung in Informatica Analyst, 11](#)
- [Informatica Developer - Übersicht, 12](#)
- [Die Tutorial-Story, 15](#)
- [Die Tutorial-Struktur, 15](#)

Informatica-Domänenübersicht

Informatica verfügt über eine dienstorientierte Architektur, die die Fähigkeit zur Skalierung von Diensten und zum gemeinsamen Nutzen von Ressourcen über mehrere Rechner aufweist. Die Informatica-Domäne ist die primäre Einheit zur Verwaltung von Diensten.

Sie können sich nach der Installation von Informatica bei Informatica Administrator anmelden. Das Administrator Tool dient zum Verwalten der Domäne und zum Konfigurieren der benötigten Anwendungsdienste, bevor Sie auf die restlichen Anwendungs-Clients zugreifen können.

Die Informatica-Domäne enthält die folgenden Komponenten:

- **Anwendungs-Clients.** Eine Gruppe von Clients, die Sie nutzen, um auf die zu Grunde liegende Informatica-Funktionalität zuzugreifen. Anwendungs-Clients stellen Anfragen an den Dienstmanager oder an Anwendungsdienste.
- **Anwendungsdienste.** Eine Gruppe von Diensten, die Funktionalität auf Serverbasis repräsentieren. Eine Informatica-Domäne kann eine Teilmenge der Anwendungsdienste enthalten. Sie erstellen und konfigurieren die Anwendungsdienste, die von den Anwendungs-Clients benötigt werden.

Anwendungsdienste umfassen Systemdienste, die eine einzelne Instanz in der Domäne aufweisen können. Die Systemdienste werden beim Erstellen der Domäne angelegt. Sie können einen Systemdienst konfigurieren und aktivieren, um die Funktionen zu verwenden, die vom Dienst bereitgestellt werden.
- **Profil-Warehouse.** Eine relationale Datenbank, die der Datenintegrationsdienst zum Speichern der Profilergebnisse verwendet.
- **Referenzdaten-Warehouse.** Eine relationale Datenbank, die Referenzdatenwerte für die Referenztabelleobjekte im Modellrepository speichert.
- **Repositorys.** Eine Gruppe relationaler Datenbanken, die Metadaten zu Objekten und Prozessen speichern, die erforderlich sind, um Benutzeranfragen von Anwendungs-Clients abzuarbeiten.

- **Dienstmanager.** Ein Dienst, der in die Domäne integriert ist, um alle Domänenvorgänge zu verwalten. Der Dienstmanager führt die Anwendungsdienste aus und ist für Domänenfunktionen wie Authentifizierung, Autorisierung und Anmeldung zuständig.
- **Arbeitsablauf-Datenbank.** Eine relationale Datenbank, die Laufzeit-Metadaten für Arbeitsabläufe speichert.

Die folgende Tabelle präsentiert die Anwendungs-Clients (außer dem Administrator Tool) sowie die Anwendungsdienste und die Repositories, die der Client benötigt:

Anwendungs-Client	Anwendungsdienste	Repositories
Informatica Analyst	<ul style="list-style-type: none"> - Analyst-Dienst - Content-Management-Dienst - Datenintegrationsdienst - Modellrepository-Dienst - Suchdienst 	Modellrepository
Informatica Developer	<ul style="list-style-type: none"> - Analyst-Dienst - Content-Management-Dienst - Datenintegrationsdienst - Modellrepository-Dienst 	Modellrepository
Metadata Manager	<ul style="list-style-type: none"> - Metadata Manager-Dienst - PowerCenter-Integrationsdienst - PowerCenter-Repository-Dienst 	<ul style="list-style-type: none"> - Metadata Manager-Repository - PowerCenter-Repository
PowerCenter® Client	<ul style="list-style-type: none"> - PowerCenter-Integrationsdienst - PowerCenter-Repository-Dienst 	PowerCenter-Repository
Webdienst-Hub-Konsole	<ul style="list-style-type: none"> - PowerCenter-Integrationsdienst - PowerCenter-Repository-Dienst - Webdienst-Hub 	PowerCenter-Repository

Auf folgenden Anwendungsdienste kann ein Informatica Anwendungs-Client nicht zugreifen:

- **PowerExchange®-Listenerdienst.** Verwaltet den PowerExchange Listener zur Verschiebung von Bulkdaten und zur Änderung der Datenerfassung. Der PowerCenter-Integrationsdienst verbindet sich über den Listenerdienst mit dem PowerExchange Listener.
- **PowerExchange-Protokollierungsdienst.** Verwaltet den PowerExchange Logger für Linux, UNIX und Windows, um sich ändernde Daten zu erfassen und in die Protokolldateien des PowerExchange Logger zu schreiben. Sich ändernde Daten können aus DB2-Wiederherstellungsprotokollen Oracle-Wiederherstellungsprotokollen, einer verteilten Microsoft-SQL-Server-Datenbank oder Datenquellen auf einem i5/OS- oder z/OS-System stammen.
- **SAP BW-Dienst.** Wartet auf RFC-Anfragen von SAP BI und fordert an, dass der PowerCenter-Integrationsdienst Arbeitsabläufe durchführt, um von SAP BI Daten zu extrahieren oder Daten auf SAP BI zu laden.

Verfügbarkeit von Funktionen

Informatica -Produkte verwenden einen gemeinsamen Satz von Anwendungen. Die Produktfunktionen, welche Sie nutzen können, sind von Ihrer Produktlizenz abhängig.

In der folgenden Tabelle finden Sie die Lizenzoptionen und die Funktionen der Anwendung, die bei jeder Option zur Verfügung stehen.

Lizenzoption	Informatica Developer-Funktionen	Informatica Analyst-Funktionen
Datenqualität	<ul style="list-style-type: none"> - Erstellen und Ausführen von Mappings - Erstellen und Ausführen von Regeln - Profiling, das die Nutzung des Enterprise-Erkennungsprofils und die Erkennung von Primärschlüssel-, Fremdschlüssel- und funktionellen Abhängigkeiten umfasst. - Wiederherstellen abgeleiteter Profilergebnisse - Arbeiten mit Scorecards - Exportieren von Objekten an PowerCenter 	<ul style="list-style-type: none"> - Profiling einschließlich Enterprise-Erkennung - Verwenden Sie die Erkennungssuche, um zu ermitteln, wo sich die Daten und Metadaten in den Profiling-Repositorys befinden - Arbeiten mit Scorecards - Wiederherstellen abgeleiteter Profilergebnisse - Erstellen und Ausführen von Profiling-Regeln - Referenztabellenverwaltung - Erstellen von Regelspezifikationen und Kompilieren von Regelspezifikationen in Mapplets - Verwaltung fehlerhafter und mehrfacher Datensätze
Datendienste	<ul style="list-style-type: none"> - Erstellen logischer Datenobjektmodelle - Erstellen und Ausführen von Mappings mit Datendienst-Umwandlungen - Erstellen von SQL-Datendiensten - Erstellen von Web-Diensten - Exportieren von Objekten an PowerCenter 	<ul style="list-style-type: none"> - Referenztabellenverwaltung
Datendienste und Profiling-Option	<ul style="list-style-type: none"> - Erstellen logischer Datenobjektmodelle - Erstellen und Ausführen von Mappings mit Datendienst-Umwandlungen - Erstellen von SQL-Datendiensten - Erstellen von Web-Diensten - Exportieren von Objekten an PowerCenter - Erstellen und Ausführen von Regeln mit Datendienst-Umwandlungen - Profiling 	<ul style="list-style-type: none"> - Referenztabellenverwaltung

Einführung in Informatica Analyst

Informatica Analyst ist ein webbasierter Anwendungs-Client-Basis, mit dem Daten in einem Unternehmen analysiert, bereinigt, standardisiert, profiliert und bewertet werden können.

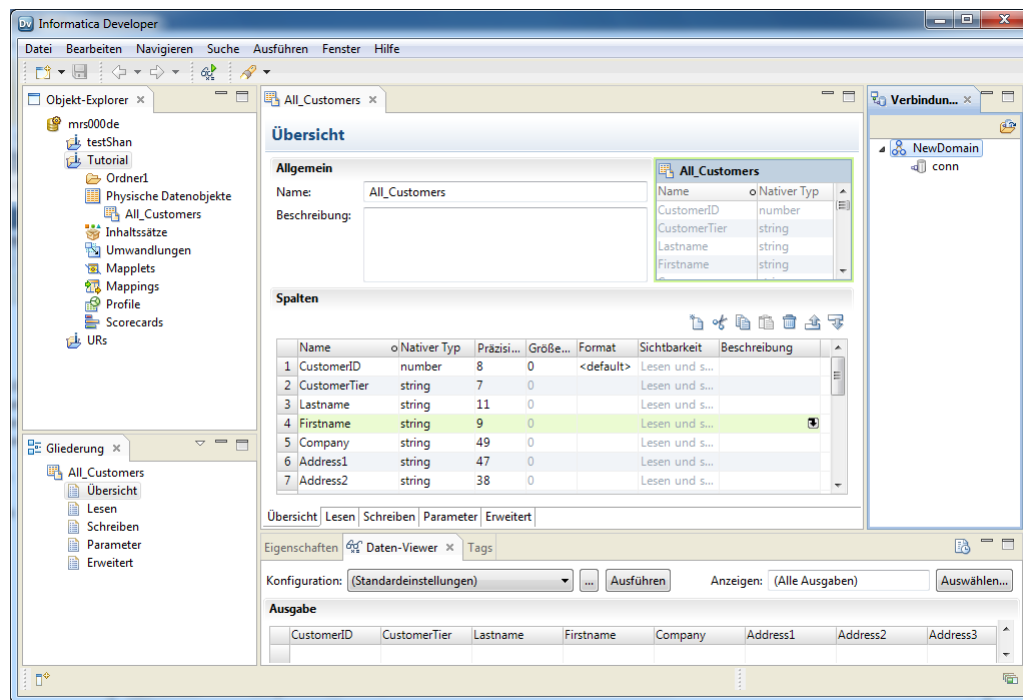
Je nach Lizenz verwenden Wirtschaftsanalysten und Entwickler das Analyst-Tool zur datenorientierten Zusammenarbeit. Sie können Spalten- und Regel-Profiling durchführen, mit Scorecards arbeiten und

fehlerhafte sowie mehrfache Datensätze verwalten. Sie können auch Referenzdaten verwalten und die Daten Entwicklern in einer Datenqualitätslösung zur Verfügung stellen.

Informatica Developer - Übersicht

Informatica Developer ist ein Anwendungs-Client, mit dem Entwickler Datenlösungen entwerfen und implementieren können.

Die folgende Abbildung zeigt das Entwicklertool:



Das Developer-Tool kann die folgenden Ansichten anzeigen:

Ansicht „Objekt-Explorer“

Zeigt Projekte, Ordner und die Objekte in den Projekten und Ordnern an.

Ansicht „Verbindungs-Explorer“

Zeigt Verbindungen zu relationalen Datenbanken an.

Ansicht „Outline“

Zeigt Objekte an, die von einem in der Ansicht **Objekt-Explorer** ausgewählten Objekt abhängig sind.

Ansicht „Referenzkarte“

Zeigt den Spickzettel an, den Sie öffnen. Um einen Spickzettel zu öffnen, klicken Sie auf **Hilfe** > **Spickzettel** und wählen Sie einen Spickzettel aus.

Ansicht „Daten-Viewer“

Zeigt Quelldaten, Profilergebnisse und eine Vorschau der Ausgabe einer Umwandlung an. .

Ansicht „Objektabhängigkeiten“

Zeigt die Objektabhängigkeiten beim Anzeigen, Ändern oder Löschen eines Objekts an.

Ansicht „Alarme“

Zeigt Verbindungsstatusalarme an.

Ansicht Validierungsprotokoll

Zeigt Objektvalidierungsfehler an.

Das Developer-Tool kann auch andere Ansichten anzeigen. Sie können Ansichten ausblenden und Ansichten an einen anderen Speicherort in der Developer Tool-Workbench verschieben. Klicken Sie auf **Fenster** >

Ansicht anzeigen, um die Ansichten auszuwählen, die Sie anzeigen möchten.

Informatica Developer -Begrüßungsseite

Die Begrüßungsseite erscheint, wenn Sie zum ersten Mal das Developer Tool öffnen. Auf der Begrüßungsseite erfahren Sie mehr über das Developer Tool, wie es eingerichtet wird und wie man beginnt, mit dem Developer Tool zu arbeiten.

Die Begrüßungsseite zeigt folgende Optionen:

- Übersicht. Klicken Sie auf Übersicht, um eine Übersicht über Datenqualitäts- und Datendienstlösungen zu erhalten.
- Erste Schritte. Klicken Sie auf Erste Schritte, um mehr darüber zu erfahren, wie das Developer Tool eingerichtet wird und auf die Lektionen von Informatica Data Quality und Informatica Data Services zugegriffen wird.
- Tutorials. Klicken Sie auf Tutorials, um Tutoriallektionen zu Datenqualitäts- und Datendienstlösungen anzuzeigen.
- Internet-Ressourcen. Klicken Sie auf die Schaltfläche „Internet-Ressourcen“, um eine Verknüpfung zu mysupport.informatica.com für den Zugriff auf die Informatica How-To-Bibliothek herzustellen. Die Informatica-Produktverwendung enthält Artikel über Informatica Data Quality, Informatica Data Services und andere Produkte von Informatica.
- Neue Funktionen. Klicken Sie auf die Schaltfläche „Neue Funktionen“, um die neuesten Funktionen im Developer Tool anzuzeigen.

Spickzettel

Das Developer-Tool umfasst Spickzettel als Teil der Online-Hilfe. Ein Spickzettel ist eine schrittweise Anleitung, die Ihnen bei der Durchführung von einem oder mehreren Tasks im Developer-Tool hilft.

Nachdem Sie einen Spickzettel vollständig ausgefüllt haben, führen Sie die Aufgaben aus und zeigen die Ergebnisse an. Nachdem Sie beispielsweise einen Spickzettel zum Importieren und zur Vorschau eines relationalen Datenobjekts vollständig ausgefüllt haben, haben Sie eine Tabelle einer relationalen Datenbank importiert und die Vorschau der Daten im Entwicklertool angezeigt.

Um auf die Spickzettel zuzugreifen, klicken Sie auf **Hilfe** > **Spickzettel**.

Datenqualität und Profiling

Verwenden Sie die Datenqualitätsfunktionen im Developer-Tool, um den Inhalt und die Struktur Ihrer Daten zu analysieren. Sie können die Daten entsprechend Ihrer Unternehmensanforderungen verbessern.

Verwenden Sie das Developer-Tool zum Konzipieren und Ausführen von Prozessen, mit denen die folgenden Ziele erreicht werden können:

- **Profilieren von Daten.** Profiling macht den Inhalt und die Struktur Ihrer Daten erkennbar. Profiling ist ein wichtiger Schritt in jedem Datenprojekt, da es Stärken und Schwächen in Ihren Daten identifizieren kann und Sie bei der Definition Ihres Projektplans unterstützt.
- **Erstellen von Scorecards,** um die Datenqualität zu prüfen. Eine Scorecard ist eine grafische Darstellung der Qualitätsmessungen in einem Profil.
- **Standardisieren von Datenwerten.** Standardisieren Sie Daten, um Fehler und Inkonsistenzen zu entfernen, die Sie finden, indem Sie ein Profil ausführen. Sie können Variationen der Zeichensetzung, Formatierung und Rechtschreibung standardisieren. Sie können beispielsweise sicherstellen, dass die Werte für Stadt, Staat und Postleitzahl konsistent sind.
- **Parsen von Datensätzen.** Parsen Sie Datensätze, um die Datensatzstruktur zu verbessern und weitere Informationen aus Ihren Daten abzuleiten. Sie können ein Einzelfeld mit Freiformdaten in Bereiche teilen, die unterschiedliche Informationstypen enthalten. Sie können Ihren Datensätzen auch Informationen hinzufügen. Sie können beispielsweise Datensätze von Kunden als persönliche oder Geschäftskunden markieren.
- **Validieren von Postadressen.** Die Adressvalidierung bewertet und verbessert die Genauigkeit und Zustellbarkeit Ihrer Postadressendaten. Die Adressvalidierung korrigiert Fehler in Adressen und komplettiert unvollständige Adressen, indem die Adressdatensätze mit Referenzdaten von nationalen Zustellern verglichen werden. Die Adressvalidierung kann auch Postinformationen hinzufügen, die die Postzustellung beschleunigen und die Zustellkosten reduzieren.
- **Nach doppelten Datensätzen suchen.** Die Analyse auf doppelte Datensätze vergleicht einen Satz von Datensätzen miteinander, um ähnliche oder übereinstimmende Werte in ausgewählten Datenspalten zu finden. Das Ähnlichkeitsniveau, welches eine gute Ähnlichkeit zwischen den Bereichswerten anzeigt, wird von Ihnen vorgegeben. Sie können auch die relative Gewichtung vorgeben, die jede Spalte in den Ähnlichkeitsberechnungen erhält. Sie können beispielsweise dem Nachnamen eine höhere Priorität als dem Vornamen geben.
- **Erstellen und Durchführen von Datenqualitätsregeln.** Informatica bietet vorgefertigte Regeln, die Sie ausführen oder bearbeiten können, um Ihre Projektziele zu erreichen. Sie können im Entwicklertool Regeln erstellen.
- **Zusammenarbeiten mit Informatica-Anwendern.** Die Regeln und Referenzdatentabellen, die Sie dem Model Repository zufügen, stehen den Anwendern im Entwicklertool und im Analysetool zur Verfügung. Benutzer können in Projekten zusammenarbeiten, und in unterschiedlichen Stadien eines Projekts können unterschiedliche Benutzer den Eigentümerstatus von Objekten übernehmen.
- **Exportieren von Mappings an PowerCenter.** Sie können Mappings an PowerCenter exportieren, um die Metadaten für die physische Datenintegration oder für die Erstellung von Internetdiensten zu verwenden.

Die Tutorial-Story

HypoStores Corporation ist eine nationale Einzelhandelsorganisation mit Firmensitz in Boston und Geschäften in mehreren Staaten. Sie integriert regelmäßig Betriebsdaten aus den landesweiten Geschäften in den Datenspeicher am Sitz des Unternehmens. Vor kurzem wurde ein Geschäft in Los Angeles eröffnet.

Zum Firmensitz gehört ein zentrales ICC-Team von Administratoren, Entwicklern und Architekten, die dafür verantwortlich sind, eine gemeinsame Datendienst-Layer für alle Verbund- und BI-Anwendungen zur Verfügung zu stellen. Die BI-Anwendungen umfassen ein CRM-System, das die Dateien mit den Kundenstammdaten enthält, die zur Abrechnung und zum Marketing verwendet werden.

HypoStores Corporation muss folgende Aufgaben durchführen, um Daten aus dem Betrieb in Los Angeles in die Daten am Firmensitz in Boston zu integrieren:

- Überprüfen der Daten von Boston und Los Angeles auf Qualitätsprobleme.
- Parsen von Informationen aus den Daten von Los Angeles.
- Standardisieren von Adressinformationen für die Daten von Boston and Los Angeles.
- Validieren der Genauigkeit der Postadresseninformationen in den Daten zum Zweck des Kundenbeziehungsmanagements (CRM).

Die Tutorial-Struktur

Die *Kurzanleitung* enthält Tutorials, die Lektionen und Aufgaben beinhalten.

Lektionen

In jeder Lektion werden Konzepte vorgestellt, die Ihnen das Verständnis der Aufgaben erleichtern, die in der Lektion bearbeitet werden. Die Lektion umfasst geschäftliche Anforderungen der gesamten Story. Die Ziele der Lektion erläutern die Aufgaben, die Sie durcharbeiten werden, um die geschäftlichen Anforderungen zu erfüllen. In jeder Lektion wird angegeben, wie viel Zeit ungefähr benötigt wird. Wenn Sie die Aufgaben in der Lektion abgeschlossen haben, können Sie die Zusammenfassung der Lektion einsehen.

Wenn die Umgebung innerhalb des Tools nicht konfiguriert ist, finden Sie in der ersten Lektion jedes Tutorials Hilfestellungen, wie dies erledigt wird.

Aufgaben

Die Aufgaben bestehen aus schrittweisen Anweisungen. Bearbeiten Sie alle Aufgaben in der angegebenen Reihenfolge, um die Lektion abzuschließen.

Informatica Analyst-Tutorial

Während dieses Tutorials meldet sich ein Analyst am Analysetool an, erstellt Projekte und Ordner, Profile und Regeln, bewertet Daten und erstellt Referenztabelle.

Welche Lektionen Sie durchführen können, hängt davon ab, ob Sie über Produkte von Informatica Data Quality oder Informatica Data Services verfügen.

In der folgenden Tabelle finden Sie die Lektionen, die Sie je nach Produkt durchführen können.

Lektion	Beschreibung	Produkt
Lektion 1. Einrichten von Informatica Analyst	Anmelden am Analysetool und Erstellen eines Projekts und Ordners für die Tutoriallektionen.	Data Quality Data Services
Lektion 2. Erstellen von Datenobjekten	Importieren einer Einfachdatei als Datenobjekt und Vorschau der Daten.	Data Quality
Lektion 3. Erstellen von Schnellprofilen	Erstellen eines Schnellprofils, um schnell eine Vorstellung von Data Quality zu erhalten.	Data Quality
Lektion 4. Erstellen von benutzerdefinierten Profilen	Erstellen eines benutzerdefinierten Profils zum Konfigurieren von Spalten sowie für Stichproben- und Drilldown-Optionen.	Data Quality
Lektion 5. Erstellen von Ausdrucksregeln	Erstellen von Expressionregeln zum Modifizieren und Profilieren von Spaltenwerten	Data Quality
Lektion 6. Erstellen und Ausführen von Scorecards	Erstellen und Ausführen einer Scorecard, um den Fortschritt der Datenqualität im Zeitverlauf zu messen.	Data Quality
Lektion 7. Erstellen von Referenztabellen aus Profilergebnissen	Erstellen einer Referenztable, die Sie zum Standardisieren von Quelldaten verwenden können.	Data Quality Data Services
Lektion 8. Erstellen von Referenztabellen	Erstellen einer Referenztable, um Beziehungen zwischen Quelldaten sowie gültigen und Standard-Werten einzurichten.	Data Quality Data Services

Informatica Developer Tool

In diesem Tutorial verwenden Sie das Developer Tool, um mehrere Vorgänge zur Datenqualität durchzuführen.

Benutzer von Informatica Data Quality verwenden das Developer Tool, um Prozesse zu entwerfen und auszuführen, die die Datenqualität verbessern. Informatica Data Quality-Benutzer verwenden das Developer Tool auch zum Erstellen und Ausführen von Profilen, die den Inhalt und die Struktur der Daten analysieren.

Arbeiten Sie folgende Lektionen im Tutorial Data Quality durch:

Lektion 1. Einrichten von Informatica Developer

Erstellen einer Verbindung zum Modellrepository, das von einem Modellrepository-Dienst in einer Domäne verwaltet wird. Erstellen eines Projekts und eines Ordners, um Arbeit für die Lektionen im Tutorial zu speichern. Auswählen eines Standard-Datenintegrationsdiensts.

Lektion 2. Importieren von physischen Datenobjekten

Sie werden Datenqualitätsprozesse für die Kundendatendateien definieren, die mit diesen Objekten assoziiert sind.

Lektion 3. Profiling von Daten

Profiling macht den Inhalt und die Struktur Ihrer Daten erkennbar.

Profiling umfasst Join-Analyse, eine Analyseform, die ermittelt, ob zwischen den beiden Datenspalten ein gültiger Join möglich ist.

Lektion 4. Parsen von Daten

Parsing erhöht den Nutzen Ihrer Datensätze und verbessert die Struktur des Datensatzes. Durch Parsen können Sie nützliche Informationen in Ihren Daten suchen und neue Informationen aus aktuellen Daten ableiten.

Lektion 5. Standardisieren von Daten

Standardisieren entfernt Datenfehler und Inkonsistenzen, die während des Profiling gefunden werden.

Lektion 6. Validieren von Adressdaten

Adressvalidierung bewertet die Genauigkeit und Zustellbarkeit Ihrer Postadressen und behebt Adressfehler und fehlende Angaben in Adressen.

Voraussetzungen für das Tutorial

Bevor Sie mit den Lektionen des Tutorials beginnen, muss die Informatica-Domäne mit mindestens einem eingerichteten Knoten laufen.

Der Installer umfasst Tutorialdateien, die Sie zur Bearbeitung der Lektionen verwenden. Diese Dateien finden Sie sowohl in Client- als auch in Server-Installationen:

- Alle Tutorialdateien finden Sie an dem folgenden Speicherort im Installationspfad des Developer Tools:

`<Informatica Installationsverzeichnis>\clients\DeveloperClient\Tutorials`

- Die Tutorialdateien finden Sie an dem folgenden Speicherort im Services-Installationspfad:

`<Informatica Installationsverzeichnis>\server\Tutorials`

Für die Tutoriallektionen benötigen Sie folgende Dateien:

- All_Customers.csv
- Boston_Customers.csv
- LA_customers.csv

Teil I: Erste Schritte mit Informatica Analyst

Dieser Teil enthält die folgenden Kapitel:

- [Lektion 1. Einrichten von Informatica Analyst, 19](#)
- [Lektion 2. Erstellen von Datenobjekten, 22](#)
- [Lektion 3. Erstellen von Standardprofilen, 25](#)
- [Lektion 4. Erstellen von Kundenprofilen, 29](#)
- [Lektion 5. Erstellen von Expressionsregeln, 33](#)
- [Lektion 6. Erstellen und Ausführen von Scorecards, 37](#)
- [Lektion 7. Erstellen von Referenztabelle aus Profilsäulen , 43](#)
- [Lektion 8. Erstellen von Referenztabelle, 46](#)

KAPITEL 2

Lektion 1. Einrichten von Informatica Analyst

Dieses Kapitel umfasst die folgenden Themen:

- [Einrichten von Informatica Analyst - Übersicht, 19](#)
- [Aufgabe 1. Anmelden bei Informatica Analyst, 20](#)
- [Aufgabe 2. Erstellen eines Projekts, 20](#)
- [Aufgabe 3. Erstellen eines Ordners, 21](#)
- [Einrichten von Informatica Analyst - Zusammenfassung, 21](#)

Einrichten von Informatica Analyst - Übersicht

Bevor Sie mit den Lektionen in diesem Tutorial beginnen können, müssen Sie das Analyst Tool einrichten. Um das Analyst Tool einzurichten, melden Sie sich beim Analyst Tool an und erstellen ein Projekt und einen Ordner, in dem Sie Ihre Arbeit speichern können.

Die Informatica-Domäne ist eine Sammlung von Knoten und Diensten, die die Informatica-Umgebung definieren. Zu den Diensten in der Domäne gehören der Analyst-Dienst und der Modellrepository-Dienst. Der Analyst-Dienst startet das Analyst Tool, und der Modellrepository-Dienst verwaltet das Modellrepository. Beim Arbeiten mit dem Analyst Tool speichert dieses die Objekte, die Sie im Modellrepository erstellen.

Sie müssen ein Projekt erstellen, bevor Sie im Analyst Tool Objekte erstellen können. Ein Projekt enthält Objekte im Analyst Tool. Ein Projekt kann auch Ordner enthalten, in denen zugehörige Objekte gespeichert werden, wie zum Beispiel Objekte, die Teile derselben Geschäftsanforderung sind.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

- Melden Sie sich beim Analyst Tool an.
- Erstellen Sie ein Projekt, um die Objekte zu speichern, die Sie im Analyst Tool erstellen.
- Erstellen Sie einen Ordner im Projekt, in dem Sie zugehörige Objekte speichern können.

Voraussetzungen

Überprüfen Sie die folgenden Voraussetzungen, bevor Sie mit dieser Lektion beginnen:

- Ein Administrator hat einen Modellrepository-Dienst und einen Analyst-Dienst im Administrator Tool konfiguriert.

- Sie haben den Hostnamen und die Portnummer für das Analyst Tool.
- Sie haben einen Benutzernamen und ein Passwort, um auf den Analyst-Dienst zuzugreifen. Sie erhalten diese Informationen von einem Domänenadministrator.

Zeitplan

Diese Lektion dauert ca. 5 bis 10 Minuten.

Aufgabe 1. Anmelden bei Informatica Analyst

Melden Sie sich beim Analyst Tool an, um mit dem Tutorial zu beginnen.

1. Starten Sie Microsoft Internet Explorer oder Google Chrome.
2. Geben Sie in der Adresszeile die URL für Informatica Analyst ein:
`http[s]://<fully qualified host name>:<port number>/analyst`
3. Wenn die Domäne die LDAP- oder native Authentifizierung verwendet, geben Sie Ihren Benutzernamen und das Passwort in der Anmeldeseite ein.
4. Wählen Sie **Nativ** oder den Namen einer spezifischen Sicherheitsdomäne aus.
Das Feld „Sicherheitsdomäne“ wird angezeigt, wenn die Informatica-Domäne die LDAP- oder Kerberos-Authentifizierung verwendet. Wenn Sie die Sicherheitsdomäne, zu der Ihr Benutzerkonto gehört, nicht kennen, wenden Sie sich an den Informatica Domänenadministrator.
5. Klicken Sie auf **Anmelden**.
Das Analyst Tool wird im Arbeitsbereich **Starten** geöffnet.

Aufgabe 2. Erstellen eines Projekts

In dieser Aufgabe erstellen Sie ein Projekt, um Objekte zu speichern, die Sie im Analyst-Tool erstellt haben. Erstellen Sie ein Tutorialprojekt, um den Ordner für das Projekt zu speichern.

1. Klicken Sie in der Kopfzeile **Verwalten** auf **Projekte**.
Der Arbeitsbereich **Projekte** wird angezeigt.
2. Klicken Sie im Menü **Aktionen** auf **Neu > Projekt**.
Das Fenster **Neues Projekt** wird geöffnet.
3. Geben Sie Ihren Namen mit vorangestelltem "Tutorial_" als Namen des Projekts ein.
4. Klicken Sie auf **OK**.

Aufgabe 3. Erstellen eines Ordners

In dieser Aufgabe erstellen Sie einen Ordner, um zugehörige Objekte zu speichern. Sie können einen Ordner in einem Projekt oder in einem anderen Ordner erstellen. Erstellen Sie einen Ordner namens „Kunden“, um die Objekte zu speichern, die zum Datenqualitätsprojekt gehören.

1. Wählen Sie im Bereich **Projekte** das Tutorialprojekt aus.
2. Klicken Sie im Menü **Aktionen** auf **Neu > Ordner**.
3. Geben Sie „Kunden“ als Ordnernamen ein.
4. Klicken Sie auf **OK**.

Der Ordner steht jetzt im Tutorialprojekt.

Einrichten von Informatica Analyst - Zusammenfassung

In dieser Lektion haben Sie gelernt, dass das Analyst-Tool Objekte in Projekten und Ordnern speichert. Ein Model Repository enthält Projekte und Ordner. Der Analyst-Dienst verwaltet das Analyst Tool. Der Modellrepository-Dienst verwaltet das Modellrepository. Der Analyst-Dienst und der Modellrepository-Dienst sind Anwendungsdienste in der Informatica-Domäne.

Sie haben sich beim Analyst Tool angemeldet und ein Projekt sowie einen Ordner erstellt.

Nun können Sie mit dem Analyst Tool weitere Lektionen in diesem Tutorial bearbeiten.

KAPITEL 3

Lektion 2. Erstellen von Datenobjekten

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen von Datenobjekten - Übersicht, 22](#)
- [Aufgabe 1. Erstellen der Einfachdatei-Datenobjekte, 23](#)
- [Aufgabe 2. Anzeigen der Datenobjekteigenschaften, 23](#)
- [Erstellen von Datenobjekten - Zusammenfassung, 24](#)

Erstellen von Datenobjekten - Übersicht

Ein Datenobjekt ist im Analyst Tool eine Darstellung von Daten, die auf einer Einfachdatei oder einer Tabelle einer relationalen Datenbank basieren. Sie erstellen eine Einfachdatei oder ein Tabellenobjekt und führen dann ein Profil mit den Daten in der Einfachdatei oder der Tabelle der relationalen Datenbank durch. Wenn Sie im Analyst Tool ein Einfachdatei-Datenobjekt erstellen, können Sie die Datei in den Einfachdatei-Cache des Rechners hochladen, auf dem das Analyst Tool läuft, oder Sie können den Speicherort für die Einfachdatei im Netzwerk angeben.

Story

HypoStores speichert die Kundendaten aus Los Angeles in Einfachdateien. HypoStores muss die Daten profilieren und analysieren und Datenqualitätsaufgaben damit durchführen.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

1. Hochladen der Einfachdatei in den Einfachdatei-Cache und Erstellen eines Datenobjekts.
2. Vorschau der Daten des Einfachdatei-Datenobjekts.

Voraussetzungen

Überprüfen Sie die folgenden Voraussetzungen, bevor Sie mit dieser Lektion beginnen:

- Sie haben Lektion 1 in diesem Tutorial abgeschlossen.
- Sie haben die Einfachdatei LA_Customers.csv. Diese Datei ist im Ordner <Installation Root Directory> \<Release Version>\clients\DeveloperClient\Tutorials gespeichert.

Zeitplan

Rechnen Sie mit 5 bis 10 Minuten für diese Aufgabe.

Aufgabe 1. Erstellen der Einfachdatei-Datenobjekte

In dieser Aufgabe erstellen Sie ein Einfachdatei-Datenobjekt aus der Datei LA_Customers.

1. Klicken Sie im Analyst-Tool auf **Neu > Einfachdatei-Datenobjekt**.
Der Assistent **Einfachdatei hinzufügen** wird eingeblendet.
2. Wählen Sie **Durchsuchen und Hochladen** aus und klicken Sie dann auf **Durchsuchen**.
3. Wechseln Sie zum Speicherort der Datei LA_Customers.csv und klicken Sie auf **Öffnen**.
4. Klicken Sie auf **Weiter**.
In der Maske **Importtyp wählen** werden die Optionen **Mit Trennzeichen** und **Feste Breite** angezeigt. Wählen Sie **Mit Trennzeichen** aus. Die Standardoption ist **Mit Trennzeichen**.
5. Klicken Sie auf **Weiter**.
6. Wählen Sie unter **Geben Sie die in Ihren Dateien verwendeten Delimiter und Textqualifikatoren an** die Option **Doppelte Anführungszeichen** als Textqualifikator aus.
7. Wählen Sie unter **Zeilen für den Import angeben** die Option **Aus erster Zeile importieren** aus, um Spaltennamen aus der ersten Zeile zu importieren, die nicht leer ist.
Der Bereich **Vorschau** wird aktualisiert und zeigt die Spaltenüberschriften aus der ersten Zeile an.
8. Klicken Sie auf **Weiter**.
In der Maske **Spaltenattribute** werden der Datentyp, die Präzision, die Größenordnung und das Format für jede Spalte angezeigt.
9. Klicken Sie auf **Weiter**.
Im Feld **Name** wird LA_Customers angezeigt.
10. Wählen Sie das Projekt **Tutorial_** und den Ordner **Kunden** aus.
11. Klicken Sie auf **Fertig stellen**.
Das Datenobjekt erscheint im Ordnerinhalt des Ordners Customers.

Aufgabe 2. Anzeigen der Datenobjekteigenschaften

In dieser Aufgabe können Sie die Eigenschaften des Datenobjekts LA_Customers anzeigen.

1. Klicken Sie auf **Öffnen**, um den Arbeitsbereich **Bibliothek** zu öffnen.
2. Klicken Sie im Arbeitsbereich **Bibliothek** im Bereich **Objekte** auf **Datenobjekte**.
Im Bereich **Datenobjekte** wird eine Liste mit Datenobjekten angezeigt.
3. Klicken Sie auf die Einfachdatei **LA_Customers**.
Der Bereich **Datenvorschau** wird mit den aus dem Datenobjekt LA_Customers abgerufenen Daten angezeigt.
4. Überprüfen Sie im Bereich **Datenvorschau** die Struktur und den Inhalt des Datenobjekts LA_Customers.
Das Analyst Tool zeigt die ersten 100 Zeilen des Einfachdatei-Datenobjekts.
5. Klicken Sie auf **Eigenschaften**.
Im Bereich **Eigenschaften** werden der Name, der Typ, die Beschreibung und der Speicherort des Datenobjekts angezeigt. Sie können auch den Spaltennamen und die Spalteneigenschaften für das Datenobjekt sehen.

Erstellen von Datenobjekten - Zusammenfassung

In dieser Lektion haben Sie erfahren, dass Datenobjekte Darstellungen von Daten sind, die auf einer Einfachdatei oder einer Quelle einer relationalen Datenbank basieren. Sie haben gelernt, dass Sie ein Einfachdatei-Datenobjekt erstellen und die darin enthaltenen Daten in einer Vorschau anzeigen können.

Sie haben eine Einfachdatei hochgeladen und ein Einfachdatei-Datenobjekt erstellt, die Vorschau der Daten des Datenobjekts angezeigt und die Eigenschaften des Datenobjekts angezeigt.

Nachdem Sie ein Datenobjekt erstellt haben, erstellen Sie in Lektion 3 ein Standardprofil für das Datenobjekt und in Lektion 4 ein benutzerdefiniertes Profil für das Datenobjekt.

KAPITEL 4

Lektion 3. Erstellen von Standardprofilen

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen von Standardprofilen – Übersicht, 25](#)
- [Aufgabe 1. Erstellen und Ausführen eines Standardprofils, 26](#)
- [Aufgabe 2. Anzeigen der Profilergebnisse in der Zusammenfassungsansicht, 27](#)
- [Erstellen von Standardprofilen – Zusammenfassung, 27](#)

Erstellen von Standardprofilen – Übersicht

Ein Profil ist die Analyse der Datenqualität, bezogen auf den Inhalt und die Struktur der Daten. Ein Standardprofil ist ein Profil, das Sie mit Standardoptionen erstellen. Mit einem Standardprofil können Sie Profilergebnisse erhalten, ohne alle Spalten und Optionen für ein Profil zu konfigurieren.

Erstellen Sie ein Standardprofil und führen Sie es aus, um die Qualität der Daten zu analysieren, wenn Sie ein Datenqualitätsprojekt starten. Wenn Sie ein Standardprofil erstellen, wählen Sie das Datenobjekt und die Spalten des Datenobjekts aus, die Sie analysieren möchten. Ein Standardprofil überspringt die Konfiguration der Profilspalten und -optionen. Das Analyst Tool führt ein Profiling mit der Live-Einfachdatei für das Einfachdatei-Datenobjekt durch.

Story

HypoStore möchte, dass Daten aus dem neu erworbenen Büro in Los Angeles in das Data Warehouse eingepflegt werden. Bevor diese Daten in das Data Warehouse eingepflegt werden können, müssen sie bereinigt werden. Sie sind der Analyst, der für die Bewertung der Datenqualität und die Weitergabe der Informationen an den Entwickler zuständig ist, der für die Bereinigung der Daten verantwortlich ist. Sie möchten die Profilergebnisse rasch anzeigen und eine Grundvorstellung der Datenqualität erhalten.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

1. Erstellen und Ausführen eines Standardprofils für das Einfachdatei-Datenobjekt LA_Customers.
2. Anzeigen der Profilergebnisse.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 und 2 in diesem Tutorial abgeschlossen.

Zeitplan

Rechnen Sie mit 5 bis 10 Minuten für diese Lektion.

Aufgabe 1. Erstellen und Ausführen eines Standardprofils

In dieser Aufgabe erstellen Sie ein Standardprofil für alle Spalten im Datenobjekt und verwenden Standardoptionen für Stichproben und Drilldown.

1. Wählen Sie im Arbeitsbereich **Bibliothek** ein Datenobjekt im Bereich **Objekte** aus.
2. Klicken Sie mit der rechten Maustaste auf das Datenobjekt und wählen Sie **Profil** aus.
Der Assistent **Neues Profil** wird eingeblendet.
3. Die Option **Einzelquelle** ist standardmäßig ausgewählt. Klicken Sie auf **Weiter**.
4. Im Bildschirm **Allgemeine Eigenschaften angeben** werden standardmäßig Name, Beschreibung und Speicherort eingetragen.
5. Klicken Sie auf **Weiter**.
6. Im Bildschirm **Quelle auswählen** wird standardmäßig das Datenobjekt angezeigt. Sie können die Spalten im Bereich **Spalten wählen** anzeigen.
7. Klicken Sie auf **Weiter**.
8. Auf dem Bildschirm **Einstellungen angeben** sind die folgenden Optionen standardmäßig ausgewählt:
 - **Spaltenprofil ausführen**
 - **Alle Zeilen** im Bereich **Spaltenprofil ausführen für**
 - **Live** im Bereich **Drilldown**
 - **Schließen Sie genehmigte Datentypen und Datendomänen in den nachfolgenden Profilausführungen aus der Datentyp- und Datendomäneninferenz aus.**
 - Verbindung **Nativ** im Bereich **Laufzeitumgebung**.
9. Klicken Sie auf **Weiter**.
10. Klicken Sie auf dem Bildschirm **Regeln und Filter angeben** auf **Speichern und ausführen**, um das Profil zu erstellen und auszuführen.

Das Analyst Tool erstellt das Profil und führt es aus. Die Profilergebnisse werden in der Zusammenfassungsansicht angezeigt.

Aufgabe 2. Anzeigen der Profilergebnisse in der Zusammenfassungsansicht

In dieser Aufgabe verwenden Sie die Zusammenfassungsansicht für das Profil LA_Customers, um einen raschen Überblick über die Profilergebnisse zu erhalten.

1. Klicken Sie im Bereich **Bibliothek > Objekte > Profile** auf das Profil **LA_Customers**.
Die Profilergebnisse werden in der Zusammenfassungsansicht angezeigt.
2. Klicken Sie in der Zusammenfassungsansicht im Bereich **Filtern nach** auf **Spalten**, um die Profilergebnisse für die Spalten anzuzeigen.
3. Bewegen Sie den Mauszeiger über die horizontalen Balkendiagramme, um die Werte als Prozentsätze anzuzeigen.
4. Im Abschnitt **Null eindeutig nicht-eindeutig %** können Sie die Nullwerte, eindeutigen Werte und nicht-eindeutigen Werte als Prozentsätze für eine Spalte anzeigen.
5. Im Abschnitt **Muster** können Sie mehrere Muster in den Spalten als horizontale Balkendiagramme anzeigen. Sie können die Musterzeichen und die Anzahl ähnlicher Muster in einer Spalte als Prozentsatz anzeigen, wenn Sie den Mauszeiger über das Balkendiagramm bewegen.
6. Im Abschnitt **Länge** können Sie die minimale und maximale Länge der Werte in der Spalte anzeigen.
7. Im Abschnitt **Wert** können Sie die minimalen und maximalen Werte in einer Spalte anzeigen.
8. Im Abschnitt **Datentyp** können Sie alle abgeleiteten Datentypen und dokumentierten Datentypen für eine Spalte anzeigen, wenn Sie den Mauszeiger über die Werte bewegen.
9. Im Abschnitt **Datendomäne** können Sie alle abgeleiteten Datendomänen für eine Spalte anzeigen, wenn Sie den Mauszeiger über die Werte bewegen.
10. Um Sonderfalldaten anzuzeigen, klicken Sie auf **Aktionen > Sonderfall erkennen**, um Sonderfälle in den Profilergebnissen zu erkennen.
11. Klicken Sie auf die Filter **Sonderfall des Musters** oder **Sonderfall der Werthäufigkeit**, um die Sonderfälle in den Profilergebnissen anzuzeigen.
12. Klicken Sie auf einen Spaltennamen, um die Profilergebnisse für die Spalte in der Detailansicht anzuzeigen.

Erstellen von Standardprofilen – Zusammenfassung

In dieser Lektion haben Sie erfahren, dass ein Standardprofil Profilergebnisse zeigt, ohne alle Stichprobenoptionen der Spalten und Zeilen für ein Profil zu konfigurieren. Sie haben gelernt, ein Standardprofil zu erstellen und auszuführen, um die Qualität der Daten zu analysieren, wenn Sie ein Datenqualitätsprojekt starten. Sie haben auch gelernt, dass das Analyst Tool Profiling mit der Live-Einfachdatei für das Einfachdatei-Datenobjekt durchführt.

Sie haben ein Standardprofil erstellt und die Ergebnisse des Profils analysiert. Sie haben weitere Informationen zu den Spalten in dem Profil erhalten, einschließlich Nullwerten und Datentypen. Sie haben auch die Spaltenwerte und Muster verwendet, um Datenqualitätsprobleme zu identifizieren.

Nachdem Sie die Ergebnisse eines Schnellprofils analysiert haben, können Sie die folgenden Aufgaben abschließen:

- Erstellen eines benutzerdefinierten Profils, um Spalten aus dem Profil auszuschließen und nur die Spalten einzuschließen, an denen Sie interessiert sind.

- Erstellen einer Expressionsregel, um virtuelle Spalten zu erstellen und sie zu profilieren.
- Erstellen einer Referenztafel, um gültige Werte für eine Spalte einzuschließen.

KAPITEL 5

Lektion 4. Erstellen von Kundenprofilen

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen von benutzerdefinierten Profilen - Übersicht, 29](#)
- [Aufgabe 1. Erstellen eines benutzerspezifischen Profils, 30](#)
- [Aufgabe 2. Ausführen des Profils, 31](#)
- [Aufgabe 3. Drilldown der Profilergebnisse, 31](#)
- [Erstellen von benutzerdefinierten Profilen - Zusammenfassung, 32](#)

Erstellen von benutzerdefinierten Profilen - Übersicht

Ein Profil ist die Analyse der Datenqualität, bezogen auf den Inhalt und die Struktur der Daten. Ein benutzerdefiniertes Profil ist ein Profil, das Sie erstellen, wenn Sie die Spalten, Stichprobenoptionen und Drilldown-Optionen für ein schnelleres Profiling konfigurieren möchten. Konfigurieren Sie Stichprobenoptionen, um die Stichprobenzeilen in der Quelle auszuwählen. Konfigurieren Sie die Drilldown-Optionen zum Drilldown auf Datenzeilen in den Quelldaten oder zwischengespeicherten Daten. Sie können das Profil in einer Hive- oder nativen Umgebung auswählen.

Sie erstellen ein Profil und führen es aus, um die Qualität der Daten zu analysieren, wenn Sie ein Datenqualitätsprojekt starten. Wenn Sie ein Profilobjekt erstellen, starten Sie mit der Auswahl des Datenobjekts und der Datenobjektspalten, für die Sie ein Profil ausführen möchten.

Story

HypoStore muss Daten aus dem neu erworbenen Büro in Los Angeles in das eigene Data Warehouse einpflegen. HypoStores möchte die Qualität der Kundenebenenendaten in der LA-Kundendatendatei bewerten. Sie sind der Analyst, der für die Bewertung der Datenqualität und die Weitergabe der Informationen an den Entwickler zuständig ist, der für die Bereinigung der Daten verantwortlich ist.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

1. Erstellen eines benutzerdefinierten Profils für das Einfachdatei-Datenobjekt und Ausschließen der Spalten mit Null-Werten.
2. Ausführen des Profils, um den Inhalt und die Struktur der Spalte CustomerTier zu analysieren.
3. Drilldown in die Zeilen für die Profilergebnisse.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1, 2 und 3 in diesem Tutorial abgeschlossen.

Zeitplan

Rechnen Sie mit 5 bis 10 Minuten für diese Lektion.

Aufgabe 1. Erstellen eines benutzerspezifischen Profils

In dieser Aufgabe erstellen Sie ein benutzerdefiniertes Profil. Wenn Sie ein benutzerdefiniertes Profil erstellen, wählen Sie das Datenobjekt und die Spalten aus, für die Sie ein Profil ausführen möchten. Sie können auch die Stichproben- und Drilldown-Optionen konfigurieren.

1. Klicken Sie auf **Neu > Profil**.

Der Assistent **Neues Profil** wird eingeblendet.

2. Die Option **Einzelquelle** ist standardmäßig ausgewählt. Klicken Sie auf **Weiter**.

3. Legen Sie im Bildschirm **Allgemeine Eigenschaften angeben** die folgenden Optionen fest:

- Geben Sie im Feld „Name“ den Wert **Profile_LA_Customers** ein.
- Wählen Sie im Feld „Speicherort“ den Ordner **Kunden** aus.

4. Klicken Sie auf **Weiter**.

5. Klicken Sie auf dem Bildschirm **Quelle auswählen** auf **Auswählen**.

Das Dialogfeld **Datenobjekt auswählen** wird eingeblendet.

6. Wählen Sie im Dialogfeld **Datenobjekt auswählen** die Option **LA_Customers** aus. Klicken Sie auf **OK**.

7. Löschen Sie auf dem Bildschirm **Quelle auswählen** die Spalten Address2, Address3, und City2.

8. Klicken Sie auf **Weiter**.

9. Legen Sie auf dem Bildschirm **Einstellungen angeben** die folgende Optionen fest:

- **Spaltenprofil ausführen**.
- Wählen Sie die Option **Statistische Probe** im Bereich **Profil ausführen für** aus.
- Wählen Sie die Option zum Ausschließen genehmigter Datentypen und Datendomänen in den nachfolgenden Profilausführungen aus der Datentyp- und Datendomäneninferenz aus.
- Wählen Sie die Option **Zwischengespeichert** im Bereich **Drilldown** aus.

10. Klicken Sie auf **Weiter**.

11. Klicken Sie im Bildschirm **Regeln und Filter angeben** auf **Speichern und beenden**, um das Profil zu erstellen und auszuführen.

Das Analyst Tool erstellt das Profil und zeigt das Profil im Arbeitsbereich **Discovery** an. Sie müssen das Profil ausführen, um die Ergebnisse anzuzeigen.

Aufgabe 2. Ausführen des Profils

In dieser Aufgabe führen Sie ein Profil aus, um an dem Datenobjekt Profilierung durchzuführen und die Profilergebnisse anzuzeigen. Das Analyst Tool führt ein Profiling mit der zwischengespeicherten Einfachdatei für das Einfachdatei-Datenobjekt durch.

1. Stellen Sie sicher, dass Sie sich im Arbeitsbereich **Discovery** befinden.
Sie können das Profil **Profile_LA_Customers_Custom** im Arbeitsbereich sehen.
2. Klicken Sie im Arbeitsbereich auf **Profile_LA_Customers_Custom**.
3. Der Profilbildschirm wird angezeigt, auf dem Sie das Profil bearbeiten oder ausführen können. Klicken Sie auf **Ausführen**.
4. Die Profilergebnisse werden in der Zusammenfassungsansicht angezeigt.

Aufgabe 3. Drilldown der Profilergebnisse

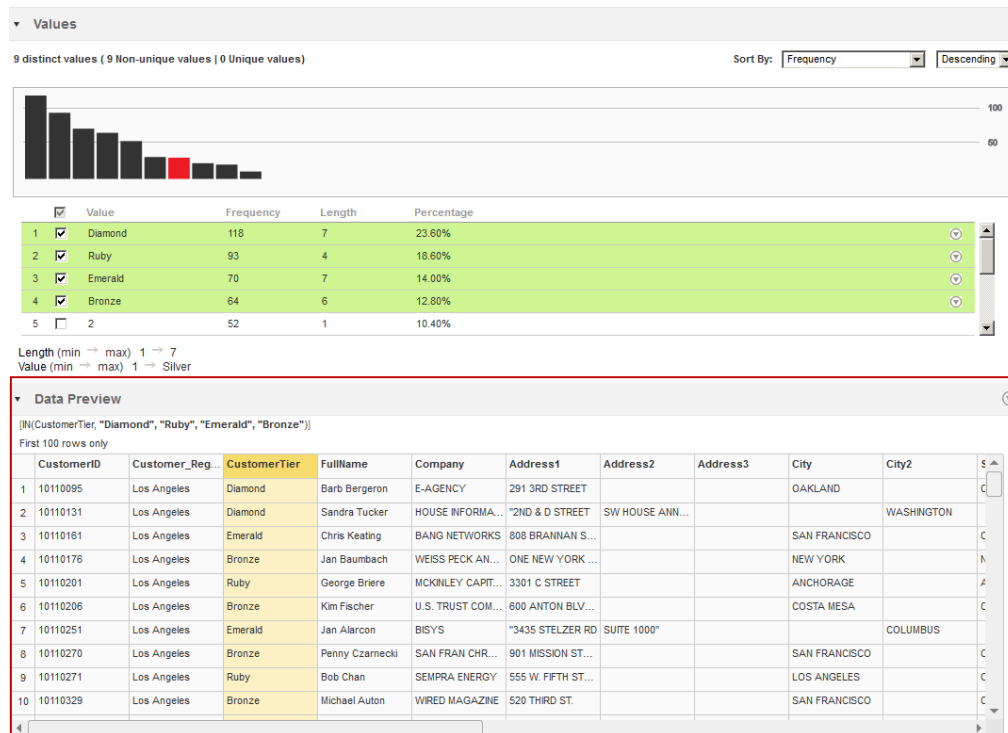
In dieser Aufgabe führen Sie ein Drilldown mit den Werten der Spalte CustomerTier durch, um die Quellzeilen in dem Datenobjekt für das Profil anzuzeigen.

1. Stellen Sie sicher, dass Sie sich in der Zusammenfassungsansicht der Profilergebnisse für das Profil **Profile_LA_Customers** befinden.
2. Klicken Sie auf die Spalte **CustomerTier**.
Die Profilergebnisse für die Spalte werden in der Detailansicht angezeigt.

- Wählen Sie in der Detailansicht die Werte „Diamond“, „Ruby“, „Emerald“ und „Bronze“ aus. Klicken Sie im Bereich **Werte** mit der rechten Maustaste auf die Werte und wählen Sie **Drilldown** aus.

Die Zeilen für die Spalten mit dem Wert „Diamond“, „Ruby“, „Emerald“ oder „Bronze“ werden im Bereich **Datenvorschau** angezeigt.

Die folgende Abbildung zeigt die Drilldown-Ergebnisse im Bereich **Datenvorschau**, wenn Sie den Drilldown für die Werte „Diamond“, „Ruby“, „Emerald“ oder „Bronze“ ausführen:



Der Bereich **Datenvorschau** zeigt die ersten 100 Zeilen für die ausgewählte Spalte an. Der Titel des Bereichs **Datenvorschau** zeigt die für die Quellspalte verwendete Logik an.

Erstellen von benutzerdefinierten Profilen - Zusammenfassung

In dieser Lektion haben Sie erfahren, dass Sie die zu profilierenden Spalten und die Stichproben- und Drilldown-Optionen konfigurieren können. Sie haben gelernt, dass Sie ein Drilldown durchführen können, um die zu Grunde liegenden Zeilen für die Spaltenwerte anzuzeigen, und dass Sie die Spalten konfigurieren können, die eingeschlossen werden, wenn Sie die Spaltenwerte anzeigen.

Sie haben ein benutzerdefiniertes Profil erstellt, das die Spalte CustomerTier enthält, das Profil ausgeführt und ein Drilldown auf die zu Grunde liegenden Zeilen für die Spalte CustomerTier in den Ergebnissen durchgeführt.

Verwenden Sie das benutzerdefinierte Profil, um in Lektion 5 eine Ausdrucksregel zu erstellen.

KAPITEL 6

Lektion 5. Erstellen von Expressionsregeln

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen von Ausdrucksregeln - Übersicht, 33](#)
- [Aufgabe 1. Erstellen von Expressionsregeln und Ausführen des Profils, 34](#)
- [Aufgabe 2. Anzeigen der Expressionsregel-Ausgabe, 34](#)
- [Aufgabe 3. Bearbeiten der Expressionsregeln, 35](#)
- [Erstellen von Expressionsregeln - Zusammenfassung, 36](#)

Erstellen von Ausdrucksregeln - Übersicht

Expressionsregeln verwenden Expressionsfunktionen und Quellspalten, um die Regellogik festzulegen. Sie können im Analyst Tool Expressionsregeln erstellen und einem Profil hinzufügen. Eine Expressionsregel kann mit einem oder mehreren Profilen assoziiert sein.

Die Ausgabe einer Expressionsregel ist eine virtuelle Spalte im Profil. Das Analyst Tool profiliert die virtuelle Spalte, wenn Sie das Profil ausführen.

Sie können Expressionsregeln verwenden, um Quellspalten zu validieren oder weitere Quellspalten bezogen auf den Wert der Quellspalten zu erstellen.

Story

HypoStore möchte, dass Daten aus dem neu erworbenen Büro in Los Angeles in das Data Warehouse eingepflegt werden. HypoStores möchte die Kundennamen analysieren und die Kundennamen in Vorname und Nachname trennen. HypoStores möchte Expressionsregeln verwenden, um eine Spalte, die Vornamen und Nachnamen enthält, in separate virtuelle Spalten zu parsen, und danach die Spalten profilieren. HypoStores möchte die Regeln auch anderen Analysten zur Verfügung stellen, die die Ausgabe dieser Regeln analysieren müssen.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

1. Erstellen von Expressionsregeln, um die Spalte FullName (Vollständiger Name) in die Spalten für Vornamen und Nachnamen zu trennen. Sie erstellen eine Regel, die den Vornamen von dem vollständigen Namen abtrennt. Sie erstellen eine weitere Regel, die den Nachnamen von dem Vornamen abtrennt. Sie erstellen diese Regeln für das Profil „Profile_LA_Customers_“.

2. Führen Sie das Profil aus, und zeigen Sie die Ausgabe der Regeln im Profil an.
3. Bearbeiten Sie die Regeln, um sie für andere Benutzer des Analyst Tools nutzbar zu machen.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben Lektionen 1 bis 4 abgeschlossen.

Zeitplan

Rechnen Sie mit 10 bis 15 Minuten für diese Lektion.

Aufgabe 1. Erstellen von Expressionsregeln und Ausführen des Profils

In dieser Aufgabe erstellen Sie zwei Expressionsregeln, um die Spalte FullName in zwei virtuelle Spalten mit den Bezeichnungen FirstName (Vorname) und LastName (Nachname) zu parsen. Die Regelnamen lauten FirstName und LastName.

1. Klicken Sie im Arbeitsbereich **Bibliothek** auf das Profil **Profile_LA_Customers**.
Die Profilergebnisse werden in der Zusammenfassungsansicht angezeigt.
2. Klicken Sie auf **Bearbeiten**.
Der **Profilassistent** wird angezeigt.
3. Klicken Sie auf den Bildschirm **Regeln und Filter angeben**.
4. Klicken Sie im Bereich **Regeln** auf **Aktionen > Regel erstellen**.
5. Geben Sie in das Feld **Name** die Option **FirstName** ein.
6. Geben Sie im Abschnitt **Ausdruck** folgenden Ausdruck ein, um den Vornamen von der Spalte **Name** zu trennen:

```
SUBSTR(FullName,1, INSTR(FullName, ' ', -1, 1) - 1)
```

7. Klicken Sie auf **Validieren**.
8. Klicken Sie auf **OK**.
9. Wiederholen Sie die Schritte 4 bis 8, um eine Regel namens **LastName** zu erstellen. Geben Sie den folgenden Ausdruck ein, um den Nachnamen von der Spalte **Name** zu trennen:

```
SUBSTR(FullName, INSTR(FullName, ' ', -1, 1), LENGTH(FullName))
```

10. Klicken Sie auf **Speichern und ausführen**, um das Profil zu speichern und auszuführen.

Aufgabe 2. Anzeigen der Expressionsregel-Ausgabe

In dieser Aufgabe zeigen Sie die Ausgabe von Ausdrucksregeln an, die nach Ausführung eines Profils Vornamen und Nachnamen trennen.

1. Klicken Sie in der Zusammenfassungsansicht auf **Bearbeiten**.
Der Profilassistent wird angezeigt.

2. Klicken Sie im Profilassistenten auf **Quelle auswählen**.
3. Aktivieren Sie auf dem Bildschirm **Quelle auswählen** das Kontrollkästchen neben **Name** auf der Symbolleiste, um alle Spalten zu löschen.
Eine der Spalten ist standardmäßig ausgewählt, da mindestens eine Spalte im Abschnitt **Spalten** ausgewählt werden muss.
4. Wählen Sie die Spalte **Vollständiger Name** und die Regeln **Vorname** und **Nachname** aus.
5. Heben Sie die Auswahl aller anderen Spalten auf.
6. Klicken Sie auf **Speichern und ausführen**.
Die Profilergebnisse werden in der Zusammenfassungsansicht angezeigt.
7. Klicken Sie auf die Regel **Vorname**. Die Profilergebnisse für die Regel werden in der Detailansicht angezeigt.
8. Wählen Sie im Bereich **Werte** einen Wert aus. Klicken Sie mit der rechten Maustaste auf den Wert und klicken Sie auf **Drilldown**.
Die Werte für die Spalte **Vollständiger Name** und die Regeln **Vorname** und **Nachname** werden im Bereich **Datenvorschau** zusammen mit anderen Spaltenwerten angezeigt. Beachten Sie, dass das Analyst Tool die Spalte **Vollständiger Name** in Vornamen und Nachnamen trennt.

Aufgabe 3. Bearbeiten der Expressionsregeln

In dieser Aufgabe machen Sie die Expressionsregeln wiederverwendbar und stellen Sie allen Benutzern des Analyst Tools zur Verfügung.

1. Klicken Sie in der Zusammenfassungsansicht für das Profil **Profile_LA_Customers** auf **Bearbeiten**.
Der Profilassistent wird angezeigt.
2. Klicken Sie auf den Bildschirm **Regeln und Filter angeben**.
3. Wählen Sie auf dem Bildschirm **Regeln und Filter angeben** die Regel **Vorname** aus und klicken Sie auf **Aktionen > Regel bearbeiten**.
Das Dialogfeld **Regel bearbeiten** wird angezeigt.
4. Wählen Sie die Option **Möchten Sie diese Regel als wiederverwendbare Regel speichern?** aus und wählen Sie einen Speicherort zum Speichern der Regel.
5. Klicken Sie auf **OK**.
6. Wählen Sie die Regel **Nachname** aus und wiederholen Sie die Schritte 3 bis 5.
7. Klicken Sie auf **Speichern und beenden**, um das Profil zu speichern.
Alle Benutzer des Analyst Tools können die Regeln FirstName und LastName zum Aufteilen einer Spalte mit Vor- und Nachnamen in getrennten Spalten verwenden.

Erstellen von Expressionsregeln - Zusammenfassung

In dieser Lektion haben Sie erfahren, dass Expressionsregeln Expressionsfunktionen und Quellspalten zum Definieren von Regellogik verwenden. Sie haben gelernt, dass die Ausgabe einer Expressionsregel eine virtuelle Spalte im Profil ist. Das Analysetool enthält die virtuelle Spalte, wenn Sie das Profil ausführen.

Sie haben zwei Expressionsregeln erstellt, diese zu einem Profil hinzugefügt und das Profil ausgeführt. Sie haben die Ausgabe der Regeln angezeigt und sie allen Benutzern des Analysetools zur Verfügung gestellt.

KAPITEL 7

Lektion 6. Erstellen und Ausführen von Scorecards

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen und Ausführen von Scorecards - Übersicht, 37](#)
- [Aufgabe 1. Erstellen einer Scorecard aus den Profilergebnissen, 38](#)
- [Aufgabe 2. Ausführen der Scorecard, 39](#)
- [Aufgabe 3. Anzeigen der Scorecard, 39](#)
- [Aufgabe 4. Bearbeiten der Scorecard, 40](#)
- [Aufgabe 5. Konfigurieren von Schwellenwerten, 41](#)
- [Aufgabe 6. Anzeigen von Score-Trenddiagrammen, 41](#)
- [Erstellen und Ausführen von Scorecards - Zusammenfassung, 41](#)

Erstellen und Ausführen von Scorecards - Übersicht

Eine Scorecard ist die grafische Darstellung von gültigen Werten für eine Spalte oder die Ausgabe einer Regel in Profilergebnissen. Verwenden Sie Scorecards zum Messen und Überwachen des Fortschritts der Datenqualität im Zeitverlauf.

Zum Erstellen einer Scorecard fügen Sie Spalten aus dem Profil als Metriken zu einer Scorecard hinzu, weisen Gewichtungen Metriken zu und konfigurieren die Score-Schwellenwerte. Sie können den Scorecards basierend auf den Quelldaten Filter hinzufügen. Um eine Scorecard auszuführen, wählen Sie die gültigen Werte für die Metrik aus und führen Sie die Scorecard aus, um die Scores für die Metriken anzuzeigen.

Scorecards zeigen die Werthäufigkeit von Spalten in einem Profil als Scores (Punkte). Scores spiegeln den Prozentsatz von gültigen Werten für eine Metrik wider.

Story

HypoStore möchte, dass Daten aus dem neu erworbenen Büro in Los Angeles in das Data Warehouse eingepflegt werden. Bevor das Unternehmen die Daten zusammenführt, muss sichergestellt werden, dass die Daten auf den verschiedenen Kundenebenen und in den verschiedenen Stadien hinsichtlich der Datenqualität analysiert werden. Sie sind der Analyst, der für die Überwachung des Fortschritts bei der Durchführung der Datenqualitätsanalyse verantwortlich ist. Sie möchten eine Scorecard aus den Spalten „Kundenebene“ und „Statusprofil“ erstellen, Schwellenwerte für die Datenqualität konfigurieren und die Score-Trenddiagramme anzeigen, um zu ermitteln, wie sich die Scores im Zeitablauf verbessern.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

1. Erstellen einer Scorecard aus den Ergebnissen des Profils **Profile_LA_Customers_Custom**, um die Scores für die Spalten **CustomerTier** und **State** anzuzeigen.
2. Ausführen der Scorecard, um die Scores für die Spalten **CustomerTier** und **State** zu erstellen.
3. Anzeigen der Scorecard, um die Scores für jede Spalte anzuzeigen.
4. Bearbeiten der Scorecard, um unterschiedliche gültige Werte für die Scores anzugeben.
5. Konfigurieren Sie Score-Schwellenwerte und führen Sie die Scorecard aus.
6. Anzeigen von Score-Trenddiagrammen, um zu ermitteln, wie sich die Scores im Zeitverlauf verbessern.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 bis 5 in diesem Tutorial abgeschlossen.

Zeitplan

Planen Sie für die vollständige Durchführung der Aufgaben in dieser Lektion 15 Minuten ein.

Aufgabe 1. Erstellen einer Scorecard aus den Profilergebnissen

In dieser Aufgabe erstellen Sie eine Scorecard aus dem Profil **Profile_LA_Customers_Custom**, um die Spaltenwerte von **CustomerTier** und **State** zu bewerten.

1. Klicken Sie im Arbeitsbereich **Bibliothek** auf das Profil **Profile_LA_Customers**.
Daraufhin wird die Zusammenfassungsansicht der Profilergebnisse angezeigt.
2. Wählen Sie in der Zusammenfassungsansicht die Spalte **CustomerTier** aus, klicken Sie mit der rechten Maustaste und wählen Sie **Hinzufügen zu > Scorecard** aus.
Der Assistent **Zu Scorecard hinzufügen** wird eingeblendet.
3. Im Assistenten **Zu Scorecard hinzufügen** ist die Option **Neue Scorecard** standardmäßig ausgewählt. Klicken Sie auf **Weiter**.
4. Geben Sie im Bildschirm **Schritt 2 von 8** als Scorecard-Namen **sc_LA_Customer** ein und navigieren Sie zum Ordner **Kunden** für den Scorecard-Speicherort.
5. Klicken Sie auf **Weiter**.
6. Wählen Sie im Bildschirm **Schritt 3 von 8** die Spalten **CustomerTier** und **Bundesland** aus, um sie zur Scorecard hinzuzufügen.
7. Klicken Sie auf **Weiter**.
8. Im Bildschirm **Schritt 4 von 8** können Sie Filter für die Metriken erstellen, bearbeiten oder löschen. In diesem Tutorial wird kein Scorecard-Filter erstellt. Klicken Sie auf **Weiter**.
9. Wählen Sie im Bildschirm **Schritt 5 von 8** im Bereich **Metriken** die Metrik **CustomerTier** aus.
10. Wählen Sie im Abschnitt **Score verwendet: Werte** alle Werte aus und klicken Sie auf **Alle hinzufügen**, um die Werte in den Abschnitt **Gültige Werte** zu verschieben.
Verwenden Sie die **Umschalttaste**, um mehrere Werte auszuwählen.

11. Markieren Sie im Bereich **Metriken** die Metrik **Bundesland** und wählen Sie die Werte aus, die im Abschnitt **Score verwendet: Werte** Staatencodes mit zwei Buchstaben haben.
12. Klicken Sie auf **Hinzufügen**, um die Werte in den Abschnitt **Gültige Werte** zu verschieben.
Sie können die Gesamtanzahl der gültigen Werte und den Prozentsatz an gültigen Werten oben im Abschnitt sehen.
13. Akzeptieren Sie für jede Metrik im Abschnitt **Metriken** die Standardeinstellungen für die Score-Schwellenwerte im Abschnitt **Metrikschwellenwerte**.
14. Klicken Sie auf **Weiter**.
15. Im Bildschirm **Schritt 6 von 8** können Sie optional eine Metrikgruppe auswählen, um sie zu den Metriken hinzuzufügen. Standardmäßig fügt das Analyst Tool die Metriken zur Metrikgruppe **Standard** hinzu.
16. Klicken Sie auf **Weiter**.
17. Doppelklicken Sie im Bildschirm **Schritt 7 von 8** auf die Spalte **Gewichtung** für die Metrik **CustomerTier** im Bereich **Standard - Metriken**.
Wenn Sie eine Scorecard ausführen, berechnet das Analyst Tool den gewichteten Durchschnitt für jede Metrikgruppe basierend auf dem Metrik-Score und der Gewichtung, die Sie jeder Metrik zuweisen.
18. Geben Sie eine Gewichtung für die Metriken **CustomerTier** und **Status** ein.
19. Klicken Sie auf **Weiter**.
20. Im Bildschirm **Schritt 8 von 8** ist die Option **Nativ** standardmäßig ausgewählt. Klicken Sie auf **Speichern** zum Erstellen der Scorecard.
Die Scorecard wird im Arbeitsbereich **Scorecards** angezeigt.

Aufgabe 2. Ausführen der Scorecard

In dieser Aufgabe führen Sie die Scorecard sc_LA_Customer aus, um die Scores für die Spalten CustomerTier und State zu erstellen.

1. Klicken Sie im Arbeitsbereich **Bibliothek** auf **Objekte > Scorecards**.
2. Wählen Sie eine Scorecard im Fenster **Scorecards** aus.
3. Klicken Sie auf **Aktionen > Öffnen**.
Die Scorecard wird angezeigt.
4. Klicken Sie auf **Aktionen > Scorecard ausführen**
Das Dialogfeld **Scorecard ausführen** wird angezeigt.
5. Überprüfen Sie die Einstellungen im Dialogfeld und klicken Sie auf **Ausführen**.
Im Arbeitsbereich **Scorecards** werden die Scores für die Spalten „CustomerTier“ und „State“ angezeigt.

Aufgabe 3. Anzeigen der Scorecard

In dieser Aufgabe zeigen Sie die Scorecard sc_LA_Customer an, um die Scores für die Spalten CustomerTier und State zu sehen.

1. Wählen Sie die Zeile **State**, die den Score für State enthält, den Sie anzeigen möchten

Im Abschnitt **sc_LA_Customer - metrics** können Sie die folgenden Eigenschaften der Scorecard anzeigen:

- Scorecard-Name
 - Gesamtanzahl der Zeilen in der Scorecard
 - Anzahl der ungültigen Zeilen
 - Score zusammen mit einem horizontalen Balkendiagramm
 - Score-Trend. Sie können auf den Score-Trend klicken, um auf dem Bildschirm **Trenddiagramm-Detail** eine grafische Darstellung anzuzeigen.
 - Gewichtung der Metrik
 - Kosten für ungültige Daten
 - Kostenentwicklung
 - Datenobjekt. Klicken Sie auf das Datenobjekt, um die Datenvorschau des Datenobjekts im **Entdeckungs**-Arbeitsbereich anzuzeigen.
 - Spalten- oder Regelname
 - Quellentyp
 - Drilldown-Symbol
2. Klicken Sie in der Zeile „Status“ auf das Drilldown-Symbol.
Die Scores, die für die Spalte „Status“ ungültig sind, werden im Bereich **Drilldown** im Abschnitt **Ungültige Zeilen** angezeigt.
 3. Wählen Sie **Gültige Zeilen**, um die für die Spalte State gültigen Zeilen anzuzeigen.
 4. Klicken Sie in der Zeile **CustomerTier** auf das Drilldown-Symbol.
Alle Scores für die Spalte **CustomerTier** sind gültig.

Aufgabe 4. Bearbeiten der Scorecard

In dieser Aufgabe bearbeiten Sie die Scorecard sc_LA_Customer, um anzugeben, dass der Wert Ruby für den Score CustomerTier ungültig ist.

1. Stellen Sie sicher, dass Sie sich im Arbeitsbereich **Scorecard** befinden und dass die Scorecard sc_LA_Customer geöffnet ist.
2. Wählen Sie **Aktionen > Bearbeiten > Metriken**.
Das Dialogfeld **Scorecard bearbeiten** wird angezeigt.
3. Wählen Sie im Abschnitt **Metriken** die Option **CustomerTier**.
4. Verschieben Sie im Abschnitt **Score verwendet: Werte** den Wert **Ruby** aus dem Abschnitt **Gültige Werte** in den Abschnitt **Verfügbare Werte**.
Akzeptieren Sie die Standardeinstellungen im Abschnitt **Metrikschwellenwerte**.
5. Klicken Sie auf **Speichern & Ausführen**, um die Änderungen an der Scorecard zu speichern und diese auszuführen.
6. Zeigen Sie den Score CustomerTier erneut an.
Der Score CustomerTier ändert sich auf 81,4 Prozent.

Aufgabe 5. Konfigurieren von Schwellenwerten

In dieser Aufgabe konfigurieren Sie Schwellenwerte für den Score State in der Scorecard sc_LA_Customer, um die akzeptablen Bereiche für die Daten in der Spalte State zu ermitteln. Werte mit einem 2-Buchstaben-Code, wie CA, sind akzeptabel, und Codes mit mehr als zwei Buchstaben, wie Calif, sind nicht akzeptabel.

1. Stellen Sie sicher, dass Sie sich im Arbeitsbereich **Scorecard** befinden und dass die Scorecard sc_LA_Customer geöffnet ist.
2. Wählen Sie **Aktionen > Bearbeiten > Metriken**.
Das Dialogfeld **Scorecard bearbeiten** wird angezeigt.
3. Wählen Sie im Abschnitt **Metriken** die Option **State**.
4. Geben Sie im Abschnitt **Metrikschwellenwerte** die folgenden Bereiche für die Scores Gut und Inakzeptabel ein: 90 bis 100 % Gut; 0 bis 50 % Inakzeptabel; 51 bis 89 % Akzeptabel.
Die Schwellenwerte stehen für die unteren Grenzen der akzeptablen und guten Bereiche.
5. Klicken Sie auf **Speichern & Ausführen**, um die Änderungen an der Scorecard zu speichern und diese auszuführen.

Zeigen Sie im Bereich **Scorecard** die Änderungen an den Score-Prozentsätzen und dem Score an, der als Balken für den Score State angezeigt wird.

Aufgabe 6. Anzeigen von Score-Trenddiagrammen

In dieser Aufgabe zeigen Sie das Trenddiagramm für den Score State an. Sie können Trenddiagramme anzeigen, um Scores im Zeitverlauf zu überwachen.

1. Stellen Sie sicher, dass Sie sich im Arbeitsbereich **Scorecard** befinden und dass die Scorecard sc_LA_Customer geöffnet ist.
2. Wählen Sie die Zeile **Status** aus.
3. Klicken Sie auf **Aktionen > Trenddiagramm anzeigen** oder klicken Sie auf den Pfeil unter der Spalte **Score-Trend**.
Das Dialogfeld **Trenddiagramm-Detail** wird angezeigt. Sie können die Schwellenwerte **Gut**, **Akzeptabel** und **Inakzeptabel** für den Score anzeigen. Die Schwellenwerte ändern sich jedes Mal, wenn Sie die Scorecard ausführen, nachdem Sie die Werte für Scores in der Scorecard bearbeitet haben.
4. Zeigen Sie auf einen beliebigen Kreis im Diagramm zum Anzeigen der gültigen Werte im Abschnitt **Gültige Werte** unten im Diagramm.
5. Klicken Sie auf **Schließen**, um zur Scorecard zurückzukehren.

Erstellen und Ausführen von Scorecards - Zusammenfassung

In dieser Lektion haben Sie gelernt, dass Sie eine Scorecard aus den Ergebnissen eines Profils erstellen können. Eine Scorecard enthält die Spalten aus einem Profil. Sie haben gelernt, dass Sie eine Scorecard ausführen können, um Scores für Spalten zu erstellen. Sie haben eine Scorecard bearbeitet, um gültige Werte

zu konfigurieren und Schwellenwerte für Scores festzulegen. Sie haben auch erfahren, wie das Score-Trenddiagramm angezeigt wird.

Sie haben eine Scorecard aus den Spalten CustomerTier und State in einem Profil erstellt, um die Datenqualität für die Spalten CustomerTier und State zu analysieren. Sie haben die Scorecard ausgeführt, um Scores für jede Spalte zu erstellen. Sie haben die Scorecard bearbeitet, um unterschiedliche gültige Werte für Scores anzugeben. Sie haben Schwellenwerte für einen Score konfiguriert und das Score-Trenddiagramm angezeigt.

KAPITEL 8

Lektion 7. Erstellen von Referenztabellen aus Profilspalten

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen von Referenztabellen aus Profilspalten - Übersicht, 43](#)
- [Aufgabe 1. Erstellen einer Referenztable aus Profilspalten, 44](#)
- [Aufgabe 2. Bearbeiten der Referenztable, 45](#)
- [Erstellen von Referenztabellen aus Profilspalten - Zusammenfassung, 45](#)

Erstellen von Referenztabellen aus Profilspalten - Übersicht

Eine Referenztable enthält Referenzdaten, die zum Standardisieren von Quelldaten verwendet werden können. Referenzdaten können gültige und Standardwerte einschließen. Erstellen Sie Referenztabellen, um Beziehungen zwischen Quelldatenwerten und den gültigen und Standardwerten einzurichten.

Sie können eine Referenztable aus den Ergebnissen eines Profils erstellen. Nachdem Sie eine Referenztable erstellt haben, können Sie sie bearbeiten, um Spalten oder Zeilen hinzuzufügen oder Standard- und gültige Werte hinzuzufügen oder zu bearbeiten. Sie können die Änderungen, die an einer Referenztable vorgenommen wurden, in einem Audit-Trail anzeigen.

Story

HypoStores möchte die Daten profilieren, um Abweichungen aufzuzeigen und die Daten mit gültigen Werten zu standardisieren. Sie sind der Analyst, der für die Standardisierung der gültigen Werte in den Daten verantwortlich ist. Sie möchten eine Referenztable auf Grundlage von gültigen Werten aus Profilspalten erstellen.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

1. Erstellen einer Referenztable aus der Spalte CustomTier im Profil Profile_LA_Customers_Custom durch Auswahl von gültigen Werten für Spalten.
2. Bearbeiten der Referenztable, um unterschiedliche gültige Werte für die Spalten zu konfigurieren.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 bis 6 in diesem Tutorial abgeschlossen.

Zeitplan

Planen Sie für die vollständige Durchführung der Aufgaben in dieser Lektion 15 Minuten ein.

Aufgabe 1. Erstellen einer Referenztable aus Profilspalten

In dieser Aufgabe erstellen Sie eine Referenztable und fügen die Spalte „CustomerTier“ aus dem Profil „Profile_LA_Customers“ zur Referenztable hinzu.

1. Klicken Sie im Arbeitsbereich **Bibliothek** auf **Objekte > Profile**.
2. Klicken Sie auf das Profil **Profile_LA_Customers**, um die Profilergebnisse in der Zusammenfassungsansicht zu öffnen.
3. Wählen Sie in der Zusammenfassungsansicht die Spalte CustomerTier aus, die Sie der Referenztable hinzufügen möchten. Klicken Sie mit der rechten Maustaste darauf und wählen Sie **Hinzufügen zu Referenztable** aus.

Das Dialogfeld **Hinzufügen zu Referenztable** wird angezeigt.

4. Wählen Sie **Eine Referenztable erstellen**.
5. Klicken Sie auf **Weiter**.
6. Geben Sie im Feld **Name** den Wert **Reftab_CustTier_HypoStores** ein.
7. Geben Sie eine Beschreibung ein und legen Sie als Standardwert 0 fest.
Das Analyst Tool verwendet den Standardwert für jeden Tabellendatensatz, der keinen Wert enthält.
8. Klicken Sie auf **Weiter**.
9. Konfigurieren Sie im Abschnitt **Spaltenattribute** die folgenden Spalteneigenschaften für die Spalte CustomerTier:

Eigenschaft	Beschreibung
Name	CustomerTier
Datentyp	Zeichenfolge
Gesamtstellenanzahl	10
Beschreibung	Referenzwerte für CustomerTier

10. Optional können Sie eine Beschreibungsspalte für Zeilen in der Referenztable erstellen. Geben Sie den Namen und die Gesamtstellenanzahl für die Spalte ein.
11. Überprüfen Sie die Werte in der Spalte CustomerTier im Abschnitt **Vorschau**.
12. Klicken Sie auf **Weiter**.

Die Referenztable Reftab_CustomerTier_HypoStores wird eingeblendet. Sie können gegebenenfalls eine Beschreibung eingeben.

13. Wählen Sie im Abschnitt **Speichern unter** Ihr Tutorialprojekt, in dem Sie die Referenztable erstellen möchten.

Der Bereich **Referenztabellen** zeigt die Referenztabellen an dem von Ihnen gewählten Speicherort.

14. Geben Sie eine optionale Audit-Anmerkung ein.
15. Klicken Sie auf **Fertig stellen**.

Aufgabe 2. Bearbeiten der Referenztable

Sie bearbeiten in dieser Aufgabe die Table Reftab_CustomerTier_HypoStores, um alternative Werte für die Kundenebenen hinzuzufügen.

1. Klicken Sie im Arbeitsbereich **Bibliothek** auf **Objekte > Referenztabellen**.
2. Klicken Sie auf die Referenztable „Reftab_CustomerTier_HypoStores“.
Die Referenztable wird im Arbeitsbereich **Design** geöffnet.
3. Um eine Zeile zu bearbeiten, wählen Sie die Zeile aus und klicken auf **Aktionen > Bearbeiten** oder klicken auf das Symbol **Bearbeiten**.
Das Dialogfeld **Zeile bearbeiten** wird angezeigt. Wählen Sie gegebenenfalls mehrere Zeilen, um jeder Zeile den gleichen alternativen Wert hinzuzufügen.
4. Geben Sie für die Zeilen „Diamond“, „Emerald“, „Gold“, „Silver“ und „Bronze“ folgende alternative Werte ein: 1, 2, 3, 4, 5.
Geben Sie eine optionale Audit-Anmerkung ein.
5. Klicken Sie auf **Übernehmen**, um die Änderungen zu übernehmen.
6. Klicken Sie auf **Schließen**
Die geänderten Referenztable Werte werden im Arbeitsbereich **Design** angezeigt.

Erstellen von Referenztabellen aus Profilspalten - Zusammenfassung

Sie haben in dieser Lektion erfahren, wie Referenztabellen aus den Ergebnissen eines Profils erstellt werden, um gültige Werte für Quelldaten zu konfigurieren.

Sie haben eine Referenztable aus einer Profilspalte erstellt, indem gültige Werte für die Spalten ausgewählt wurden. Sie haben die Referenztable bearbeitet, um unterschiedliche gültige Werte für die Spalten zu konfigurieren.

KAPITEL 9

Lektion 8. Erstellen von Referenztabellen

Dieses Kapitel umfasst die folgenden Themen:

- [Erstellen von Referenztabellen - Übersicht, 46](#)
- [Aufgabe 1. Erstellen einer Referenztable, 47](#)
- [Erstellen von Referenztabellen - Zusammenfassung, 47](#)

Erstellen von Referenztabellen - Übersicht

Eine Referenztable enthält Referenzdaten, die zum Standardisieren von Quelldaten verwendet werden können. Referenzdaten können gültige und Standardwerte einschließen. Erstellen Sie Referenztabellen, um Beziehungen zwischen den Quelldatenwerten und den gültigen und Standard-Werten einzurichten.

Eine Referenztable kann manuell mit dem Referenztabellen-Editor erstellt werden. Verwenden Sie die Referenztable, um die Quelldaten zu definieren und zu standardisieren. Sie können die Referenztable gemeinsam mit einem Entwickler nutzen, um sie im Standardisierer und in Lookup-Transformationen im Developer Tool zu verwenden.

Story

HypoStores möchte Daten mit gültigen Werten standardisieren. Sie sind der Analyst, der für die Standardisierung der gültigen Werte in den Daten verantwortlich ist. Sie möchten eine Referenztable erstellen, um Standard-Tier-Codes für Kunden zu definieren, die die LA-Kundendaten referenzieren. Danach können Sie die Referenztable gemeinsam mit einem Entwickler nutzen.

Ziele

In dieser Lektion führen Sie folgende Aufgabe durch:

- Erstellen einer Referenztable mit dem Referenztabellen-Editor, um Standard-Tier-Codes für Kunden zu definieren, die die LA-Kundendaten referenzieren.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 und 2 in diesem Tutorial abgeschlossen.

Zeitplan

Rechnen Sie mit 10 Minuten, um die Aufgabe in dieser Lektion abzuschließen.

Aufgabe 1. Erstellen einer Referenztablelle

In dieser Aufgabe erstellen Sie die Referenztablelle Reftab_CustomerTier_Codes, um die gültigen Werte für die CustomerTier-Daten zu standardisieren.

1. Klicken Sie auf **Neue > Referenztablelle**.

Der Assistent **Neue Referenztablelle** wird eingeblendet.

2. Wählen Sie **Referenztablellen-Editor verwenden**.

3. Klicken Sie auf **Weiter**.

4. Klicken Sie für jede Spalte, die Sie in der Referenztablelle hinzufügen möchten, auf das Symbol **Neue Spalte hinzufügen**, und konfigurieren Sie die Spalteneigenschaften für jede Spalte.

Fügen Sie folgende Spaltennamen hinzu: CustomerID (Kunden-ID), CustomerTier (Kundenebene) und Status. Sie können die Spalten neu ordnen oder Spalten löschen.

5. Geben Sie optional eine Beschreibung ein und setzen Sie den Standardwert auf **0**.

Das Analyst-Tool verwendet den Standardwert für jeden Tabellendatensatz, der keinen Wert enthält.

6. Klicken Sie auf **Weiter**.

7. Geben Sie im Feld **Name** den Wert **Reftab_CustomerTier_Codes** ein.

8. Wählen Sie im Abschnitt **Ordner** im Tutorialprojekt den Ordner **Kunden**.

9. Klicken Sie auf **Fertig stellen**.

Die Referenztablelle wird im Arbeitsbereich **Design** angezeigt.

10. Wählen Sie im Menü **Aktionen** die Option **Zeile hinzufügen**, um jede Spalte der Referenztablelle mit den folgenden vier Werten zu füllen:

CustomerID = LA1, LA2, LA3, LA4

CustomerTier = 1, 2, 3, 4.

Status = Aktiv, Inaktiv

Erstellen von Referenztablellen - Zusammenfassung

In dieser Lektion haben Sie erfahren, wie Referenztablellen mit dem Referenztablellen-Editor erstellt werden, um gültige Standardwerte zu erstellen, die mit Quelldaten verwendet werden.

Sie haben eine Referenztablelle unter Verwendung des Referenztablellen-Editors erstellt, um Kundenebenenwerte für die LA-Kundendaten zu standardisieren.

Teil II: Inbetriebnahme des Informatica Developer

Dieser Teil enthält die folgenden Kapitel:

- [Lektion 1. Einrichten von Informatica Developer, 49](#)
- [Lektion 2: Importieren von physischen Datenobjekten., 53](#)
- [Lektion 3. Ausführen eines Profils auf Quelldaten, 62](#)
- [Lektion 4. Parsen von Daten, 67](#)
- [Lektion 5. Standardisieren von Daten , 74](#)
- [Lektion 6. Validieren von Adressdaten, 80](#)

KAPITEL 10

Lektion 1. Einrichten von Informatica Developer

Dieses Kapitel umfasst die folgenden Themen:

- [Einrichten von Informatica Developer - Übersicht, 49](#)
- [Aufgabe 1 Informatica Developer starten, 50](#)
- [Aufgabe 2 Eine Domäne hinzufügen, 50](#)
- [Aufgabe 3 Ein Model Repository hinzufügen, 51](#)
- [Aufgabe 4 Ein Projekt erstellen, 51](#)
- [Aufgabe 5 Einen Ordner erstellen, 51](#)
- [Aufgabe 6 Standard-Data Integration Service auswählen., 52](#)
- [Einrichten von Informatica Developer - Zusammenfassung, 52](#)

Einrichten von Informatica Developer - Übersicht

Bevor Sie mit den Lektionen in diesem Tutorial beginnen können, müssen Sie das Developer Tool starten und einrichten. Fügen Sie eine Domäne hinzu, um das Developer Tool einzurichten. Sie fügen ein Modellrepository hinzu, das sich in der Domäne befindet, und Sie erstellen ein Projekt und einen Ordner, um Ihre Arbeit zu speichern. Außerdem wählen Sie einen Standard-Datenintegrationsdienst.

Die Informatica-Domäne ist eine Sammlung von Knoten und Diensten, die die Informatica-Umgebung definieren. Zu den Diensten in der Domäne gehören der Modellrepository-Dienst und der Datenintegrationsdienst.

Der Modellrepository-Dienst verwaltet das Modellrepository. Das Modellrepository ist eine relationale Datenbank, die die Metadaten für Projekte speichert, die Sie im Developer Tool erstellen. In einem Projekt werden Objekte gespeichert, die Sie im Developer Tool erstellen. Ein Projekt kann auch Ordner enthalten, in denen verwandte Objekte gespeichert werden, wie Objekte, die Teile der gleichen Geschäftsumgebung sind.

Der Datenintegrationsdienst führt Datenintegrationsaufgaben im Developer Tool aus.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

- Developer Tool starten und zur Developer Tool-Workbench gehen.
- Hinzufügen einer Domäne im Developer Tool.

- Hinzufügen eines Modellrepository, damit Sie ein Projekt erstellen können.
- Erstellen eines Projekts, um die Objekte zu speichern, die Sie im Developer Tool erstellen.
- Erstellen eines Ordners im Projekt, in dem Sie verwandte Objekte speichern können.
- Auswählen eines Standard-Datenintegrationsdiensts zur Durchführung von Datenintegrationsaufgaben.

Voraussetzungen

Überprüfen Sie die folgenden Voraussetzungen, bevor Sie mit dieser Lektion beginnen:

- Sie haben das Developer Tool installiert.
- Sie haben einen Domänennamen, einen Hostnamen und eine Portnummer, um sich mit einer Domäne zu verbinden. Sie erhalten diese Informationen von einem Domänenadministrator.
- Ein Domänenadministrator hat im Administrator Tool einen Modellrepository-Dienst konfiguriert.
- Sie haben einen Benutzernamen und ein Passwort, um auf den Modellrepository-Dienst zuzugreifen. Sie können diese Informationen von einem Domänenadministrator erhalten.
- Ein Domänenadministrator hat einen Datenintegrationsdienst konfiguriert.
- Der Datenintegrationsdienst wird ausgeführt.

Zeitplan

Rechnen Sie mit 5 bis 10 Minuten, um die Aufgaben in dieser Lektion abzuschließen.

Aufgabe 1 Informatica Developer starten

Starten Sie das Entwicklertool, um mit dem Tutorial zu beginnen.

1. Starten Sie das Entwicklertool.
Die **Begrüßungsseite** des Entwicklertools wird eingeblendet.
2. Klicken Sie auf **Workbench**.
Die Workbench des Entwicklertools wird eingeblendet.

Aufgabe 2 Eine Domäne hinzufügen

In dieser Aufgabe fügen Sie im Entwicklertool eine Domäne hinzu, um auf ein Model Repository zuzugreifen.

1. Klicken Sie auf **Fenster > Einstellungen**.
Das Dialogfeld **Einstellungen** wird eingeblendet
2. Wählen Sie **Informatica > Domänen**.
3. Klicken Sie auf **Hinzufügen**.
Das Dialogfeld **Neue Domäne** wird eingeblendet
4. Geben Sie den Domänennamen, den Hostnamen und die Portnummer ein.
5. Klicken Sie auf **Fertigstellen**.
6. Klicken Sie auf **OK**.

Aufgabe 3 Ein Model Repository hinzufügen

In dieser Aufgabe fügen Sie ein Modell-Repository zum Speichern von Projekten und Ordnern hinzu.

1. Klicken Sie auf **Datei > Mit Repository verbinden**.
Das Dialogfeld **Mit Repository verbinden** wird eingeblendet.
2. Klicken Sie auf **Durchsuchen**, um einen Model Repository Service auszuwählen.
3. Klicken Sie auf **OK**.
4. Klicken Sie auf **Weiter**.
5. Geben Sie Ihren Benutzernamen und Ihr Passwort ein.
6. Wählen Sie einen Namespace aus.
7. Klicken Sie auf **Fertigstellen**.
Das Modell-Repository wird in der Ansicht **Object Explorer** angezeigt.

Aufgabe 4 Ein Projekt erstellen

In dieser Aufgabe erstellen Sie ein Projekt, um Objekte zu speichern, die Sie im Entwicklertool erstellt haben. Sie können ein Projekt für alle Tutorials in dieser Anleitung erstellen.

1. Wählen Sie in der Ansicht **Object Explorer** einen Model Repository Service.
2. Klicken Sie auf **Datei > Neu > Projekt**.
Das Dialogfeld **Neues Projekt** wird eingeblendet.
3. Geben Sie Ihren Namen mit vorangestelltem "Tutorial_" als Namen des Projekts ein.
4. Klicken Sie auf **Fertigstellen**.
Das Projekt wird unter dem Model Repository Service in der Ansicht **Object Explorer** angezeigt.

Aufgabe 5 Einen Ordner erstellen

In dieser Aufgabe erstellen Sie einen Ordner, um verwandte Objekte zu speichern. Sie können einen Ordner für alle Tutorials in dieser Anleitung erstellen.

1. Wählen Sie in der Ansicht **Object Explorer** das Projekt, dem Sie die Ordner hinzufügen möchten.
2. Klicken Sie auf **Datei > Neu > Ordner**.
3. Geben Sie einen Namen für den Ordner ein.
4. Klicken Sie auf **Fertigstellen**.
Das Entwicklertool fügt den Ordner unter dem Projekt in der Ansicht **Object Explorer** ein. Erweitern Sie das Projekt, um den Ordner zu sehen.

Aufgabe 6 Standard-Data Integration Service auswählen.

In dieser Aufgabe wählen Sie einen Standard-Data Integration Service aus, damit Sie Mappings ausführen und die Vorschau von Daten anzeigen können.

1. Klicken Sie auf **Fenster > Einstellungen**.
Das Dialogfeld **Einstellungen** wird eingeblendet
2. Wählen Sie **Informatica > Data Integration Services**.
3. Erweitern Sie die Domäne.
4. Wählen Sie einen Data Integration Service.
5. Klicken Sie auf **Als Standard definieren**.
6. Klicken Sie auf **OK**.

Einrichten von Informatica Developer - Zusammenfassung

In dieser Lektion haben Sie erfahren, dass der Model Repository Service und der Data Integration Service zu der Informatica-Domäne gehören. Der Model Repository Service verwaltet das Model Repository. Ein Model Repository enthält Projekte und Ordner. Der Data Integration Service führt Datenintegrationsaufgaben aus.

Sie haben das Entwicklertool gestartet und eingerichtet. Sie haben im Entwicklertool eine Domäne und ein Model Repository hinzugefügt und ein Projekt sowie einen Ordner erstellt. Sie haben außerdem einen Standard-Data Integration Service ausgewählt.

Sie können nun das Entwicklertool verwenden, um weitere Lektionen in diesem Tutorial durchzuarbeiten.

KAPITEL 11

Lektion 2: Importieren von physischen Datenobjekten.

Dieses Kapitel umfasst die folgenden Themen:

- [Importieren von physischen Datenobjekten - Übersicht, 53](#)
- [Aufgabe 1 Importieren des Flatfile-Datenobjekts Boston_Customers, 54](#)
- [Aufgabe 2 Import des Einfachdatei-Datenobjekts LA_Customers, 60](#)
- [Aufgabe 3. Importieren des Flatfile-Datenobjekts All_Customers, 61](#)
- [Importieren von physischen Datenobjekten - Zusammenfassung, 61](#)

Importieren von physischen Datenobjekten - Übersicht

Ein physisches Datenobjekt ist eine Darstellung von Daten, die auf einer Einfachdatei oder einer Tabelle einer relationalen Datenbank basieren. Sie können eine Einfachdatei oder Tabelle einer relationalen Datenbank als physisches Datenobjekt importieren, um es als Quelle oder Ziel in einem Mapping einzusetzen.

Story

HypoStores Corporation speichert Kundendaten aus dem Büro in Los Angeles und dem Büro in Boston in Einfachdateien. Sie möchten mit diesen Kundendaten im Developer Tool arbeiten. Dazu müssen Sie jede Einfachdatei als physisches Datenobjekt importieren.

Ziele

In dieser Lektion importieren Sie Einfachdateien als physische Datenobjekte. Sie können das Quelldateiverzeichnis auch so festlegen, dass der Datenintegrationsdienst die Quelldaten aus dem richtigen Verzeichnis lesen kann.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben Lektion 1 in diesem Tutorial abgeschlossen.

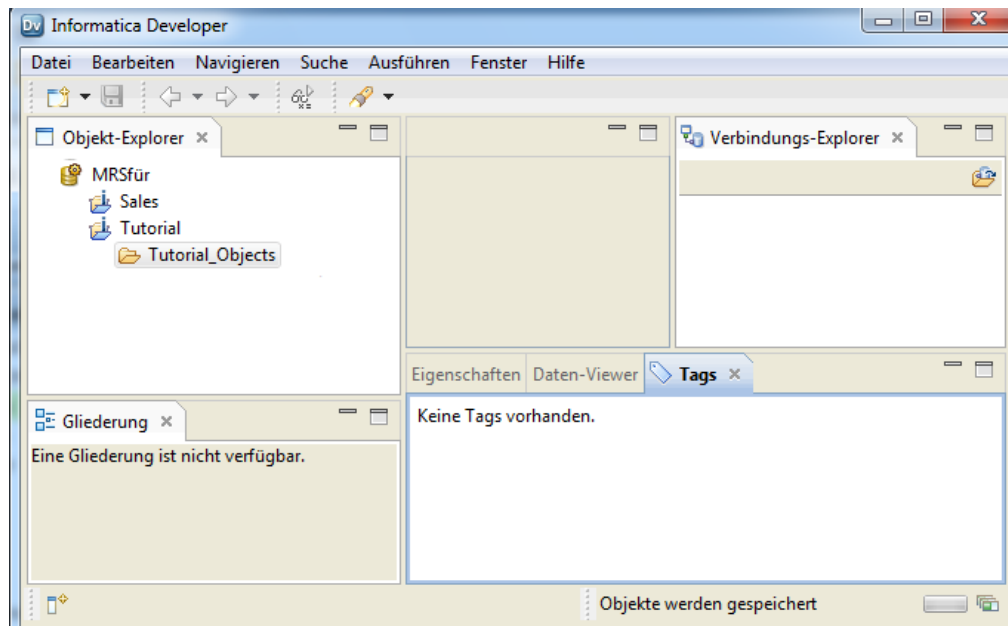
Zeitplan

Rechnen Sie mit 10 bis 15 Minuten, um die Aufgaben in dieser Lektion durchzuführen.

Aufgabe 1 Importieren des Flatfile-Datenobjekts Boston_Customers

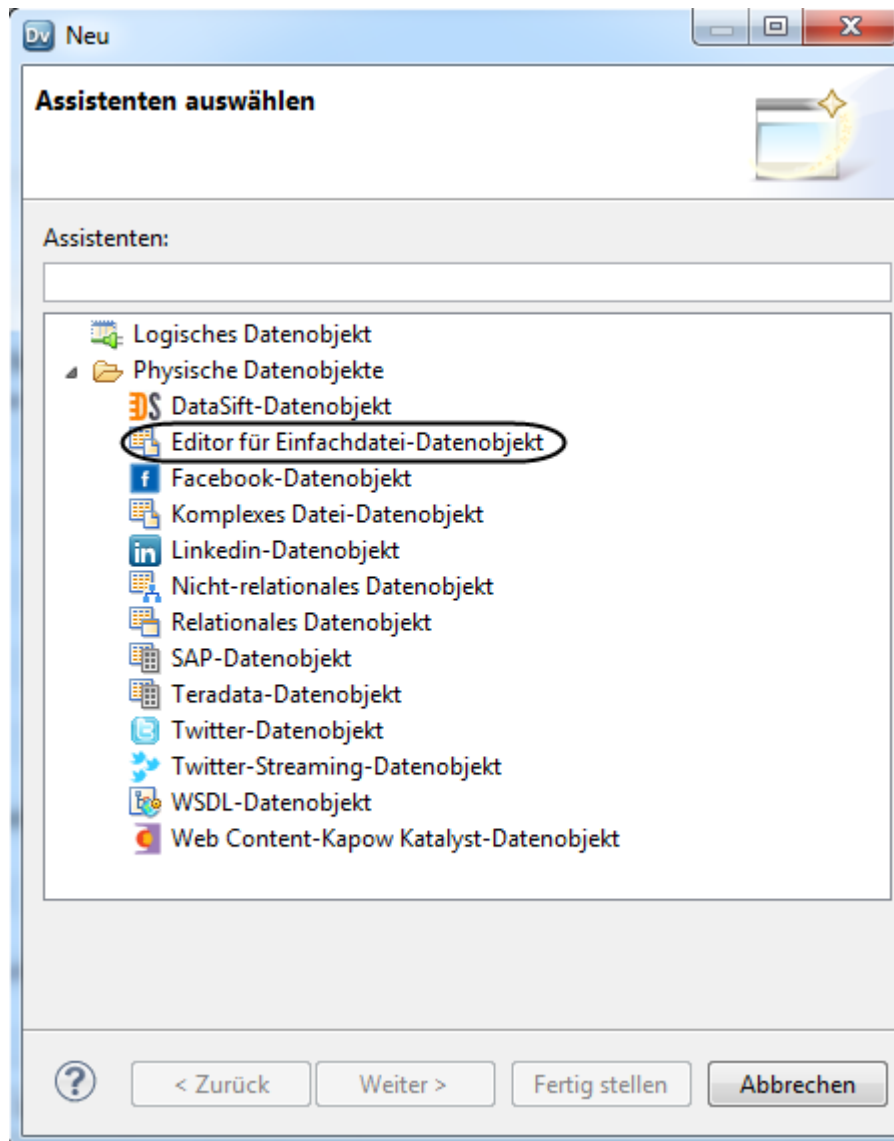
In dieser Aufgabe importieren Sie ein physisches Datenobjekt aus einer Datei, die Kundendaten aus dem Büro in Boston enthält.

1. Wählen Sie in der Ansicht **Objekt-Explorer** den Ordner Tutorial_Objects.



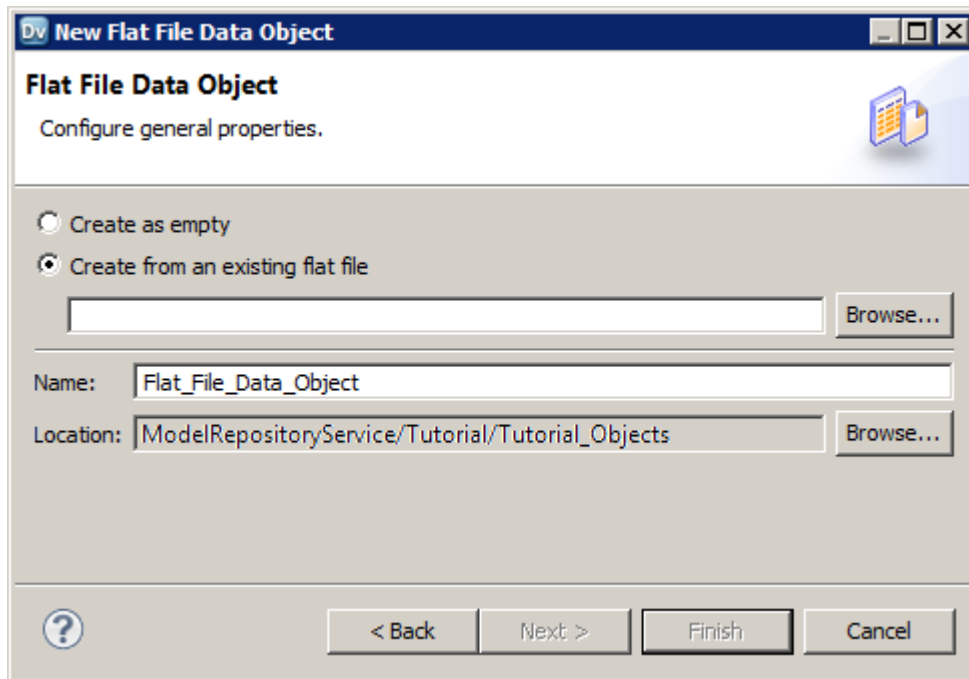
2. Klicken Sie mit der rechten Maustaste auf den Ordner Tutorial_Objects, und wählen Sie **Neu > Datenobjekt**.

Das Dialogfeld **Neu** wird eingeblendet.



3. Wählen Sie **Physische Datenobjekte** > **Einfachdatei-Datenobjekt** aus und klicken Sie auf **Weiter**.

Das Dialogfeld **Neues Einfachdatei-Datenobjekt** wird eingeblendet.



4. Wählen Sie **Aus vorhandener Einfachdatei erstellen**.
5. Klicken Sie auf **Durchsuchen** und navigieren Sie zu Boston_Customers.csv im folgenden Verzeichnis auf dem Computer des Developer-Tools: `<Informatica installation directory>\clients\DeveloperClient\Tutorials`
6. Klicken Sie auf **Öffnen**.
Der Assistent benennt das Datenobjekt "Boston_Customers".
7. Klicken Sie auf **Weiter**.
8. Stellen Sie sicher, dass die Codepage auf **MS Windows Latin 1 (ANSI)**, eine **Obermenge von Latin 1** und das Format auf **Delimitiert** festgelegt sind.

Das Dialogfeld **Neues Einfachdatei-Datenobjekt** zeigt die Standard-Codepage, das Format und eine Vorschau der Einfachdatei.

Neues Einfachdatei-Datenobjekt

Einfachdatei-Datenobjekt
Konfigurieren Sie und Codepage und Format.

Codepage: MS Windows Latin 1 (ANSI), superset of Latin1

Format

☒ Begrenzt (durch Delimiter getrennte Felder)

☐ Feste Breite (in Spalten ausgerichtete Felder)

Max. Anzahl Zeilen in der Vorschau: 500

	Feld1	Feld2	Feld3	
1	CustomerID	Customer Region	CustomerTier	Las
2	10110102	Boston	Bronze	Bro
3	10110105	Boston		An
4	10110106	Boston	Emerald	Bo
5	10110107	Boston	Ruby	Ch
6	10110109	Boston	Gold	Ac

? < Zurück Weiter > Fertig stellen Abbrechen

9. Klicken Sie auf **Weiter**.
10. Wählen Sie **Spaltennamen aus erster Zeile importieren**.

Das Dialogfeld **Neues Einfachdatei-Datenobjekt** zeigt die Spaltennamen in der Vorschau der Einfachdateidaten.

Einfachdatei-Datenobjekt
Konfigurieren Sie die Eigenschaften des durch Delimiter getrennten Formats.

Delimiter
☐ Tabulator ☐ Semikolon ☒ Komma
☐ Leerzeichen ☐ Andere:

Textqualifizierer
☒ Keine Anführungszeichen ☐ Einfache Anführungszeichen ☐ Doppelte Anführungszeichen

Vorschau-Optionen
☒ Spaltennamen aus erster Zeile importieren Import beginnen in Zeile: 1
Zeilen-Delimiter: \012 LF (\n) ☐ Aufeinander folgende Delimiter wie eins behandeln
Escape-Zeichen: ☐ Escape-Zeichen in Daten beibehalten

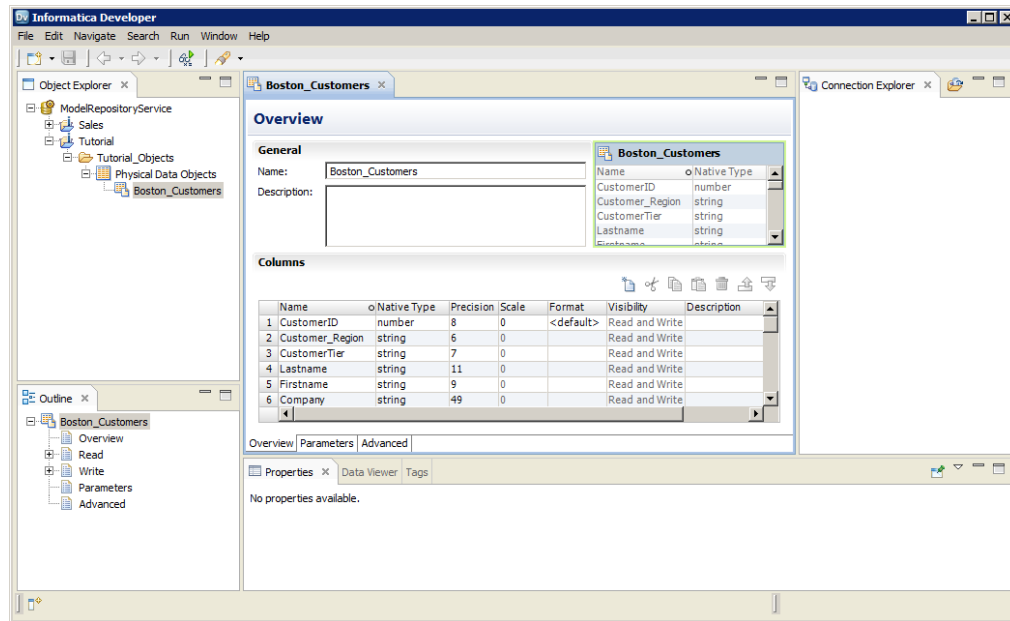
Max. Anzahl Zeilen in der Vorschau: 500

	CustomerID	Customer_Region	CustomerTier	Lastname
1	10110102	Boston	Bronze	Brosseau
2	10110105	Boston		Anderson
3	10110106	Boston	Emerald	Boonstra
4	10110107	Boston	Ruby	Chan

1. Die Option **Spaltennamen aus erster Zeile importieren**
2. Spaltennamen

11. Klicken Sie auf **Fertig stellen**.

Das physische Datenobjekt Boston_Customers wird unter dem Ordner der physischen Datenobjekte im Ordner Tutorial_Objects angezeigt. Die Ansicht **Übersicht** zeigt den Dateinhalt und wird im Editor geöffnet.



12. Klicken Sie auf die Ansicht **Erweitert**.

Die Ansicht **Erweitert** zeigt die Eigenschaften des physischen Datenobjekts an.

13. Führen Sie in der Ansicht **Erweitert** einen Bildlauf zum Abschnitt **Laufzeit: Lesen** durch.
14. Setzen Sie im Abschnitt **Laufzeit: Lesen** das **Quelldateiverzeichnis** auf das folgende Verzeichnis des Computers, auf dem der Datenintegrationsdienst läuft: <Informatica-Installationsverzeichnis> \server\Tutorials

Der Datenintegrationsdienst sucht die Quelldatei im Serververzeichnis auf dem Computer, auf dem der Datenintegrationsdienst ausgeführt wird. Die Server-Installation enthält eine Kopie der Tutorialdateien. Der Datenintegrationsdienst kann Dateien erst dann vom Client-Installationsverzeichnis lesen, wenn die Zugriffsberechtigungen für die Quelldatei und das Verzeichnis geändert wurden. Die folgende Abbildung zeigt ein Beispiel-Quelldateiverzeichnis:

Advanced	
Name	Value
Runtime : Read	
Input type	File
Source type	Direct
Source file name	Boston_Customers.csv
Source file directory	\\MyMachine\Informatica\10.0\server\Tutorials

Hinweis: Der Computer des Developer-Tools muss Zugriff auf das Quelldateiverzeichnis auf dem Computer haben, auf dem der Datenintegrationsdienst ausgeführt wird. Wenn das Developer-Tool nicht auf das Quelldateiverzeichnis zugreifen kann, kann das Developer-Tool keine Vorschau der Daten in der Quelldatei aufrufen oder Mappings ausführen, die Zugriff auf Daten in der Quelldatei haben. Wenn Sie mehrere Datenintegrationsdienste ausführen, ist für jeden Datenintegrationsdienst ein separates Quelldateiverzeichnis verfügbar.

15. Klicken Sie auf die Ansicht **Daten-Viewer**.
16. Klicken Sie in der Ansicht **Daten-Viewer** auf **Ausführen**.
Der Datenintegrationsdienst liest die Daten aus der Datei Boston_Customers und zeigt die Ergebnisse im Ausgabefenster.
17. Klicken Sie auf **Datei > Speichern**, um das physische Datenobjekt Boston_Customers zu speichern.

Aufgabe 2 Import des Einfachdatei-Datenobjekts LA_Customers

In dieser Aufgabe importieren Sie ein physisches Datenobjekt aus einer Einfachdatei, die Kundendaten aus dem Büro in Los Angeles enthält.

1. Wählen Sie in der Ansicht **Object Explorer** das Tutorialprojekt aus.
2. Klicken Sie auf **Datei > Neu > Datenobjekt**.
Das Dialogfeld **Neu** wird eingeblendet.
3. Wählen Sie **Physische Datenobjekte > Einfachdatei-Datenobjekt**, und klicken Sie auf **Weiter**.
Das Dialogfeld **Neues Einfachdatei-Datenobjekt** wird eingeblendet.
4. Wählen Sie **Aus vorhandener Einfachdatei erstellen**.
5. Klicken Sie auf **Durchsuchen**, und gehen Sie im Verzeichnis `<Informatica Installation Directory>\clients\DeveloperClient\Tutorials` zu `LA_Customers.csv`.
6. Klicken Sie auf **Öffnen**.
Der Assistent benennt das Datenobjekt LA_Customers.
7. Klicken Sie auf **Weiter**.
8. Vergewissern Sie sich, dass die Codepage MS Windows Latin 1 (ANSI) ist, eine Obermenge von Latin 1.
9. Vergewissern Sie sich, dass das Format "delimited" ist.
10. Klicken Sie auf **Weiter**.
11. Vergewissern Sie sich, dass das Komma als Delimiter festgelegt ist.
12. Wählen Sie **Spaltennamen aus erster Zeile importieren**.
13. Klicken Sie auf **Fertigstellen**.
Das physische Datenobjekt LA_Customers erscheint im Tutorialprojekt unter Physische Datenobjekte.
14. Klicken Sie auf die Ansicht **Lesen**, und wählen Sie die Ausgabeumwandlung.
15. Klicken Sie auf die Registerkarte **Laufzeit** in der Ansicht **Eigenschaften**.
16. Richten Sie das Quelldateiverzeichnis auf das folgende Verzeichnis des Rechners ein, auf dem der Data Integration Service läuft: `<Informatica Installation Directory>\server\Tutorials`.
17. Klicken Sie auf **Datei > Speichern**.

Aufgabe 3. Importieren des Flatfile-Datenobjekts All_Customers

In dieser Aufgabe importieren Sie ein physisches Datenobjekt aus einer Einfachdatei, die Kunden-Auftragsdaten aus den Büros in Los Angeles und Bosten kombiniert.

1. Wählen Sie in der Ansicht **Object Explorer** das Tutorialprojekt aus.
2. Klicken Sie auf **Datei > Neu > Datenobjekt**.
Das Dialogfeld **Neu** wird angezeigt.
3. Wählen Sie **Physische Datenobjekte > Einfachdatei-Datenobjekt** aus und klicken Sie auf **Weiter**.
Das Dialogfeld **Neue Einfachdatei-Datenquelle** wird eingeblendet.
4. Wählen Sie **Aus vorhandener Einfachdatei erstellen**.
5. Klicken Sie auf **Durchsuchen**, und gehen Sie im Verzeichnis `<Informatica-Installationsverzeichnis>\clients\DeveloperClient\Tutorials` zu `All_Customers.csv`.
6. Klicken Sie auf **Öffnen**.
Der Assistent benennt das Datenobjekt "All_Customers".
7. Klicken Sie auf **Weiter**.
8. Vergewissern Sie sich, dass die Codepage MS Windows Latin 1 (ANSI) ist, eine Obermenge von Latin 1.
9. Vergewissern Sie sich, dass das Format "delimited" ist.
10. Klicken Sie auf **Weiter**.
11. Vergewissern Sie sich, dass das Komma als Delimiter festgelegt ist.
12. Wählen Sie **Spaltennamen aus erster Zeile importieren**.
13. Klicken Sie auf **Fertigstellen**.
Das physische Datenobjekt All_Customers erscheint im Tutorialprojekt unter **Physische Datenobjekte**.
14. Klicken Sie auf die Ansicht **Lesen**, und wählen Sie die Ausgabeumwandlung.
15. Klicken Sie auf die Registerkarte **Laufzeit** in der Ansicht **Eigenschaften**.
16. Richten Sie das Quelldateiverzeichnis auf das folgende Verzeichnis des Rechners ein, auf dem der Datenintegrationsdienst läuft: `<Informatica Installation Directory>\server\Tutorials`.
17. Klicken Sie auf **Datei > Speichern**.

Importieren von physischen Datenobjekten - Zusammenfassung

In dieser Lektion haben Sie erfahren, dass physische Datenobjekte Darstellungen von Daten sind, die auf einer Einfachdatei oder einer Tabelle einer relationalen Datenbank basieren.

Sie haben physische Datenobjekte aus Einfachdateien erstellt. Sie können das Quelldateiverzeichnis auch so festlegen, dass der Data Integration Service die Quelldaten aus dem richtigen Verzeichnis lesen kann.

Sie verwenden die Datenobjekte als Mapping-Quellen in den Datenqualitätslektionen.

KAPITEL 12

Lektion 3. Ausführen eines Profils auf Quelldaten

Dieses Kapitel umfasst die folgenden Themen:

- [Profiling von Daten - Übersicht, 62](#)
- [Aufgabe 1. Durchführen einer Join-Analyse für zwei Datenquellen, 63](#)
- [Aufgabe 2. Anzeigen der Ergebnisse einer Join-Analyse, 64](#)
- [Aufgabe 3. Ausführen eines Profils für eine Datenquelle, 65](#)
- [Aufgabe 4. Anzeigen der Spalten-Profiling-Ergebnisse, 65](#)
- [Profiling von Daten - Zusammenfassung, 66](#)

Profiling von Daten - Übersicht

Ein Profil ist ein Satz von Metadaten, die den Inhalt und die Struktur eines Datensatzes beschreiben.

Profiling und Datenerkennung stellen oftmals den ersten Schritt in einem Projekt dar. Sie können ein Profil ausführen, um die Struktur der Daten zu bewerten und sich zu vergewissern, dass Datenspalten mit den Typen von Informationen befüllt sind, die Sie erwarten. Wenn ein Profil Probleme in den Daten aufzeigt, können Sie Schritte in Ihrem Projekt festlegen, um diese Probleme zu beheben. Wenn ein Profil beispielsweise aufzeigt, dass eine Spalte Werte enthält, die über der erwarteten Länge liegen, können Sie Datenqualitätsprozesse entwerfen, um die problematischen Werte zu entfernen oder zu entschärfen.

Ein Profil, das die Datenqualität von ausgewählten Spalten analysiert, wird als Spaltenprofil bezeichnet.

Hinweis: Sie können auch das Developer Tool verwenden, um Beziehungen von Primärschlüssel, Fremdschlüssel und funktionaler Abhängigkeit zu erkennen und Join-Bedingungen an Datenspalten zu analysieren.

Ein Spaltenprofil liefert die folgenden Fakten über Daten:

- Die Anzahl der eindeutigen Werte und Nullwerte in jeder Spalte, angegeben als Zahl und als Prozentwert.
- Das Muster der Daten in jeder Spalte und die Häufigkeit, mit der diese Werte vorkommen.
- Statistiken zu den Spaltenwerten, wie Maximal- und Minimallänge der Werte und erster und letzter Wert in jeder Spalte.
- Bei Join-Analyse-Profilen wird der Überlappungsgrad zwischen zwei Datenspalten als Venn-Diagramm und als Prozentwert angezeigt. Verwenden Sie Join-Analyse-Profile, um mögliche Profile mit Spalten-Joinbedingungen zu identifizieren.

Sie können in jedem Stadium eines Projekts ein Spaltenprofil ausführen und überprüfen, ob die Änderungen an den Daten Ihren Projektzielen entsprechen. Sie können ein Spaltenprofil an einer Umwandlung in einem Mapping ausführen, um die Wirkung anzuzeigen, die die Umwandlung auf die Daten haben wird.

Story

HypoStores möchte sich vergewissern, dass die Kundendaten frei von Fehlern, Inkonsistenzen und doppelten Informationen sind. Bevor HypoStores die Prozesse entwickelt, um die Datenqualitätsziele zu erreichen, muss die Qualität ihrer Quelldatendateien gemessen werden und bestätigt werden, dass die Daten verarbeitet werden können.

Ziele

In dieser Lektion führen Sie folgende Aufgaben durch:

- Durchführen einer Join-Analyse an der Datenquelle `Boston_Customers` und der Datenquelle `LA_Customers`.
- Anzeigen der Ergebnisse der Join-Analyse, um zu bestimmen, ob die Daten aus den beiden Büros erfolgreich zusammengeführt werden können oder nicht.
- Ausführen eines Spaltenprofils an der Datenquelle `All_Customers`.
- Anzeigen der Ergebnisse des Spaltenprofils, um die Werte und Muster anzusehen, die in den Daten enthalten sind.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 und 2 in diesem Tutorial abgeschlossen.

Zeitbedarf

- Rechnen Sie mit 20 Minuten für diese Lektion.

Aufgabe 1. Durchführen einer Join-Analyse für zwei Datenquellen

In dieser Aufgabe führen Sie eine Join-Analyse der Datenquellen „Boston_Customers“ und „LA_Customers“ durch, um die Join-Bedingungen anzuzeigen.

1. Wählen Sie den Tutorial-Ordner aus und klicken Sie auf **Datei > Neu > Profil**.
2. Wählen Sie **Enterprise-Erkennungsprofil** aus.
3. Klicken Sie auf **Weiter**.
4. Geben Sie im Feld **Name** die Bezeichnung **Tutorial_Profile** ein.
5. Klicken Sie auf **Fertig stellen**.
Das Tutorial_Profile-Profil wird im Objektexplorer angezeigt.
6. Ziehen Sie die Datenquellen **Boston_Customers** und **LA_Customers** in den Editor auf der rechten Seite.
Tipp: Halten Sie die Shift-Taste gedrückt, um mehrere Datenobjekte auszuwählen.
7. Klicken Sie mit der rechten Maustaste auf einen Datenobjektnamen und wählen Sie **Join-Profil** aus.
Der Assistent **Neues Join-Profil** wird angezeigt.
8. Geben Sie im Feld **Name** die Bezeichnung **JoinAnalysis** ein.

9. Überprüfen Sie, ob Boston_Customers und LA_Customers als Datenobjekte angezeigt werden und klicken Sie auf **Weiter**.
10. Überprüfen Sie, ob die Spalte **CustomerID** in beiden Datenquellen ausgewählt ist. Scrollen Sie im Assistentenbereich nach unten, um die Spalten in beiden Datensätzen anzuzeigen. Klicken Sie auf **Weiter**.
11. Klicken Sie auf **Hinzufügen**, um Join-Bedingungen hinzuzufügen.
Das Dialogfeld **Join-Bedingung** wird angezeigt.
12. Klicken Sie im Abschnitt **Spalten** auf **Zeile hinzufügen**.
13. Doppelklicken Sie auf die erste Zeile in der linken Spalte und wählen Sie **CustomerID** aus.
14. Doppelklicken Sie auf die erste Zeile in der rechten Spalte und wählen Sie **CustomerID** aus.
15. Klicken Sie auf **OK** und dann auf **Fertigstellen**.
16. Wenn Sie vom Developer-Tool zum Speichern der Änderungen aufgefordert werden, klicken Sie auf **Ja**.
Das Developer-Tool führt das Profil aus.

Hinweis: Schließen Sie das Profil nicht. In der nächsten Aufgabe zeigen Sie die Ergebnisse des Profils an.

Aufgabe 2. Anzeigen der Ergebnisse einer Join-Analyse

In dieser Aufgabe zeigen Sie die Ergebnisse der Join-Analyse in der Ansicht der Join-Ergebnisse des Profils „JoinAnalysis“ an.

1. Klicken Sie auf die Registerkarte **JoinAnalysis** im Editor.
2. Klicken Sie im Abschnitt **Join-Ergebnis** auf die erste Zeile.
Im Abschnitt **Details** werden ein Venn-Diagramm und ein Schlüssel, der die Ergebnisse der Join-Analyse detailliert angibt, angezeigt.
3. Überprüfen Sie, ob die Spalte **Zeilen verbinden** für die Anzahl der Zeilen, die einen Join enthalten, Null enthält.
Dieser Wert gibt an, dass CustomerID-Felder keine Duplikate aufweisen. Sie können die beiden Datenquellen erfolgreich zusammenführen.
4. Um die CustomerID-Werte für das Datenobjekt LA_Customers anzuzeigen, doppelklicken Sie im Venn-Diagramm auf den Kreis mit der Bezeichnung **LA_Customers**.
Tipp: Doppelklicken Sie auf die Kreise im Venn-Diagramm, um die Datenzeilen anzuzeigen. In Fällen, in denen sich im Venn-Diagramm Kreise überschneiden, doppelklicken Sie auf den Schnittpunkt, um Datenwerte anzuzeigen, die in beiden Datensätzen enthalten sind.
Der Daten-Viewer zeigt die CustomerID-Werte an, die im Datenobjekt „LA_Customers“ enthalten sind.

Aufgabe 3. Ausführen eines Profils für eine Datenquelle

In dieser Aufgabe führen Sie ein Profil für die Datenquelle „All_Customers“ aus, um die Struktur und den Inhalt der Daten anzuzeigen.

1. Navigieren Sie in der Ansicht **Objekt-Explorer** zu den Datenobjekten in Ihrem Tutorial-Projekt.
2. Wählen Sie die Datenquelle **All_Customers** aus.
3. Klicken Sie auf **Datei > Neu > Profil**.
Das Dialogfeld **Neu** wird angezeigt.
4. Wählen Sie **Profil** aus.
5. Klicken Sie auf **Weiter**.
6. Geben Sie im Feld **Name** die Bezeichnung **All_Customers** ein.
7. Klicken Sie auf **Fertig stellen**.

Das Profil „All_Customers“ wird im Editor geöffnet und das Profil wird ausgeführt.

Aufgabe 4. Anzeigen der Spalten-Profiling-Ergebnisse

In dieser Aufgabe zeigen Sie die Ergebnisse des Spalten-Profiling für das Datenobjekt „All_Customers“ an und untersuchen die Werte und die in den Daten enthaltenen Muster.

1. Klicken Sie auf **Fenster > Ansicht anzeigen > Fortschritt**, um den Fortschritt des Profils „All_Customers“ anzuzeigen.
Die Fortschritt-Ansicht wird geöffnet.
2. Wenn die Ansicht „Fortschritt“ angibt, dass die Ausführung des Profils „All_Customers“ abgeschlossen ist, klicken Sie im Editor auf die Ansicht **Ergebnisse**.
3. Klicken Sie im Abschnitt **Spalten-Profiling** auf die Spalte **CustomerTier**.
Im Abschnitt **Details** werden alle Werte, die in der Spalte „CustomerTier“ enthalten sind, sowie Informationen darüber angezeigt, wie häufig der Wert in dem Datensatz vorkommt.
4. Doppelklicken Sie im Abschnitt **Details** auf **Ruby**.
Der Daten-Viewer wird ausgeführt und zeigt die Datensätze, in denen die Spalte „CustomerTier“ den Wert **Ruby** enthält.
5. Klicken Sie im Abschnitt **Spalten-Profiling** auf die Spalte **OrderAmount**.
6. Klicken Sie im Abschnitt **Details** auf die Liste **Anzeigen** und wählen Sie **Muster** aus.
Der Abschnitt **Details** zeigt die Muster an, die sich in der Spalte „OrderAmount“ befinden. Die Zeichenfolge **9((5)** in der Spalte „Muster“ bezieht sich auf Datensätze, die fünfstellige Bestellmengen enthalten. Die Zeichenfolge **9((4** bezieht sich auf Datensätze, die vierstellige Mengen enthalten.
7. Doppelklicken Sie in der Spalte **Muster** auf die Zeichenfolge **9(4)**.
Der Daten-Viewer wird ausgeführt und zeigt die Datensätze an, in denen die Spalte „OrderAmount“ eine vierstellige Bestellmenge enthält.

8. Klicken Sie im Abschnitt **Details** auf die Liste **Anzeigen** und wählen Sie **Statistiken** aus.

Im Abschnitt **Details** werden Statistiken für die Spalte „OrderAmount“ einschließlich dem Durchschnittswert, der Standardabweichung, maximalen und Mindestlängen, den fünf häufigsten Werten und den fünf seltensten Werten angezeigt.

Profiling von Daten - Zusammenfassung

In dieser Lektion haben Sie erfahren, dass ein Profil Informationen über den Inhalt und die Struktur der Daten liefert.

Sie haben gelernt, dass Sie mit zwei Datenobjekten eine Join-Analyse durchführen können und den Überlappungsgrad zwischen den Datenobjekten anzeigen können. Sie haben auch erfahren, dass Sie ein Spaltenprofil an einem Datenobjekt ausführen können und Werte, Muster und Statistiken anzeigen können, die jede Spalte in dem Datenobjekt betreffen.

Sie haben das Profil `JoinAnalysis` erstellt, um zu ermitteln, ob Daten aus dem Datenobjekt `Boston_Customers` mit den Daten im Datenobjekt `LA_Customers` zusammengeführt werden können. Sie haben die Ergebnisse dieses Profils angezeigt und ermittelt, dass alle Werte in der Spalte `CustomerID` eindeutig sind und die Datenobjekte erfolgreich zusammengeführt werden können.

Sie haben das Profil `All_Customers` erstellt und ein Spaltenprofil mit dem Datenobjekt `All_Customers` ausgeführt. Sie haben die Ergebnisse dieses Profils angezeigt, um Werte, Muster und Statistiken für die Spalten im Datenobjekt `All_Customers` anzuzeigen. Sie haben schließlich den Daten-Viewer ausgeführt, um Zeilen anzuzeigen, die die von Ihnen ausgewählten Werte und Muster enthalten, wodurch Sie die Qualität der Daten prüfen konnten.

KAPITEL 13

Lektion 4. Parsen von Daten

Dieses Kapitel umfasst die folgenden Themen:

- [Parsen von Daten - Übersicht, 67](#)
- [Aufgabe 1. Erstellen eines Target-Datenobjekts, 68](#)
- [Aufgabe 2. Erstellen eines Mappings zum Parsen von Daten, 70](#)
- [Aufgabe 3. Ausführen eines Profils für die Parser-Umwandlung, 72](#)
- [Aufgabe 4. Ausführen des Mappings, 72](#)
- [Aufgabe 5. Anzeigen der Mapping-Ausgabe, 72](#)
- [Parsen von Daten - Zusammenfassung, 73](#)

Parsen von Daten - Übersicht

Sie parsen Daten, um ein oder mehrere Datenelemente in einem Eingabefeld zu kennzeichnen und jedes Datenelement in ein anderes Ausgabefeld zu schreiben.

Das Parsen ermöglicht Ihnen, die Informationen in den einzelnen Spalten besser zu kontrollieren. Nehmen Sie beispielsweise ein Datenfeld mit dem vollen Namen einer Person, `Bob Smith`. Anhand der Parser-Umwandlung können Sie den vollständigen Namen in separate Datenspalten für den Vornamen und den Nachnamen aufteilen. Nachdem Sie die Daten in neue Spalten geparkt haben, können Sie für jede Spalte benutzerspezifische Datenqualitäts-Operationen erstellen.

Sie können die Parser-Umwandlung auf die Verwendung von *Tokensätzen* zum Parsen von Datenspalten in Komponenten-Strings konfigurieren. Ein Tokensatz kennzeichnet Datenelemente, z. B. Wörter, Postleitzahlen, Telefonnummern und Sozialversicherungsnummern.

Mit der Parser-Umwandlung können Sie ebenfalls Daten parsen, die mit Referenztabelleinträgen oder von Ihnen eingegebenen benutzerspezifischen regulären Expressionen übereinstimmen.

Story

Bei HypoStores muss das Format der Kundendatendateien der Niederlassung in Los Angeles mit dem Format der Datendateien der Bostoner Niederlassung übereinstimmen. Die Kundendaten der Niederlassung Los Angeles speichern den Kundennamen in einer Spalte `FullName`, während die Kundendaten der Bostoner Niederlassung den Kundennamen in separaten Spalten für den `FirstName` und `LastName` speichern. Bei HypoStores müssen die Daten der Spalten Los Angeles `FullName` in Vornamen und Nachnamen geparkt werden, damit das Format der Daten für Los Angeles mit dem Format der Bostoner Daten übereinstimmt.

Zielsetzungen

In dieser Lektion bearbeiten Sie folgende Aufgaben:

- Erstellen und Konfigurieren eines Datenobjekts `LA_Customers_tgt`, damit es geparte Daten enthält.
- Erstellen eines Mapping zum Parsen der Spalte `FullName` in separate Spalten für `FirstName` und `LastName`.
- Hinzufügen des Datenobjekts `LA_Customers` zum Mapping, um die Verbindung zu den Quelldaten herzustellen.
- Hinzufügen des Datenobjekts `LA_Customers_tgt` zum Mapping, um ein Target-Datenobjekt zu erstellen.
- Hinzufügen einer Parser-Umwandlung zum Mapping und Konfigurieren dieser Umwandlung, um einen Tokensatz zum Parsen vollständiger Namen in Vor- und Nachnamen zu verwenden.
- Ausführen eines Profils für die Parser-Umwandlung, um die Daten vor dem Generieren der Target-Datenquelle zu überprüfen.
- Ausführen des Mapping, um geparte Namen zu generieren.
- Ausführen des Daten-Viewers zur Anzeigen der Mapping-Ausgabe.

Voraussetzungen

Überprüfen Sie die folgende Voraussetzung, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 und 2 in diesem Tutorial abgeschlossen.

Zeitaufwand

Planen Sie für die vollständige Durchführung der Aufgaben in dieser Lektion 20 Minuten ein.

Aufgabe 1. Erstellen eines Target-Datenobjekts

In dieser Aufgabe erstellen Sie ein Datenobjekt `LA_Customers_tgt`, in das Sie geparte Namen schreiben können.

Zum Erstellen eines Target-Datenobjekts gehen Sie folgendermaßen vor:

1. Erstellen Sie ein Datenobjekt `LA_Customers_tgt` basierend auf der Datei `LA_Customers.csv`.
2. Konfigurieren Sie die Lesen- und Schreiben-Optionen für das Datenobjekt einschließlich der Dateispeicherorte und Dateinamen.
3. Fügen Sie die Spalten `Firstname` und `Lastname` zum Datenobjekt `LA_Customers_tgt` hinzu.

Schritt 1. Erstellen eines Datenobjekts `LA_Customers_tgt`

In diesem Schritt erstellen Sie ein Datenobjekt `LA_Customers_tgt` basierend auf der Datei `LA_Customers.csv`.

1. Klicken Sie auf **Datei > Neu > Datenobjekt**.
Das Fenster **Neu** wird geöffnet.
2. Wählen Sie **Flatfile-Datenobjekt**, und klicken Sie auf **Weiter**.
3. Vergewissern Sie sich, dass **Aus einem vorhandenen Flatfile erstellen** ausgewählt wurde.
4. Klicken Sie auf **Durchsuchen**, und navigieren Sie zu `LA_Customers.csv` im folgenden Verzeichnis:
<Informatica-Installationsverzeichnis>\clients\DeveloperClient\Tutorials

5. Klicken Sie auf **Öffnen**.
6. Geben Sie im Feld Name `LA_Customers_tgt` ein.
7. Klicken Sie auf **Weiter**.
8. Klicken Sie auf **Weiter**.
9. Wählen Sie im Abschnitt **Vorschauoptionen Spaltennamen aus erster Zeile importieren**, und klicken Sie auf **Weiter**.
10. Klicken Sie auf **Fertigstellen**.
Das Datenobjekt `LA_Customers_tgt` erscheint im Editor.

Schritt 2. Konfigurieren der Lesen- und Schreiben-Optionen

In diesem Schritt konfigurieren Sie die Lesen- und Schreiben-Optionen für das Datenobjekt `LA_Customers_tgt` einschließlich Dateispeicherort und Dateinamen.

1. Vergewissern Sie sich, dass das Datenobjekt `LA_Customers_tgt` im Editor geöffnet ist.
2. Wählen Sie im Editor die Ansicht **Lesen**.
3. Klicken Sie auf **Fenster > Ansicht zeigen > Eigenschaften**.
4. Wählen Sie in der Ansicht **Eigenschaften** die Ansicht **Laufzeit**.
5. In der Spalte **Wert** doppelklicken Sie auf den Quelldateinamen und -typ `LA_Customers_tgt.csv`.
6. Doppelklicken Sie in der Spalte **Wert**, um das Quelldateiverzeichnis hervorzuheben.
7. Klicken Sie mit der rechten Maustaste auf den hervorgehobenen Namen, und wählen Sie **Kopieren** aus.
8. Wählen Sie im Editor die Ansicht **Schreiben**.
9. Wählen Sie in der Ansicht **Eigenschaften** die Ansicht **Laufzeit** aus.
10. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Ausgabedateiverzeichnis**.
11. Klicken Sie mit der rechten Maustaste, und wählen Sie **Einfügen**, um den Verzeichnisspeicherort einzufügen, den Sie aus der Ansicht **Lesen** kopiert haben.
12. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Kopfzeilenoptionen**, und wählen Sie **Ausgabefeldnamen**.
13. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Ausgabedateinamen** und den Typ `LA_Customers_tgt.csv`.
14. Zum Speichern des Datenobjekts klicken Sie auf **Datei > Speichern**.

Schritt 3. Hinzufügen von Spalten zum Datenobjekt

In diesem Schritt fügen Sie die Spalten `Vorname` und `Nachname` zum Datenobjekt `LA_Customers_tgt` hinzu.

1. Gehen Sie in der Ansicht Objekt-Explorer zu den Datenobjekten in Ihrem Tutorial-Projekt.
2. Doppelklicken Sie auf das Datenobjekt `LA_Customers_tgt`.
Das Datenobjekt `LA_Customers_tgt` wird im Editor geöffnet.
3. Überprüfen Sie, dass die Ansicht **Übersicht** ausgewählt wurde.
4. Wählen Sie die Spalte **FullName**, und klicken Sie die Schaltfläche **Neu** an, um eine Spalte hinzuzufügen.
Eine Spalte mit dem Namen `FullName1` wird angezeigt.
5. Benennen Sie die Spalte in `Firstname` um. Klicken Sie auf das Feld **Präzision**, und geben Sie "30" ein.
6. Wählen Sie die Spalte `Firstname`, und klicken Sie die Spalte **Neu** an, um eine Spalte hinzuzufügen.

Eine Spalte mit dem Namen `FirstName1` wird angezeigt.

- Benennen Sie die Spalte in `LastName` um. Klicken Sie auf das Feld **Präzision**, und geben Sie "30" ein.
- Klicken Sie auf **Datei > Speichern**, um das Datenobjekt zu speichern.

Aufgabe 2. Erstellen eines Mappings zum Parsen von Daten

In dieser Aufgabe erstellen Sie ein Mapping und konfigurieren es für die Arbeit mit Datenobjekten und einer Parser-Umwandlung.

Um ein Mapping zum Parsen von Daten anzulegen, gehen Sie bitte folgendermaßen vor:

- Erstellen Sie ein Mapping.
- Fügen Sie dem Mapping Quell- und Targetobjekte hinzu.
- Fügen Sie eine Parser-Umwandlung zum Mapping hinzu.
- Konfigurieren Sie die Parser-Umwandlung zum Parsen der Quellenspalte mit dem vollständigen Kundennamen in separate Targetspalten mit Vor- und Nachnamen.

Schritt 1. Erstellen eines Mappings

In diesem Schritt erstellen Sie das Mapping und geben ihm einen Namen.

- Wählen Sie in der Ansicht Objekt-Explorer das Tutorial-Projekt aus.
- Klicken Sie auf **Datei > Neu > Mapping**.

Das Fenster **Neues Mapping** wird geöffnet.

- Geben Sie in das Feld **Name** `ParserMapping` ein.
- Klicken Sie auf **Fertigstellen**.

Das Mapping wird im Editor geöffnet.

Schritt 2. Hinzufügen von Datenobjekten zum Mapping

In diesem Schritt fügen Sie das Datenobjekt `LA_Customers` und das Datenobjekt `LA_Customers_tgt` zum Mapping hinzu.

- Gehen Sie in der Ansicht Objekt-Explorer zu den Datenobjekten in Ihrem Tutorial-Projekt.
- Wählen Sie das Datenobjekt `LA_Customers` aus, und ziehen Sie es in den Editor.

Das Fenster **Physisches Datenobjekt zum Mapping hinzufügen** wird geöffnet.

- Vergewissern Sie sich, das **Lesen** ausgewählt ist, und klicken Sie auf **OK**.

Das Datenobjekt erscheint im Editor.

- Gehen Sie in der Ansicht Objekt-Explorer zu den Datenobjekten in Ihrem Tutorial-Projekt.
- Wählen Sie das Datenobjekt `LA_Customers_tgt` aus, und ziehen Sie es in den Editor.

Das Fenster **Physisches Datenobjekt zum Mapping hinzufügen** wird geöffnet.

- Wählen Sie **Schreiben**, und klicken Sie auf **OK**.

Das Datenobjekt erscheint im Editor.

- Wählen Sie die Ports `CustomerID`, `CustomerTier` und `FullName` im Datenobjekt `LA_Customers`. Ziehen Sie die Ports in den Port `CustomerID` im Datenobjekt `LA_Customers_tgt`.

Tipp: Halten Sie die Taste **Strg** gedrückt, um mehrere Ports auszuwählen.

Die Ports des Datenobjekts `LA_Customers` stellen die Verbindung mit entsprechenden Ports im Datenobjekt `LA_Customers_tgt` her.

Schritt 3. Hinzufügen einer Parser-Umwandlung zum Mapping

In diesem Schritt fügen Sie eine Parser-Umwandlung zum Mapping `ParserMapping` hinzu.

- Wählen Sie den Editor, der das Mapping `ParserMapping` enthält.
- Wählen Sie aus der Umwandlungspalette die Parser-Umwandlung aus.
- Klicken Sie auf den Editor.
Das Fenster **Neue Parser-Umwandlung** wird geöffnet.
- Vergewissern Sie sich, dass **Token Parser** ausgewählt wurde, und klicken Sie auf **Fertigstellen**.
Die Parser-Umwandlung erscheint im Editor.
- Wählen Sie den Port `FullName` im Datenobjekt `LA_Customers` aus, und ziehen Sie den Port in die Gruppe `Eingabe` der Parser-Umwandlung.

Der Port `FullName` erscheint in der Parser-Umwandlung und ist mit dem Port `FullName` im Datenobjekt verbunden.

Schritt 4. Konfigurieren der Parser-Umwandlung

In diesem Schritt konfigurieren Sie die Parser-Umwandlung auf Parsen der Spalte mit dem vollständigen Kundennamen in separate Spalten, die den Vornamen und den Nachnamen enthalten.

- Wählen Sie den Editor aus, der das Mapping `ParserMapping` enthält.
- Klicken Sie auf die Parser-Umwandlung.
- Klicken Sie auf **Fenster > Ansicht zeigen > Eigenschaften**.
- Wählen Sie in der Ansicht `Eigenschaften` die Ansicht **Strategien** aus.
- Klicken Sie auf **Neu**. Der Assistent für die neue Strategie wird aufgerufen.
- Klicken Sie auf den Auswahlpfeil in der Eingabespalte, und wählen Sie den Port `FullName`.
- Wählen Sie den Leerschritt-Delimiter `[\s]`.
- Klicken Sie auf **Weiter**.
- Wählen Sie die Operation **Parsen mit Tokensatz**, und klicken Sie auf **Weiter**.
- Wählen Sie **Fester Token-Satz (nur Einzelausgabe)**, und wählen Sie den Tokensatz **Undefiniert**.
- Klicken Sie auf das Feld **Ausgaben**, und wählen Sie **Neu** aus.
- Im Dialogfenster **Operationsausgaben** ändern Sie den Ausgabenamen in `Undefined_Output`.
- Klicken Sie auf **Fertigstellen**.
- Klicken Sie in der Parser-Umwandlung auf den Port `Undefined_Output`, und ziehen Sie ihn in den Port `FirstName` im Datenobjekt `LA_customers_tgt`.
Zwischen den Ports erscheint eine Verbindung.
- Klicken Sie in der Parser-Umwandlung auf den Port `OverflowField`, und ziehen Sie ihn in den Port `LastName` im Datenobjekt `LA_customers_tgt`.

Zwischen den Ports erscheint eine Verbindung.

16. Klicken Sie zum Speichern des Mappings auf **Datei > Speichern**.

Aufgabe 3. Ausführen eines Profils für die Parser-Umwandlung

In dieser Aufgabe führen Sie ein Profil auf der Parser-Umwandlung aus, um zu überprüfen, dass Sie die Parser-Umwandlung so konfiguriert haben, dass der vollständige Name korrekt geparkt wird.

1. Wählen Sie den Editor aus, der das Mapping `ParserMapping` enthält.
2. Klicken Sie mit der rechten Maustaste auf die Umwandlung, und wählen Sie **Profil jetzt** aus.
Das Profil wird gestartet und im Editor geöffnet.
3. Klicken Sie im Editor auf die Ansicht **Ergebnisse**, um das Ergebnis der Profilierungsoperation anzuzeigen..
4. Wählen Sie die Spalte `Undefined_output`, um Informationen über die Spalte im Abschnitt **Details** anzuzeigen.
Die in der Spalte `Undefined_output` enthaltenen Werte erscheinen im Abschnitt **Details** zusammen mit der Frequenz und der Prozentstatistik für jeden Wert.
5. Zeigen Sie die Daten an, und überprüfen Sie, dass in der Spalte `Undefined_output` nur Vornamen stehen.

Aufgabe 4. Ausführen des Mappings

Bei dieser Aufgabe führen Sie das Mapping aus, um die Mapping-Ausgabe zu erstellen.

1. Wählen Sie den Editor aus, der das Mapping `ParserMapping` enthält.
2. Klicken Sie auf **Ausführen > Mapping ausführen**.
Das Mapping wird ausgeführt und schreibt Ausgaben in die Datei `LA_Customers_tgt.csv`.

Aufgabe 5. Anzeigen der Mapping-Ausgabe

In dieser Aufgabe führen Sie den Daten-Viewer aus, um die Mapping-Ausgabe anzuzeigen.

1. Gehen Sie in der Ansicht Objekt-Explorer zum Datenobjekt `LA_Customers_tgt` in Ihrem Tutorial-Projekt, und doppelklicken Sie auf das Datenobjekt.
Das Datenobjekt wird im Editor geöffnet.
2. Klicken Sie auf **Fenster > Ansicht zeigen > Daten-Viewer**.
Die Ansicht Daten-Viewer wird geöffnet.
3. Klicken Sie in der Ansicht Daten-Viewer auf **Ausführen**.
Der Daten-Viewer wird ausgeführt, und die Daten werden angezeigt.

4. Vergewissern Sie sich, dass in den Spalten `FirstName` und `LastName` korrekt geparste Daten angezeigt werden.

Parsen von Daten - Zusammenfassung

Sie haben in dieser Lektion gelernt, dass das Parsen von Daten die Datenelemente in einem Eingabefeld kennzeichnet und jedes Element in eine neue Spalte schreibt.

Sie haben gelernt, dass Sie Daten mithilfe der Parser-Umwandlung parsen. Außerdem haben Sie gelernt, dass Sie ein Profil für eine Umwandlung in einem Mapping erstellen können, um die Ausgabe dieser Umwandlung zu analysieren. Schließlich haben Sie gelernt, dass Sie die Mapping-Ausgabe mit dem Daten-Viewer anzeigen können.

Sie haben das Datenobjekt `LA_Customers_tgt` erstellt und so konfiguriert, dass es die geparste Ausgabe enthält. Sie haben ein Mapping erstellt, um die Daten zu parsen. In diesem Mapping haben Sie eine Parser-Umwandlung mit einem Tokensatz zum Parsen von Vornamen und Nachnamen aus der Spalte `FullName` der Kundendatei von Los Angeles konfiguriert. Sie haben das Mapping so konfiguriert, dass die geparsten Daten in die Spalten `Firstname` und `Lastname` im Datenobjekt `LA_Customers_tgt` geschrieben werden. Sie haben außerdem ein Profil ausgeführt, um die Ausgabe der Umwandlung vor dem Ausführen des Mappings anzuzeigen. Schließlich haben Sie das Mapping ausgeführt, um mit dem Daten-Viewer die neuen Datenspalten im Datenobjekt `LA_Customers_tgt` anzuzeigen.

KAPITEL 14

Lektion 5. Standardisieren von Daten

Dieses Kapitel umfasst die folgenden Themen:

- [Standardisieren von Daten - Übersicht, 74](#)
- [Aufgabe 1. Erstellen eines Target-Datenobjekts, 75](#)
- [Aufgabe 2 Erstellen eines Mappings zum Standardisieren von Daten, 76](#)
- [Aufgabe 3. Ausführen des Mappings, 79](#)
- [Aufgabe 4. Anzeigen der Mapping-Ausgabe, 79](#)
- [Standardisieren von Daten - Zusammenfassung, 79](#)

Standardisieren von Daten - Übersicht

Das Standardisieren von Daten dient der Qualitätsverbesserung durch Entfernen von Fehlern und Inkonsistenzen in den Daten.

Zur Qualitätsverbesserung sind Daten zu standardisieren, die folgende Wertetypen enthalten:

- Falsche Werte
- Werte mit korrekten Informationen im falschen Format
- Werte, von denen Sie neue Informationen ableiten möchten

Mit der Standardisierer-Umwandlung können Sie diese Daten in Werten suchen. Sie können einen der folgenden Suchoperationstypen verwenden:

- **Text.** Suchen nach von Ihnen eingegebenen benutzerspezifischen Strings. Entfernen Sie diese Strings, oder ersetzen Sie sie durch benutzerdefinierten Text.
- **Referenztabelle.** Suchen nach in einer von Ihnen auszuwählenden Referenztabelle enthaltenen Strings. Entfernen Sie diese Strings, oder ersetzen Sie sie durch Referenztabelleneinträge bzw. benutzerdefinierten Text.

Sie können die Standardisierer-Umwandlung beispielsweise auf Standardisierung der die benutzerdefinierten Strings `Street` und `St.` enthaltenden Adressdaten durch den Ersatz-String `ST.` konfigurieren. Die Standardisierer-Umwandlung ersetzt die Suchbegriffe durch den Begriff `ST.` und schreibt das Ergebnis in eine neue Datenspalte.

Story

HypoStores muss seine Kundenadressdaten so standardisieren, dass alle Adressen die Begriffe konsistent verwenden. Die Adressdaten im Datenobjekt `All_Customers` enthalten inkonsistent formatierte Einträge für normale Begriffe wie `Street`, `Boulevard`, `Avenue`, `Drive` und `Park`.

Zielsetzungen

In dieser Lektion führen Sie folgende Aufgaben durch:

- Erstellen und Konfigurieren eines Datenobjekts `All_Customers_Stdz_tgt` so, dass es standardisierte Daten enthält.
- Erstellen eines Mappings zum Standardisieren der Adressbegriffe `Street`, `Boulevard`, `Avenue`, `Drive` und `Park` in ein konsistentes Format.
- Hinzufügen des Datenobjekts `All_Customers` zum Mapping zum Verbinden der Quelldaten.
- Hinzufügen des Datenobjekts `All_Customers_Stdz_tgt` zum Mapping zwecks Erstellung eines Datenobjekts.
- Hinzufügen einer Standardisierer-Umwandlung zum Mapping und dessen Konfiguration auf Standardisierung der Adressbegriffe.
- Ausführen des Mappings zum Generieren standardisierter Adressdaten.
- Ausführen des Daten-Viewers zur Anzeigen der Mapping-Ausgabe.

Voraussetzungen

Überprüfen Sie, ob folgende Voraussetzung erfüllt ist, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 und 2 in diesem Tutorial abgeschlossen.

Zeitaufwand

Planen Sie für diese Lektion 20 Minuten ein.

Aufgabe 1. Erstellen eines Target-Datenobjekts

In dieser Aufgabe erstellen Sie ein Datenobjekt `All_Customers_Stdz_tgt`, in das Sie standardisierte Daten schreiben können.

Zum Erstellen eines Target-Datenobjekts gehen Sie folgendermaßen vor:

1. Erstellen Sie ein Datenobjekt `All_Customers_Stdz_tgt` basierend auf der Datei `All_Customers.csv`.
2. Konfigurieren Sie die Lesen- und Schreiben-Optionen für das Datenobjekt einschließlich Dateispeicherorten und Dateinamen.

Schritt 1. Erstellen eines Datenobjekts `All_Customers_Stdz_tgt`

In diesem Schritt erstellen Sie ein Datenobjekt `All_Customers_Stdz_tgt` basierend auf der Datei `All_Customers.csv`.

1. Klicken Sie auf **Datei > Neu > Datenobjekt**.
Das Fenster **Neu** wird geöffnet.
2. Wählen Sie **Flatfile-Datenobjekt**, und klicken Sie auf **Weiter**.
3. Vergewissern Sie sich, dass **Aus einem vorhandenen Flatfile erstellen** ausgewählt wurde.

4. Klicken Sie auf **Durchsuchen**, und navigieren Sie zu `All_Customers.csv` im Verzeichnis `<Informatica-Installationsverzeichnis>\clients\DeveloperClient\Tutorials`.
5. Klicken Sie auf **Öffnen**.
6. Geben Sie im Feld Name `All_Customers_Stdz_tgt` ein.
7. Klicken Sie auf **Weiter**.
8. Klicken Sie auf **Weiter**.
9. Wählen Sie im Abschnitt **Vorschauoption Spaltennamen aus der ersten Zeile importieren**, und klicken Sie auf **Weiter**.
10. Klicken Sie auf **Fertigstellen**.
Im Editor erscheint das Datenobjekt `All_Customers_Stdz_tgt`.

Schritt 2. Konfigurieren der Lesen- und Schreiben-Optionen

In diesem Schritt konfigurieren Sie die Lesen- und Schreiben-Optionen für das Datenobjekt `All_Customers_Stdz_tgt`, einschließlich der Dateispeicherorte und Dateinamen.

1. Überprüfen Sie, dass das Datenobjekt `All_Customers_Stdz_tgt` im Editor geöffnet ist.
2. Wählen Sie im Editor die Ansicht **Lesen**.
3. Klicken Sie auf **Fenster > Ansicht zeigen > Eigenschaften**.
4. Wählen Sie in der Ansicht **Eigenschaften** die Ansicht **Laufzeit**.
5. Doppelklicken Sie in der Spalte **Wertauf** den Quelldateinamen und -typ `All_Customers_Stdz_tgt.csv`.
6. In der Spalte **Wert** doppelklicken Sie auf den Eintrag **Quelldateiverzeichnis**.
7. Klicken Sie mit der rechten Maustaste auf den hervorgehobenen Namen, und wählen Sie **Kopieren** aus.
8. Wählen Sie im Editor die Ansicht **Schreiben**.
9. Wählen Sie in der Ansicht **Eigenschaften** die Ansicht **Laufzeit** aus.
10. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Ausgabedateiverzeichnis**.
11. Klicken Sie mit der rechten Maustaste, und wählen Sie **Einfügen**, um den Verzeichnisspeicherort einzufügen, den Sie aus der Ansicht **Lesen** kopiert haben.
12. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Kopfzeilenoptionen**, und wählen Sie **Ausgabefeldnamen**.
13. In der Spalte **Wert** doppelklicken Sie auf den Eintrag **Ausgabedateiname** und den Typ `All_Customers_Stdz_tgt.csv`.
14. Klicken Sie zum Speichern des Datenobjekts auf **Datei > Speichern**.

Aufgabe 2 Erstellen eines Mappings zum Standardisieren von Daten

In dieser Aufgabe erstellen Sie ein Mapping und konfigurieren das Mapping auf die Verwendung von Datenobjekten und einer Standardisierer-Umwandlung.

Um ein Mapping zum Standardisieren von Daten zu erstellen, gehen Sie folgendermaßen vor:

1. Erstellen Sie ein Mapping.

2. Fügen Sie dem Mapping Quell- und Target-Objekte hinzu.
3. Fügen Sie eine Standardisierer-Umwandlung zum Mapping hinzu.
4. Konfigurieren Sie die Standardisierer-Umwandlung auf die Standardisierung normaler Adressbegriffe in konsistente Formate.

Schritt 1. Erstellen eines Mappings

In diesem Schritt erstellen Sie das Mapping und geben ihm einen Namen.

1. Wählen Sie das Schulungsprojekt in der Ansicht **Objekt-Explorer** aus.
2. Klicken Sie auf **Datei > Neu > Mapping**.
Das Fenster **Neues Mapping** wird aufgerufen.
3. Geben Sie im Feld **Name** `StandardizerMapping` ein.
4. Klicken Sie auf **Fertigstellen**.
Das Mapping wird im Editor geöffnet.

Schritt 2. Hinzufügen von Datenobjekten zum Mapping

In diesem Schritt fügen Sie das Datenobjekt `All_Customers` und das Datenobjekt `All_Customers_Stdz_tgt` zum Mapping hinzu.

1. Gehen Sie in der Ansicht **Objekt-Explorer** zu den Datenobjekten in Ihrem Schulungsprojekt.
2. Wählen Sie das Datenobjekt `All_Customers` aus, und ziehen Sie es in den Editor.
Das Fenster **Physisches Datenobjekt zum Mapping hinzufügen** wird geöffnet.
3. Vergewissern Sie sich, dass **Lesen** ausgewählt wurde, und klicken Sie auf **OK**.
Das Datenobjekt erscheint im Editor.
4. Gehen Sie in der Ansicht **Objekt-Explorer** zu den Datenobjekten in Ihrem Schulungsprojekt.
5. Wählen Sie das Datenobjekt `All_Customers_Stdz_tgt` aus, und ziehen Sie es in den Editor.
Das Fenster **Physisches Datenobjekt zum Mapping hinzufügen** wird geöffnet.
6. Wählen Sie **Schreiben** aus, und klicken Sie auf **OK**.
Das Datenobjekt erscheint im Editor.
7. Wählen Sie alle Ports im Datenobjekt `All_Customers`. Ziehen Sie die Ports zum Port `CustomerID` im Datenobjekt `All_Customers_Stdz_tgt`.
Tipp: Halten Sie die Hochstelltaste gedrückt, um mehrere Spalten auszuwählen. Möglicherweise müssen Sie in der Portliste herunterscrollen, um alle auszuwählen.
Die Ports des Datenobjekts `All_Customers` stellen die Verbindung zu den entsprechenden Ports im Datenobjekt `All_Customers_Stdz_tgt` her.

Schritt 3. Hinzufügen einer Standardisierer-Umwandlung zum Mapping

In diesem Schritt fügen Sie eine Standardisierer-Umwandlung zum Standardisieren von Strings in den Adressdaten hinzu.

1. Wählen Sie den Editor aus, der das Mapping `StandardizerMapping` enthält.
2. In der Umwandlungspalette wählen Sie die Standardisierer-Umwandlung aus.

3. Klicken Sie auf den Editor.
Eine Standardisierer-Umwandlung mit dem Namen `NewStandardizer` erscheint im Mapping.
4. Um die Standardisierer-Umwandlung umzubenennen, doppelklicken Sie auf die Titelleiste der Umwandlung und den Typ `AddressStandardizer`.
5. Wählen Sie den Port `Address1` im Datenobjekt `All_Customers` aus, und ziehen Sie den Port zur Gruppe **Eingabe** der Standardisierer-Umwandlung.
Ein Port mit dem Namen `Address1` erscheint in der Eingabegruppe. Der Port stellt die Verbindung zum Port `Address1` im Datenobjekt `All_Customers` her.

Hinweis: Sie fügen einen Ausgabeport zur Umwandlung hinzu, wenn Sie eine Standardisierungsstrategie konfigurieren.

Schritt 4. Konfigurieren der Standardisierer-Umwandlung

In diesem Schritt konfigurieren Sie die Standardisierer-Umwandlung zum Standardisieren von Adressbegriffen in den Quelldaten.

Hinweis: In dieser Aufgabe definieren Sie fünf Standardisierungs-Operationen. Jede Operation ersetzt einen String in der Eingabespalte durch einen neuen String.

1. Wählen Sie den Editor, der das Mapping `StandardizerMapping` enthält.
2. Klicken Sie auf die Standardisierer-Umwandlung.
3. Klicken Sie auf **Fenster > Ansicht zeigen > Eigenschaften**.
4. Wählen Sie in der Ansicht Eigenschaften **Strategien**.
5. Klicken Sie auf **Neu**. Der Assistent für die neue Strategie wird aufgerufen.
6. Klicken Sie auf den Auswahlpfeil in der Eingabespalte, und wählen Sie den Eingabeport `Address1`.
Im Ausgabefeld steht als Ausgabeport `Address1`.
7. Wählen Sie die Leerzeichen- und Komma-Delimiter `[\s]` und `[,]`. Wählen Sie optional die Optionen zum Entfernen nachgelagerter Leerzeichen.
8. Klicken Sie auf **Weiter**.
9. Wählen Sie die Operation **Kundenspezifische Strings ersetzen**, und klicken Sie auf **Weiter**.
10. Klicken Sie unter **Eigenschaften** auf **Neu**.
11. Bearbeiten Sie die Felder **Benutzerdef. Strings** und **Ersetzen durch**, sodass sie das erste String-Paar aus der folgenden Tabelle enthalten:

Benutzerdef. Strings	Ersetzen durch
STREET (STRASSE)	ST. (ST.)
BOULEVARD	BLVD.
AVENUE	AVE.
DRIVE	Dr.
PARK	PK.

12. Wiederholen Sie die Schritte 9 bis 12, um Standardisierungs-Operationen für alle Strings in der Tabelle zu definieren.

13. Ziehen Sie den Ausgabeport `Address1` in den Port `Address1` des Datenobjekts `All_Customers_Stdz_tgt`.
14. Um das Mapping zu speichern, klicken Sie auf **Datei > Speichern**.

Aufgabe 3. Ausführen des Mappings

Im Rahmen dieser Aufgabe führen Sie das Mapping aus, um standardisierte Adressen in das Ausgabedatenobjekt zu schreiben.

1. Wählen Sie den Editor, der das Mapping `StandardizerMapping` enthält.
2. Klicken Sie auf **Ausführen > Mapping ausführen**.

Das Mapping wird ausgeführt und schreibt die Ausgabe in die Datei `All_Customers_Stdz_tgt.csv`.

Aufgabe 4. Anzeigen der Mapping-Ausgabe

In dieser Aufgabe führen Sie den Daten-Viewer zum Anzeigen der Mapping Ausgabe aus und überprüfen, dass die Mapping-Ausgabedaten korrekt standardisiert wurden.

1. Gehen Sie in der Ansicht **Objekt-Explorer** zum Datenobjekt `All_Customers_Stdz_tgt` in Ihrem Tutorial-Projekt, und doppelklicken Sie auf das Datenobjekt.

Das Datenobjekt wird im Editor geöffnet.

2. Klicken Sie auf **Fenster > Ansicht zeigen > Daten-Viewer**.

Die Ansicht Daten-Viewer wird geöffnet.

3. Klicken Sie in der Ansicht Daten-Viewer auf **Ausführen**.

Der Daten-Viewer zeigt die Mapping-Ausgabe an.

4. Vergewissern Sie sich, dass in der Spalte `Address1` korrekt standardisierte Daten angezeigt werden. Beispielsweise sollten alle Instanzen des Strings `STREET` durch den String `ST` ersetzt werden.

Standardisieren von Daten - Zusammenfassung

In dieser Lektion haben Sie gelernt, dass Sie Daten standardisieren können, um Fehler und Inkonsistenzen in den Daten zu entfernen.

Sie haben gelernt, dass Sie mit einer Standardisierer-Umwandlung Strings in einer Eingabespalte standardisieren können. Außerdem haben Sie erfahren, wie man Mapping-Ausgaben mit dem Daten-Viewer anzeigt.

Sie haben das Datenobjekt `All_Customers_Stdz_tgt` erstellt und so konfiguriert, dass es standardisierte Ausgaben enthält. Sie haben ein Mapping zum Standardisieren der Daten erstellt. In diesem Mapping haben Sie eine Standardisierer-Umwandlung zum Standardisieren der Spalte `Address1` im Datenobjekt `All_Customers` konfiguriert. Sie haben das Mapping so konfiguriert, dass die standardisierte Ausgabe in das Datenobjekt `All_Customers_Stdz_tgt` geschrieben wird. Schließlich haben Sie das Mapping ausgeführt und die standardisierten Daten im Datenobjekt `All_Customers_Stdz_tgt` mit dem Daten-Viewer angezeigt.

KAPITEL 15

Lektion 6. Validieren von Adressdaten

Dieses Kapitel umfasst die folgenden Themen:

- [Validieren von Adressdaten - Übersicht, 80](#)
- [Aufgabe 1 Erstellen eines Target-Datenobjekts , 81](#)
- [Aufgabe 2 Erstellen eines Mappings zum Validieren von Adressen, 83](#)
- [Aufgabe 3. Konfigurieren der Adressvalidierungsumwandlung, 84](#)
- [Aufgabe 4. Ausführen des Mappings, 87](#)
- [Aufgabe 5. Anzeigen der Mapping-Ausgabe, 88](#)
- [Validieren von Adressdaten - Zusammenfassung, 90](#)

Validieren von Adressdaten - Übersicht

Unter Adressvalidierung versteht man die Evaluierung und Verbesserung der Qualität von Postadressen. Dabei wird die Adressqualität durch Vergleichen der Eingabeadressen mit einem Referenzdatensatz gültiger Adressen evaluiert. Die Adressqualität wird verbessert, indem falsche Adresswerte erkannt und anhand des Referenzdatensatzes Felder mit richtigen Werten erstellt werden.

Eine Adresse ist gültig, wenn sie zustellfähig ist. Eine Adresse kann gut formatiert sein und Straße, Ort und Postleitzahl enthalten, aber wenn die Daten keine zustellungsfähige Adresse ergeben, ist die Adresse nicht gültig. Das Developer Tool prüft die Zustellungsfähigkeit der Eingabeadressen anhand von Adress-Referenz-Datensätzen. Informatica stellt Adressreferenz-Datensätze zur Verfügung.

Ein Adressreferenz-Datensatz enthält Daten, die alle zustellungsfähigen Adressen eines Landes enthalten. Der Adressvalidierungsprozess sucht den Referenz-Datensatz für die Adresse, die den Eingabeadressdaten am genauesten entspricht. Findet der Prozess eine annähernde Übereinstimmung im Referenzdatensatz, schreibt er neue Werte für etwaige falsche oder unvollständige Daten. Bei diesem Prozess wird ein Satz alphanumerischer Codes erstellt, die den Typ der gefundenen Übereinstimmung von Eingabeadresse und Referenzadresse beschreiben. Dabei kann die Adresse neu strukturiert werden. Außerdem können Informationen hinzugefügt werden, die in der Eingabeadresse fehlen, wie ein vierstelliges Postleitzahlensuffix für eine Adresse in den Vereinigten Staaten.

Mit der Adressvalidierer-Umwandlung können Sie Adressvalidierungsprozesse im Developer Tool aufbauen. Diese Multi-Gruppen-Umwandlung enthält einen Satz vordefinierter Eingabeports und Ausgabepports, die allen möglichen Feldern in einer Eingabeadresse entsprechen. Beim Konfigurieren einer Adressvalidierer-Umwandlung stellen Sie den Standard-Referenz-Datensatz ein und erstellen anhand der Umwandlungsports

eine Eingabe- und Ausgabeadressvorlage. In dieser Lektion konfigurieren Sie die Umwandlung für die Validierung der US-Adressdaten.

Story

HypoStores erfordert richtige und vollständige Daten, damit gewährleistet ist, dass seine Direktwerbungskampagnen und andere Verbraucher-Postsendungen seine Kunden erreichen. Richtige und vollständige Adressdaten helfen, die Postversandkosten der Organisation zu senken. Außerdem müssen die Daten der HypoStores-Kunden Adressen in druckbarem Format enthalten, das flexibel genug ist, um Adressen unterschiedlicher Länge bearbeiten zu können.

Um diesen geschäftlichen Anforderungen gerecht zu werden, erstellt das HypoStores ICC-Team ein Adressvalidierungs-Mapping im Developer Tool.

Zielsetzungen

In dieser Lektion führen Sie folgende Aufgaben durch:

- Erstellen eines Target-Datenobjekts, das die validierten Adressfelder und Match-Codes enthält.
- Erstellen eines Mappings mit einem Quelldatenobjekt, einem Target-Datenobjekt und einer Adressvalidierer-Umwandlung.
- Konfigurieren der Adressvalidierer-Umwandlung für die Validierung der Adressdaten Ihrer Kunden.
- Ausführen des Mappings zum Validieren der Adressdaten und Überprüfen der Match-Code-Ausgaben zur Kontrolle der Gültigkeit der Adressdaten.

Voraussetzungen

Überprüfen Sie, ob die folgenden Voraussetzungen erfüllt sind, bevor Sie mit dieser Lektion beginnen:

- Sie haben die Lektionen 1 und 2 in diesem Tutorial abgeschlossen.
- Die US-Adressreferenzdaten wurden in der Domäne installiert und mit dem Administrator Tool registriert. Zwecks Überprüfung, ob auf Ihrem System US-Adressdaten installiert sind, setzen Sie sich bitte mit Ihrem Informatica-Administrator in Verbindung. Die Referenzdaten werden mit dem Datenqualitätsinhalt-Installer installiert.

Zeitaufwand

Bitte veranschlagen Sie für diese Lektion 25 Minuten.

Aufgabe 1 Erstellen eines Target-Datenobjekts

In dieser Aufgabe erstellen Sie ein Target-Datenobjekt, konfigurieren die Schreiben-Optionen und fügen Ports hinzu.

Um das Target-Datenobjekt zu erstellen und zu konfigurieren, führen Sie die folgenden Schritte aus:

1. Erstellen eine Datenobjekts All_Customers_av_tgt basierend auf der Datei All_Customers.csv.
2. Konfigurieren der Lesen- und Schreiben-Optionen für das Datenobjekt einschließlich der Datei-Speicherorte und Dateinamen.
3. Hinzufügen von Ports zum Datenobjekt, um die von der Adress-Validierer-Umwandlung generierten Match-Code-Werte zu erhalten.

Schritt 1. Erstellen des Datenobjekts All_Customers_av_tgt

In diesem Schritt erstellen Sie ein Datenobjekt `All_Customers_av_tgt` basierend auf der Datei `All_Customers.csv`.

1. Klicken Sie auf **Datei > Neu > Datenobjekt**.
Das Fenster **Neu** wird aufgerufen.
2. Wählen Sie **Flatfile-Datenobjekt**, und klicken Sie auf **Weiter**.
3. Vergewissern Sie sich, dass **Aus einem vorhandenen Flatfile erstellen** ausgewählt wurde. Klicken Sie auf **Durchsuchen** neben dieser Auswahl, suchen Sie die Datei `All_Customers.csv`, und klicken Sie auf **Öffnen**.
4. Geben Sie im Feld `Name` `All_Customers_av_tgt` ein.
5. Klicken Sie auf **Weiter**.
6. Klicken Sie auf **Weiter**.
7. Wählen Sie im Abschnitt **Vorschauoptionen** **Spaltennamen aus erster Zeile importieren**, und klicken Sie auf **Weiter**.
8. Klicken Sie auf **Fertigstellen**.
Im Editor erscheint das Datenobjekt `All_Customers_av_tgt`.

Schritt 2. Konfigurieren der Lesen- und Schreiben-Optionen

In diesem Schritt konfigurieren Sie die Lesen- und Schreiben-Optionen für das Datenobjekt `All_Customers_av_tgt`, einschließlich Target-Dateiname und Speicherort.

1. Überprüfen Sie, dass das Datenobjekt `All_Customers_av_tgt` im Editor geöffnet ist.
2. Wählen Sie im Editor die Ansicht **Lesen**.
3. Wählen Sie **Fenster > Ansicht zeigen > Eigenschaften**.
4. Wählen Sie in der Ansicht **Eigenschaften** die Ansicht **Laufzeit** aus.
5. Doppelklicken Sie in der Spalte **Wert** auf den Quellendateinamen und -typ `All_Customers_av_tgt.csv`.
6. In der Spalte **Wert** doppelklicken Sie, um den Quelldatei-Verzeichnispfad zu markieren.
7. Klicken Sie mit der rechten Maustaste auf den markierten Pfad und Namen, und wählen Sie **Kopieren**.
8. Wählen Sie im Editor die Ansicht **Schreiben**.
9. Wählen Sie in der Ansicht **Eigenschaften** die Ansicht **Laufzeit** aus.
10. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Ausgabedateiverzeichnis**.
11. Klicken Sie mit der rechten Maustaste auf diesen Eintrag, und wählen Sie **Einfügen**, um den aus der Ansicht **Lesen** kopierten Pfad hinzuzufügen.
12. Doppelklicken Sie in der Spalte **Wert** auf den Eintrag **Kopfzeilenoptionen**, und wählen Sie **Ausgabefeldnamen**.
13. In der Spalte **Wert** doppelklicken Sie auf den Eintrag **Ausgabedateiname**, und geben Sie `All_Customers_av_tgt.csv` ein.
14. Zum Speichern der Datenobjekte wählen Sie **Datei > Speichern** aus.

Schritt 3. Hinzufügen von Ports zum Datenobjekt

In diesem Schritt fügen Sie zwei Ports zum Datenobjekt `All_Customers_av_tgt` hinzu, damit die Adressvalidierer-Umwandlung Match-Codewerte in die Target-Datei schreiben kann. Geben Sie den Ports die Namen `MailabilityScore` und `MatchCode`.

Der Wert `MailabilityScore` beschreibt die Zustellbarkeit der Eingabeadresse. Der Wert `MatchCode` beschreibt den Match-Typ, den die Umwandlung zwischen Eingabeadresse und Referenzdatenadresse ergibt.

1. Gehen Sie in der Ansicht **Objekt-Explorer** zu den Datenobjekten in Ihrem Tutorial-Projekt.
2. Doppelklicken Sie auf das Datenobjekt `All_Customers_av_tgt`.
Das Datenobjekt `All_Customers_av_tgt` wird im Editor geöffnet.
3. Vergewissern Sie sich, dass **Übersicht** ausgewählt ist.
4. Wählen Sie den endgültigen Port in der Portliste aus. Dieser Port trägt den Namen `MiscDate`.
5. Klicken Sie auf **Neu**.
Ein Port mit dem Namen `MiscDate1` wird angezeigt.
6. Benennen Sie den Port `MiscDate1` in `MailabilityScore` um.
7. Wählen Sie den Port `MailabilityScore`.
8. Klicken Sie auf **Neu**.
Ein Port mit dem Namen `MailabilityScore1` wird angezeigt.
9. Ändern Sie den Namen des Ports `MailabilityScore1` in `MatchCode`.
10. Klicken Sie zum Speichern des Datenobjekts auf **Datei > Speichern**.

Aufgabe 2 Erstellen eines Mappings zum Validieren von Adressen

In dieser Aufgabe erstellen Sie ein Mapping und fügen Datenobjekte und eine Adressvalidierer-Umwandlung hinzu.

Zum Erstellen des Mappings und Hinzufügen der benötigten Objekte gehen Sie folgendermaßen vor:

1. Erstellen Sie ein Mapping-Objekt.
2. Fügen Sie Quellen- und Target-Datenobjekte zum Mapping hinzu.
3. Fügen Sie eine Adressvalidierer-Umwandlung zum Mapping hinzu.

Schritt 1. Erstellen eines Mappings

In diesem Schritt erstellen Sie das Mapping und geben ihm einen Namen.

1. Wählen Sie in der Ansicht **Objekt-Explorer** das Tutorial-Projekt aus.
2. Wählen Sie **Datei > Neu > Mapping**.
Das Fenster **Neues Mapping** wird geöffnet.
3. Geben Sie im Feld **Name** `ValidateAddresses` ein.
4. Klicken Sie auf **Fertigstellen**.
Das Mapping wird im Editor geöffnet.

Schritt 2. Hinzufügen von Datenobjekten zum Mapping

In diesem Schritt fügen Sie Quellen- und Target-Datenobjekte zum Mapping hinzu.

`All_Customers` ist das Quelldatenobjekt für das Mapping. Die Adressvalidierer-Umwandlung liest Daten aus diesem Objekt. `All_Customers_av_tgt` ist das Daten-Target-Objekt für das Mapping. Dieses Objekt liest Daten aus der Adressvalidierer-Umwandlung.

1. Gehen Sie in der Ansicht **Objekt-Explorer** zu den Datenobjekten in Ihrem Schulungsprojekt.
2. Wählen Sie das Datenobjekt `All_Customers` aus, und ziehen Sie es in den Editor.
Das Fenster **Physisches Datenobjekt zum Mapping hinzufügen** wird geöffnet.
3. Vergewissern Sie sich, dass **Lesen** ausgewählt ist, und klicken Sie auf **OK**.
Das Datenobjekt erscheint im Editor.
4. Gehen Sie in der Ansicht **Objekt-Explorer** zu den Datenobjekten in Ihrem Tutorial-Projekt.
5. Wählen Sie das Datenobjekt `All_Customers_av_tgt` aus, und ziehen Sie es in den Editor.
Das Fenster **Physisches Datenobjekt zum Mapping hinzufügen** wird geöffnet.
6. Wählen Sie **Schreiben** aus, und klicken Sie auf **OK**.
Das Datenobjekt erscheint im Editor.
7. Klicken Sie auf **Speichern**.

Schritt 3. Hinzufügen einer Adressvalidierer-Umwandlung zum Mapping

In diesem Schritt fügen Sie eine Adressvalidierer-Umwandlung zu dem Mapping hinzu, das die Quelle und die Datenobjekte enthält.

Sobald dieser Schritt abgeschlossen ist, können Sie die Umwandlung konfigurieren und ihre Ports mit den Datenobjekten verbinden.

1. Wählen Sie den Editor, der das Mapping `ValidateAddresses` enthält.
2. Wählen Sie in der Umwandlungspalette die Adressvalidierer-Umwandlung.
3. Klicken Sie auf den Editor.
Die Adressvalidierer-Umwandlung erscheint im Editor.

Aufgabe 3. Konfigurieren der Adressvalidierungsumwandlung

In dieser Aufgabe konfigurieren Sie die Adressvalidierer-Umwandlung auf das Lesen und Validieren von Adressen aus der Datenquelle `All_Customers`.

Hinweis: Die Adressvalidierer-Umwandlung enthält eine Reihe vordefinierter Eingabe- und Ausgabe-Ports. Wählen Sie die benötigten Ports aus, und verbinden Sie sie mit den Objekten im Mapping.

Zum Konfigurieren der Umwandlung gehen Sie bitte folgendermaßen vor:

1. Wählen Sie das Standardland für die Adressvalidierung aus.
2. Konfigurieren Sie die Eingabe-Ports für die Umwandlung.

3. Konfigurieren Sie die Ausgabe-Ports für die Umwandlung.
4. Verbinden Sie unbenutzte Quellen-Ports mit dem Daten-Target.

Schritt 1. Festlegen des Standardlands für Adressvalidierung

In diesem Schritt wählen Sie das Standardland für die Adressvalidierung aus. Die Referenzdaten-Dateien für die Adresse, die bei der Adressvalidierer-Umwandlung verwendet werden, sind nach Land organisiert. Wenn Sie das Standardland auswählen, geben Sie den Länderdatensatz an, der bei der Umwandlung auf jede Eingabeadresse angewendet wird, die keine Länderinformationen enthalten.

1. Wählen Sie die Adressvalidierer-Umwandlung im Editor aus.
2. Klicken Sie unter **Eigenschaften** auf **Allgemeine Einstellungen**.
3. Wählen Sie im Menü **Standardland** den Eintrag `Vereinigte Staaten` aus.

Schritt 2. Konfigurieren der Eingabeports der Adressvalidierer-Umwandlung

In diesem Schritt wählen Sie die Umwandlungs-Eingabeports aus und verbinden diese Ports mit dem Datenobjekt `All_Customers_av`.

Die Adressvalidierer-Umwandlung enthält mehrere Gruppen vordefinierter Eingabe-Ports. Wählen Sie Eingabeports aus, die den Feldern in Ihrer Eingabeadresse entsprechen, und fügen Sie diese Ports zur Umwandlung hinzu.

Halten Sie die Strg-Taste gedrückt, wenn Sie in den folgenden Schritten zum Auswählen mehrerer Ports in einem Arbeitsgang Ports auswählen.

1. Wählen Sie die Adressvalidierer-Umwandlung im Editor aus.
2. Klicken Sie unter **Eigenschaften** auf **Vorlagen**.
3. Erweitern Sie die Portgruppe **Basismodell**.
4. Erweitern Sie die Eingabeportgruppe **Hybrid**, und wählen Sie folgende Ports aus:

Port-Name	Beschreibung
Lieferadresszeile 1	Straßenadressdaten, wie Straßenname und Hausnummer.
Vollständiger Name der Region 1	Name des Ortes.
Postleitzahl 1	Postcode bzw. PLZ.
Provinz 1	Name der Provinz oder des Bundesstaates.
Name des Landes	Name des Landes bzw. Abkürzung.

Hinweis: Halten Sie die Strg-Taste gedrückt, um mehrere Ports in einem Arbeitsgang auszuwählen.

5. Klicken Sie auf die Symbolleiste oberhalb der Liste der Port-Namen auf **Port zu Umwandlung hinzufügen**. Diese Symbolleiste wird angezeigt, wenn Sie **Vorlagen** auswählen.
Die ausgewählten Ports erscheinen in der Umwandlung im Mapping-Editor.

6. Verbinden Sie die Quellen-Ports folgendermaßen mit den Adressvalidierer-Umwandlungsports:

Quellen-Port	Adressvalidierer-Umwandlungsport
Adresse 1	Lieferadresszeile 1
Ort	Vollständiger Name der Region 1
PLZ	Postleitzahl 1
Bundesland	Provinz 1
Land	Name des Landes

Schritt 3. Konfigurieren der Ausgabe-Ports der Adressvalidierer-Umwandlung

In diesem Schritt wählen Sie die Umwandlungs-Ausgabeports und verbinden diese Ports mit dem Datenobjekt `All_Customers_av_tgt`.

Die Adressvalidierer-Umwandlung enthält mehrere Gruppen vordefinierter Ausgabeports. Wählen Sie die Ports aus, die die benötigte Adressstruktur festlegen, und fügen Sie diese Ports der Umwandlung hinzu.

Sie können ebenfalls Ports auswählen, die Informationen über den für jede Adresse erzielten Validierungstyp enthalten.

1. Wählen Sie die Adressvalidierer-Umwandlung im Mapping-Editor.
2. Klicken Sie unter **Eigenschaften** auf **Vorlagen**.
3. Erweitern Sie die Portgruppe **Basismodell**.
4. Erweitern Sie die Ausgabeportgruppe **Adresselemente**, und wählen Sie folgenden Port aus:

Port-Name	Beschreibung
Vollständiger Name der Straße 1	Straßenadressdaten, wie Straßenname und Hausnummer.

5. Erweitern Sie die Ausgabeportgruppe **LastLine Elements**, und wählen Sie die folgenden Ports aus:

Port-Name	Beschreibung
Vollständiger Name der Region 1	Name des Ortes.
Postleitzahl 1	Postcode oder PLZ.
Provinz Abkürzung 1	Kennung der Provinz oder des Bundesstaates.

Hinweis: Halten Sie die Strg-Taste gedrückt, um mehrere Ports in einem Arbeitsgang auszuwählen.

6. Erweitern Sie die Ausgabeportgruppe **Land**, und wählen Sie den folgenden Port aus:

Port-Name	Beschreibung
Name des Landes 1	Name des Landes.

7. Erweitern Sie die Ausgabeportgruppe **Status-Info**, und wählen Sie folgende Ports aus:

Port-Name	Beschreibung
Zustellbarkeits-Score	Score, der die Chance auf eine erfolgreiche Postauslieferung bewertet.
Match-Code	Code, der den Ähnlichkeitsgrad zwischen Eingabeadresse und Referenzdaten darstellt.

8. Klicken Sie auf der Symbolleiste oberhalb der Liste der Port-Namen auf **Port zu Umwandlung hinzufügen**.

Diese Symbolleiste wird angezeigt, wenn Sie **Vorlagen** auswählen.

9. Verbinden Sie die Addressvalidierer-Umwandlungsports folgendermaßen mit den Ports

`All_Customers_av_tgt`:

Addressvalidierer-Umwandlungsport	Target-Port
Vollständiger Name der Straße 1	Address1 (Adresse 1)
Vollständiger Name der Region 1	Ort
Postleitzahl 1	PLZ
Provinz Abkürzung 1	Bundesstaat
Name des Landes 1	Land
Zustellbarkeits-Score	MailabilityScore
Match-Code	MatchCode

Schritt 4. Verbinden unbenutzter Datenquellenports mit dem Daten-Target

In diesem Schritt verbinden Sie unbenutzte Ports der Datenquelle `All_Customers` mit dem Daten-Target.

- Verbinden Sie die unbenutzten Ports der Datenquelle mit den gleichnamigen Ports des Daten-Targets.

Aufgabe 4. Ausführen des Mappings

In dieser Aufgabe führen Sie das Mapping aus, um die Mapping-Ausgabe zu erstellen.

1. Wählen Sie den Editor aus, der das Mapping `ValidateAddresses` enthält.
2. Wählen Sie **Ausführen > Mapping ausführen**.

Das Mapping wird ausgeführt und die Ausgabe in die Datei `All_Customers_av_tgt.csv` geschrieben.

Aufgabe 5. Anzeigen der Mapping-Ausgabe

In dieser Aufgabe führen Sie den Daten-Viewer aus, um die Mapping-Ausgabe anzuzeigen. Überprüfen Sie die Qualität Ihrer validierten Adressen, indem Sie die in die Spalten "Zustellbarkeits-Score" und "Match-Code" geschriebenen Werte im Target-Datenobjekt untersuchen.

Der MatchCode-Wert ist ein alphanumerischer Code, der den Typ der Validierung darstellt, dem das Mapping die Adresse unterzogen hat.

Der Wert "Zustellbarkeits-Score" ist ein einstelliger Wert, der die Zustellbarkeit der Adresse zusammenfasst.

1. Suchen Sie in der Object-Explorer-Ansicht das Datenobjekt `All_Customers_av_tgt` in Ihrem Tutorial-Projekt und doppelklicken Sie auf das Datenobjekt.

Das Datenobjekt wird im Editor geöffnet.

2. Wählen Sie **Fenster > Ansicht zeigen > Daten-Viewer** aus.

Der Daten-Viewer wird geöffnet.

3. Klicken Sie im Daten-Viewer auf **Ausführen**.

Der Daten-Viewer zeigt die Mapping-Ausgabe an.

4. Führen Sie einen Bildlauf über die Mapping-Ergebnisse aus, sodass die Spalten „Zustellbarkeits-Score“ und „Match-Code“ angezeigt werden.

5. Überprüfen Sie die Werte in der Spalte „Zustellbarkeits-Score“.

Die möglichen Scores sind 0 bis 5. Bei Adressen mit höheren Scores ist die erfolgreiche Zustellung wahrscheinlicher.

6. Überprüfen Sie die Werte in der Spalte „Match-Code“.

Der Match-Code ist ein alphanumerischer Code. Das alphanumerische Zeichen gibt den von der Umwandlung durchgeführten Validierungstyp an, und die Ziffer bezeichnet die Qualität der Endadresse.

In der folgenden Tabelle werden die üblichen Match-Code-Werte beschrieben:

Code	Beschreibung
A1	Beim Adresscode-Lookup wurde eine unvollständige oder eine vollständige Adresse für den Eingabecode gefunden.
A0	Beim Adresscode-Lookup wurde keine Adresse für den Eingabecode gefunden.
C4	Korrigiert. Alle postalisch relevanten Elemente wurden überprüft.
C3	Korrigiert. Bestimmte Elemente können nicht überprüft werden.
C2	Korrigiert, aber der Zustellstatus ist aufgrund von fehlenden Referenzdaten unklar.
C1	Korrigiert, aber der Zustellstatus ist aufgrund von Fehlern durch die Benutzerstandardisierung unklar.
I4	Daten können nicht vollständig berichtet werden, aber es ist ein einzelnes Match mit einer Adresse in den Referenzdaten vorhanden.
I3	Daten können nicht vollständig berichtet werden, aber es sind mehrere Matches mit Adressen in den Referenzdaten vorhanden.

Code	Beschreibung
I2	Daten können nicht berichtigt werden. Batchmodus gibt teilweise vorgeschlagene Adressen zurück.
I1	Daten können nicht berichtigt werden. Batchmodus kann keine Adresse vorschlagen.
N7	Validierungsfehler. Es wurde keine Validierung durchgeführt, da die Einzelzeilenvalidierung nicht entsperrt ist.
N6	Validierungsfehler. Es wurde keine Validierung durchgeführt, da Einzelzeilenvalidierung für das Zielland nicht unterstützt wird.
N5	Validierungsfehler. Es wurde keine Validierung durchgeführt, da die Referenzdatenbank veraltet ist.
N4	Validierungsfehler. Es wurde keine Validierung durchgeführt, da die Referenzdaten beschädigt oder schlecht formatiert sind.
N3	Validierungsfehler. Es wurde keine Validierung durchgeführt, da die Länderdaten nicht entsperrt werden können.
N2	Validierungsfehler. Es wurde keine Validierung durchgeführt, da die benötigte Referenzdatenbank nicht verfügbar ist.
N1	Validierungsfehler. Es wurde keine Validierung durchgeführt, da das Land nicht erkannt oder nicht unterstützt wird.
Q3	Vorschlagslistenmodus. Die Adressvalidierung kann nicht eine oder mehrere vollständige Adressen aus den Adressreferenzdaten abrufen, die der Eingabeadresse entsprechen.
Q2	Vorschlagslistenmodus. Die Adressvalidierung kann die Eingabeadresselemente und die Elemente aus den Adressreferenzdaten kombinieren, um eine vollständige Adresse zu erstellen.
Q1	Vorschlagslistenmodus. Die Adressvalidierung kann keine vollständige Adresse vorschlagen. Um einen Vorschlag für eine vollständige Adresse zu generieren, fügen Sie der Eingabeadresse Daten hinzu.
Q0	Vorschlagslistenmodus. Es sind nicht genügend Eingabedaten zum Generieren eines Vorschlags vorhanden.
RB	Land aus Abkürzung erkannt. Erkennt zwei- und dreistellige ISO-Ländercodes. Kann auch allgemeine Abkürzungen wie „GER“ für Deutschland erkennen.
RA	Land aus der Einstellung „Land erzwingen“ in der Umwandlung erkannt.
R9	Land aus der Einstellung „Standardland“ in der Umwandlung erkannt.
R8	Anhand des Ländernamens erkanntes Land.
R7	Anhand des Ländernamens erkanntes Land. Die Umwandlung hat jedoch Fehler in den Länderdaten ermittelt.

Code	Beschreibung
R6	Anhand der Gebietsdaten erkanntes Land.
R5	Anhand der Provinzdaten erkanntes Land.
R4	Anhand der Großstadtdaten erkanntes Land.
R3	Anhand des Adressformats erkanntes Land.
R2	Anhand eines Skripts erkanntes Land.
R1	Nicht erkanntes Land, da mehrere Entsprechungen zur Verfügung stehen.
R0	Land nicht erkannt.
S4	Parsing-Modus. Die Adresse wurde ordnungsgemäß geparkt.
S3	Parsing-Modus. Die Adresse wurde mit mehreren Ergebnissen geparkt.
S1	Parsing-Modus. Aufgrund einer Nichtübereinstimmung beim Eingabeformat ist ein Parsing-Fehler aufgetreten.
V4	Überprüft. Die Eingabedaten sind korrekt. Die Adressvalidierung prüfte alle postalisch relevanten Elemente und die Eingaben stimmten vollständig überein.
V3	Überprüft. Die Eingabedaten sind korrekt, aber einige oder alle Elemente wurden standardisiert, oder die Eingabe enthält veraltete Namen oder Synonyme.
V2	Überprüft. Die Eingabedaten sind korrekt, aber einige Elemente konnten aufgrund unvollständiger Referenzdaten nicht überprüft werden.
V1	Überprüft. Die Eingabedaten sind korrekt, aber die Benutzerstandardisierung hat negative Auswirkungen auf die Zustellbarkeit. Die Postleitzahl ist beispielsweise zu kurz.

Validieren von Adressdaten - Zusammenfassung

In dieser Lektion haben Sie gelernt, dass bei der Adressvalidierung Eingabeadressdaten mit Referenzdaten verglichen werden und die genauest mögliche Adressversion zurückgegeben wird.

Sie haben gelernt, dass bei der Adressvalidierung außerdem Statusinformationen über die Qualität jeder Adresse zurückgegeben werden.

Weiterhin haben Sie erfahren, dass man als Benutzer des Administrator Tools den Data Quality Content Installer ausführt, um Adressreferenzdaten zu installieren.

Sie haben ebenfalls gelernt, dass die Adressvalidatorumwandlung eine Multi-Gruppen-Umwandlung ist, und dass Sie die Eingabe- und Ausgabeports für die Umwandlung aus den Portgruppen auswählen. Die ausgewählten Eingabeports legen den Inhalt der Adresse fest, die validiert wird. Die Ausgabeports legen den Inhalt des abschließenden Adressdatensatzes fest.

ANHANG A

Häufig gestellte Fragen (FAQ)

Dieser Anhang umfasst die folgenden Themen:

- [Informatica Analyst - Häufig gestellte Fragen \(FAQ\), 91](#)
- [Informatica Developer - Häufig gestellte Fragen \(FAQ\), 91](#)

Informatica Analyst - Häufig gestellte Fragen (FAQ)

Lesen Sie sich die häufig gestellten Fragen (FAQ) durch. Sie finden dort Antworten auf Fragen, die Sie möglicherweise zu Informatica Analyst haben.

Kann ich ein einziges Benutzerkonto zum Zugriff auf das Administrator-Tool, das Developer-Tool und das Analyst-Tool verwenden?

Ja Sie können einem Benutzer die Berechtigung zum Zugriff auf alle drei Tools geben. Es müssen keine separaten Benutzerkonten für jede Client-Anwendung erstellt werden.

Wo sind meine Referenzdaten gespeichert?

Sie können das Developer-Tool und das Analyst-Tool zum Erstellen und gemeinsamen Nutzen von Referenzdatenobjekten verwenden. Das Modellrepository speichert die Metadaten der Referenzdatenobjekte. In der Referenzdaten-Datenbank werden die Werte der Referenztabelldaten gespeichert. Konfigurieren Sie die Referenzdaten-Datenbank im Content-Management-Dienst.

Informatica Developer - Häufig gestellte Fragen (FAQ)

Lesen Sie sich die häufig gestellten Fragen (FAQ) durch. Sie finden dort Antworten auf Fragen, die Sie möglicherweise zu Informatica Developer haben.

Was ist der Unterschied zwischen einer Quelle und einem Ziel in PowerCenter und einem physischen Datenobjekt im Developer Tool?

In PowerCenter erstellen Sie eine Quelldefinition, die eine Mapping-Quelle einschließt. Sie erstellen eine Zieldefinition, die ein Mapping-Ziel einschließt. Bei einem Neustart wird eventuell wieder eine neue Adresse vom Server zugewiesen. Sie als Mapping-Quelle oder -Ziel verwenden können.

Was ist der Unterschied zwischen einem Mapping im Developer Tool und einem Mapping in PowerCenter?

Ein PowerCenter Mapping gibt an, wie Daten zwischen Quellen und Targets verschoben werden. Ein Developer Tool Mapping gibt an, wie Daten zwischen der Mapping-Eingabe und -Ausgabe verschoben werden.

Ein PowerCenter Mapping muss eine oder mehrere Quelld Definitionen, Quellen-Qualifikatoren und Zieldefinitionen aufweisen. Ein PowerCenter Mapping kann auch Verknüpfungen, Umwandlungen und Mapplets enthalten.

Ein Developer Tool Mapping muss Mapping-Eingabe und -Ausgabe enthalten. Ein Developer Tool Mapping kann auch Umformungen und Mapplets enthalten.

Das Developer Tool beinhaltet folgende Mapping-Typen:

- Mapping, das Daten zwischen Quellen und Zielen verschiebt. Dieser Mapping-Typ unterscheidet sich von einem PowerCenter Mapping nur darin, dass es keine Verknüpfungen verwenden kann und keinen Quellen-Qualifikator verwendet.
- Zuordnung logisches Datenobjekt. Ein Mapping in einem logischen Datenobjektmodell. Eine Zuordnung logisches Datenobjekt kann ein logisches Datenobjekt als Mapping-Eingabe und ein Datenobjekt als Mapping-Ausgabe enthalten. Sie kann auch ein oder mehrere logische Datenobjekte als Mapping-Eingabe und logische Datenobjekte als Mapping-Ausgabe enthalten.
- Virtuelle Tabellenzuordnung. Ein Mapping in einem SQL-Datendienst. Die Zuordnung enthält ein Datenobjekt als Mapping-Eingabe und eine virtuelle Tabelle als Mapping-Ausgabe.
- Virtuelles gespeichertes Prozedur-Mapping. Definiert einen Unternehmenslogiksatz in einem SQL-Datendienst. Es enthält eine Eingabeparameterumwandlung oder ein physisches Datenobjekt als Mapping-Eingabe und eine Ausgabeparameterumwandlung oder physisches Datenobjekt als Mapping-Ausgabe.

Was ist der Unterschied zwischen einem Mapplet in PowerCenter und einem Mapplet im Developer Tool?

Ein Mapplet in PowerCenter und im Developer Tool ist ein wiederverwendbares Objekt, das einen Satz von Umwandlungen enthält. Sie können die Umwandlungslogik in mehreren Mappings nutzen.

Ein PowerCenter-Mapplet kann Quelld Definitionen oder Eingabeumwandlungen als Mapplet-Eingabe enthalten. Es muss Ausgabeumwandlungen als Mapplet-Ausgabe enthalten.

Ein Developer Tool Mapplet kann Datenobjekte oder Eingabeumwandlungen als Mapplet-Eingabe enthalten. Es kann Datenobjekte oder Ausgabeumwandlungen als Mapplet-Ausgabe enthalten. Ein Mapping im Developer Tool kann auch die folgenden Merkmale aufweisen:

- Sie können ein Mapplet als eine Regel validieren. Sie können eine Regel in einem Profil verwenden.
- Ein Mapplet kann andere Mapplets enthalten.

Was ist der Unterschied zwischen einem Mapplet und einer Regel?

Sie können ein Mapplet als eine Regel validieren. Eine Regel ist Geschäftslogik, die Bedingungen definiert, welche auf Quelldaten angewendet werden, wenn ein Profil ausgeführt wird. Sie können ein Mapplet als eine Regel validieren, wenn das Mapplet folgende Anforderungen erfüllt:

- Es enthält eine Eingabe- und Ausgabeumwandlung.
- Das Mapplet enthält keine aktiven Umwandlungen.
- Sie gibt keine Kardinalität zwischen Eingabegruppen an.

INDEX

B

benutzerdefinierte Profile erstellen
Übersicht [29](#)

D

Datenobjekte erstellen
Übersicht [22](#)

E

Einrichten des Analyst Tool
Übersicht [19](#)
Einrichten des Developer Tool
Übersicht [49](#)
Erstellen von Ausdrucksregeln
Übersicht [33](#)

P

physische Datenobjekte importieren
Übersicht [53](#)

Profiling von Daten
Übersicht [62](#)

R

Referenztabellen
Übersicht [46](#)
Referenztabellen aus Spalten erstellen
Übersicht [43](#)

S

Scorecards erstellen
Übersicht [37](#)
Standardprofile erstellen
Übersicht [25](#)