



Informatica®  
10.2.1

# Profiling Getting Started Guide

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.

Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

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.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

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](mailto: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: 2018-06-26

# Inhalt

<b>Einleitung .....</b>	<b>6</b>
Informatica-Ressourcen. ....	6
Informatica-Netzwerk. ....	6
Informatica-Wissensdatenbank. ....	6
Informatica-Dokumentation. ....	7
Informatica-Produktverfügbarkeitsmatrizen. ....	7
Informatica Velocity. ....	7
Informatica Marketplace. ....	7
Globaler Kundensupport von Informatica. ....	7
 <b>Kapitel 1: Erste Schritte - Übersicht.....</b>	 <b>8</b>
Profiling - Übersicht. ....	8
Einführung in Informatica Analyst. ....	9
Informatica Analyst-Benutzeroberfläche. ....	10
Informatica Developer-Benutzeroberfläche. ....	12
Informatica Developer -Begrüßungsseite. ....	13
Spickzettel. ....	13
Informatica Developer-Aufgaben. ....	14
Die Tutorial-Story. ....	14
Die Tutorial-Struktur. ....	14
Voraussetzungen des Tutorials. ....	15
Informatica Analyst-Tutorial. ....	15
Informatica Developer Tool. ....	16
 <b>Teil I: Erste Schritte mit Informatica Analyst.....</b>	 <b>17</b>
 <b>Kapitel 2: Lektion 1. Einrichten von Informatica Analyst.....</b>	 <b>18</b>
Einrichten von Informatica Analyst - Übersicht. ....	18
Aufgabe 1. Anmelden bei Informatica Analyst. ....	19
Aufgabe 2. Erstellen eines Projekts. ....	19
Aufgabe 3. Erstellen eines Ordners. ....	20
Einrichten von Informatica Analyst - Zusammenfassung. ....	20
 <b>Kapitel 3: Lektion 2. Erstellen von Datenobjekten.....</b>	 <b>21</b>
Erstellen von Datenobjekten - Übersicht. ....	21
Aufgabe 1. Erstellen des Einfachdatei-Datenobjekts. ....	22
Aufgabe 2. Anzeigen der Datenobjekteigenschaften. ....	22
Erstellen von Datenobjekten - Zusammenfassung. ....	23

## **Kapitel 4: Lektion 3. Erstellen von Standardprofilen. .... 24**

Erstellen von Standardprofilen – Übersicht. ....	24
Aufgabe 1. Erstellen und Ausführen eines Standardprofils. ....	25
Aufgabe 2. Anzeigen der Profilergebnisse in der Zusammenfassungsansicht. ....	26
Aufgabe 3. Anzeigen der Profilergebnisse in der Detailansicht. ....	26
Erstellen von Standardprofilen – Zusammenfassung. ....	27

## **Kapitel 5: Lektion 4. Erstellen von Kundenprofilen. .... 28**

Erstellen von benutzerdefinierten Profilen - Übersicht. ....	28
Aufgabe 1. Erstellen eines benutzerspezifischen Profils. ....	29
Aufgabe 2. Ausführen des Profils. ....	30
Aufgabe 3. Drilldown der Profilergebnisse. ....	30
Erstellen von benutzerdefinierten Profilen - Zusammenfassung. ....	31

## **Kapitel 6: Lektion 5. Erstellen von Expressionsregeln. .... 32**

Erstellen von Ausdrucksregeln - Übersicht. ....	32
Aufgabe 1. Erstellen von Expressionsregeln und Ausführen des Profils. ....	33
Aufgabe 2. Anzeigen der Expressionsregel-Ausgabe. ....	33
Aufgabe 3. Bearbeiten der Expressionsregeln. ....	34
Erstellen von Expressionsregeln - Zusammenfassung. ....	35

## **Kapitel 7: Lektion 6. Erstellen und Ausführen von Scorecards. .... 36**

Erstellen und Ausführen von Scorecards - Übersicht. ....	36
Aufgabe 1. Erstellen einer Scorecard aus den Profilergebnissen. ....	37
Aufgabe 2. Ausführen der Scorecard. ....	38
Aufgabe 3. Anzeigen der Scorecard. ....	38
Aufgabe 4. Bearbeiten der Scorecard. ....	39
Aufgabe 5. Konfigurieren von Schwellenwerten. ....	40
Aufgabe 6. Anzeigen von Score-Trenddiagrammen. ....	40
Erstellen und Ausführen von Scorecards - Zusammenfassung. ....	40

## **Kapitel 8: Lektion 7. Erstellen von Referenztabelle aus Profilspalten ..... 42**

Erstellen von Referenztabelle aus Profilspalten - Übersicht. ....	42
Aufgabe 1. Erstellen einer Referenztabelle aus Profilspalten. ....	43
Aufgabe 2. Bearbeiten der Referenztabelle. ....	44
Erstellen von Referenztabelle aus Profilspalten - Zusammenfassung. ....	44

## **Kapitel 9: Lektion 8. Erstellen von Referenztabelle. .... 45**

Erstellen von Referenztabelle - Übersicht. ....	45
Aufgabe 1. Erstellen einer Referenztabelle. ....	46
Erstellen von Referenztabelle - Zusammenfassung. ....	46

<b>Teil II: Inbetriebnahme des Informatica Developer.....</b>	<b>47</b>
<b>Kapitel 10: Lektion 1. Einrichten von Informatica Developer. ....</b>	<b>48</b>
Einrichten von Informatica Developer - Übersicht. ....	48
Aufgabe 1 Informatica Developer starten. ....	49
Aufgabe 2 Eine Domäne hinzufügen. ....	49
Aufgabe 3 Ein Model Repository hinzufügen. ....	50
Aufgabe 4 Ein Projekt erstellen. ....	50
Aufgabe 5 Einen Ordner erstellen. ....	50
Aufgabe 6. Auswählen eines Standard-Datenintegrationsdiensts. ....	51
Einrichten von Informatica Developer - Zusammenfassung. ....	51
<b>Kapitel 11: Lektion 2: Importieren von physischen Datenobjekten.....</b>	<b>52</b>
Importieren von physischen Datenobjekten - Übersicht. ....	52
Aufgabe 1 Importieren des Flatfile-Datenobjekts Boston_Customers. ....	53
Aufgabe 2 Import des Einfachdatei-Datenobjekts LA_Customers. ....	59
Aufgabe 3. Importieren des Flatfile-Datenobjekts All_Customers. ....	60
Importieren von physischen Datenobjekten - Zusammenfassung. ....	60
<b>Kapitel 12: Lektion 3. Ausführen eines Profils auf Quelldaten. ....</b>	<b>61</b>
Profiling von Daten - Übersicht. ....	61
Aufgabe 1. Durchführen einer Join-Analyse für zwei Datenquellen. ....	62
Aufgabe 2. Anzeigen der Ergebnisse einer Join-Analyse. ....	63
Aufgabe 3. Ausführen eines Profils für eine Datenquelle. ....	64
Aufgabe 4. Anzeigen der Spalten-Profiling-Ergebnisse. ....	64
Profiling von Daten - Zusammenfassung. ....	65
<b>Index. ....</b>	<b>66</b>

# Einleitung

Die Informatica *Profiling-Kurzanleitung* wurde für Entwickler und Analysten von Datenqualität und Datendiensten verfasst. Sie enthält ein Tutorial, um Einsteigern die Anwendung von Informatica Developer und Informatica Analyst zu erleichtern. Diese Anleitung geht davon aus, dass Ihnen die Datenqualitätskonzepte, Konzepte von Einfachdateien 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](mailto:KB_Feedback@informatica.com).

## Informatica-Dokumentation

Navigieren Sie zur Informatica-Wissensdatenbank unter [https://kb.informatica.com/\\_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx](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](mailto: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](mailto: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:

- [Profiling - Übersicht, 8](#)
- [Einführung in Informatica Analyst, 9](#)
- [Informatica Developer-Benutzeroberfläche, 12](#)
- [Die Tutorial-Story, 14](#)
- [Die Tutorial-Struktur, 14](#)

## Profiling - Übersicht

Verwenden Sie das Profiling, um nach Inhalt, Qualität und Struktur von Datenquellen einer Anwendung, eines Schemas oder eines Unternehmens zu suchen. Der Inhalt der Datenquelle schließt Werthäufigkeiten und Datentypen ein. Die Struktur der Datenquelle schließt Schlüssel und funktionale Abhängigkeiten ein.

Als Teil des Erkennungsprozesses können Sie Profile erstellen und ausführen. Ein Profil ist ein Repository-Objekt, das alle Datenunregelmäßigkeiten über Datenquellen im Unternehmen und versteckte Datenprobleme, die Datenprojekte gefährden können, findet und analysiert. Die Ausführung eines Profils für eine beliebige Datenquelle im Unternehmen vermittelt Ihnen ein gutes Verständnis für die Stärken und Schwächen ihrer Daten und Metadaten.

Mithilfe von Informatica Analyst und Informatica Developer können Sie die Quell- und Metadaten analysieren. Analytiker und Entwickler können diese Tools nutzen, um zusammenzuarbeiten, um Datenqualitätsprobleme zu identifizieren und um Datenbeziehungen zu analysieren. Je nach Ihrer Arbeitsplatzrolle können Sie die Funktionen des Analyst Tools oder des Developer Tools nutzen. Der Profilierungsgrad, den Sie durchführen können, unterscheidet sich je nach verwendetem Tool.

Die folgenden Aufgaben können sowohl im Developer Tool als auch im Analyst Tool ausgeführt werden.

- Durchführen von Spalten-Profiling. Der Prozess schließt das Erkennen der Anzahl an eindeutigen Werten, Null-Werten und Datenmustern in einer Spalte ein.
- Ausführen der Datendomänenerkennung. Sie können kritische Datenmerkmale innerhalb eines Unternehmens erkennen.
- Pflegen Sie Profilergebnisse, einschließlich Datentypen, Datendomänen, Primär- und Fremdschlüssel.
- Erstellen von Scorecards, um die Datenqualität zu überwachen.
- Wählen Sie ein Betriebssystemprofil zum Erstellen und Ausführen von Spaltenprofilen, Enterprise-Erkennungsprofilen und Scorecards basierend auf den Berechtigungen des im Betriebssystemprofil angegebenen Benutzers aus.



- Verwenden Sie Repository-Objektsperren, damit andere Benutzer keine Arbeiten überschreiben können.
- Verwenden Sie das Versionsverwaltungssystem, um mehrere Versionen eines Profils zu speichern.
- Erstellen und Zuweisen von Tags zu Datenobjekten.
- Schlagen Sie die Bedeutung von Objektnamen als Geschäftsbegriffe in Business Glossary Desktop nach. Beispielsweise können Sie die Bedeutung eines Spaltennamens oder eines Profilnamens nachschlagen, um dessen geschäftliche Anforderungen und gegenwärtige Implementierung zu verstehen.

Die folgenden Aufgaben können im Developer Tool ausgeführt werden.

- Erkennen des Grads der potenziellen Joins zwischen zwei Datenspalten in einer Datenquelle.
- Bestimmen des Prozentsatzes sich überschneidender Daten zwischen Spaltenpaaren einer Datenquelle oder mehreren Datenquellen.
- Vergleichen der Ergebnisse des Spalten-Profiling.
- Generieren eines Mapping-Objekts aus einem Profil.
- Erkennen von Primärschlüsseln in einer Datenquelle.
- Erkennen von Fremdschlüsseln in einem Satz von einer oder mehreren Datenquellen.
- Erkennen von funktionalen Abhängigkeiten zwischen Spalten in einer Datenquelle.
- Ausführen von Datenerkennungs-Tasks über eine große Anzahl von Datenquellen über mehrere Verbindungen hinweg. Zu den Datenerkennungs-Tasks gehören Spaltenprofil, Rückschluss von Primärschlüssel- und Fremdschlüssel-Beziehungen, Datendomänenerkennung und Generierung einer konsolidierten grafischen Übersicht über die Beziehungen zwischen den Daten.

Die folgenden Aufgaben können im Analyst Tool ausgeführt werden:

- Ausführen der Enterprise-Erkennung auf vielen Datenquellen über mehrere Verbindungen hinweg. Sie können eine konsolidierte Zusammenfassung für die Erkennungsergebnisse der Spalte Metadaten und Datendomänen anzeigen.
- Ausführen der Erkennungssuche, um zu ermitteln, wo sich die Daten und Metadaten im Unternehmen befinden. Sie können nach bestimmten Objekten wie Datenobjekten, Regeln und Profilen suchen. Bei der Erkennungssuche werden Objekte gesucht und Beziehungen zu anderen Objekten in den Datenbanken und Schemata des Unternehmens ermitteln.
- Anzeigen der Profilergebnisse für eine historische Profilausführung.
- Vergleichen Sie die Profilergebnisse für zwei Profilausführungen in einem Spaltenprofil.
- Die Scorecard-Herkunft kann für jede Scorecard-Metrik oder -Metrikgruppe angezeigt werden.
- Zeigen Sie das Scorecard-Dashboard an.
- Hinzufügen von Kommentaren zu einem Profil oder zu Spalten in einem Profil.
- Zuweisen von Tags zu einem Profil oder zu Spalten in einem Profil.
- Zuweisen von Geschäftsbegriffen zu Spalten in einem Profil.

## 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.

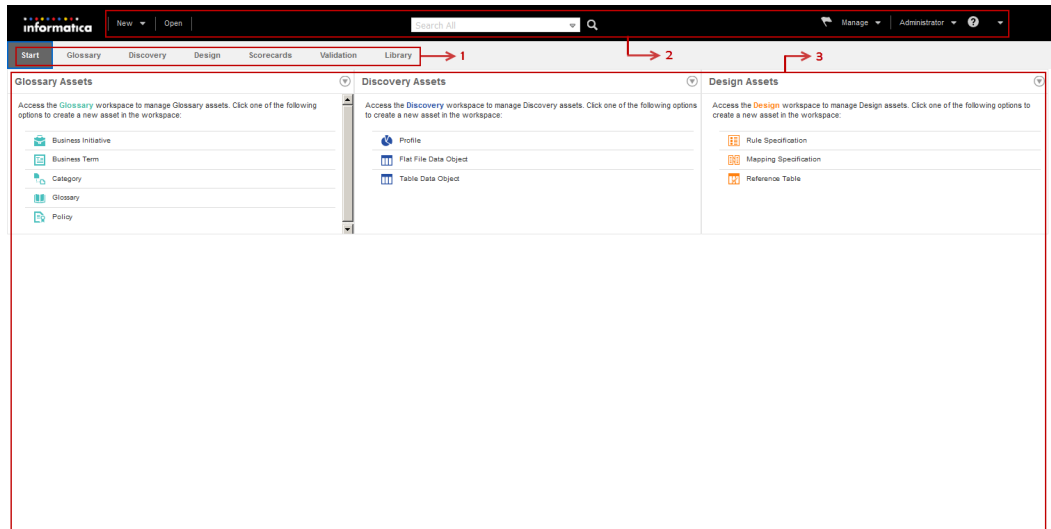
Je nach Lizenz verwenden Wirtschaftsanalysten und Entwickler das Analyst-Tool zur datenorientierten Zusammenarbeit. Sie können Spalten- und Regel-Profiling durchführen, mit Scorecards arbeiten und

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

## Informatica Analyst-Benutzeroberfläche

Die Benutzeroberfläche des Analyst Tools besteht aus mehreren Arbeitsbereichen, die Sie zum Erstellen von Datenintegrations- und Datenqualitäts-Datenverbundlösungen verwenden können. Ein Arbeitsbereich ist eine Webseite, auf die Sie basierend auf der Lizenz und Berechtigung zugreifen können.

Die folgende Abbildung zeigt die Benutzeroberfläche des Analyst Tools:



1. Zugriffsbereich des Arbeitsbereichs
2. Kopfzeilenbereich
3. Arbeitsbereichsregisterkarten

Das Analyst Tool wird im Arbeitsbereich **Starten** geöffnet. Der Arbeitsbereich **Starten** listet die Arbeitsbereiche auf, für die Sie über die Lizenz und Berechtigung zur Verwendung über Zugriffsbereiche des Arbeitsbereichs verfügen.

Das Analyst Tool zeigt die folgenden Header-Einträge an:

### Neu

Erstellen Sie Assets in den Arbeitsbereichen **Glossar**, **Erkennung** und **Design**.

### Öffnen

Öffnen Sie den Arbeitsbereich **Bibliothek**.

### Benachrichtigungsalarm

Zeigen Sie Benachrichtigungen für Glossar-Assets an.

### Verwalten

Öffnen Sie temporäre Arbeitsbereiche und **Benachrichtigungen**. Sie können die Arbeitsbereiche **Verbindung**, **Datendomänen**, **Jobstatus**, **Projekt** und **Business Glossary-Sicherheit** öffnen.

### Benutzername

Geben Sie Benutzereinstellungen zum Ändern des Passworts und zum Abmelden aus dem Analyst Tool ein.

## Hilfe

Greifen Sie im aktuellen Arbeitsbereich auf die Hilfe zu.

Das Analyst Tool enthält die folgenden permanenten Arbeitsbereiche:

## Start

Greifen Sie auf andere Arbeitsbereiche zu, für die Sie über die Lizenz zum Zugriff auf die Zugriffsbereiche des Arbeitsbereichs verfügen. Wenn Sie über die Lizenz zum Durchführen der Ausnahmeverwaltung verfügen, werden Ihre Aufgaben im Bereich **Meine Aufgaben** des Arbeitsbereichs angezeigt.

## Glossar

Definieren und beschreiben Sie Geschäftskonzepte, die für Ihr Unternehmen eine wichtige Rolle spielen. Sie können Geschäftsbegriffe, Kategorien, Glossare und Richtlinien erstellen sowie verwalten.

## Erkennung

Analysieren Sie die Qualität von Daten und Metadaten in Quellsystemen. Sie können Profile, Einfachdateidatenobjekte und Tabellendatenobjekte erstellen und verwalten.

## Design

Entwerfen Sie die Geschäftslogik für die Zusammenarbeit von Analysten und Entwicklern. Sie können Mapping-Spezifikationen, Referenztabellen und Regeldefinitionen erstellen sowie verwalten.

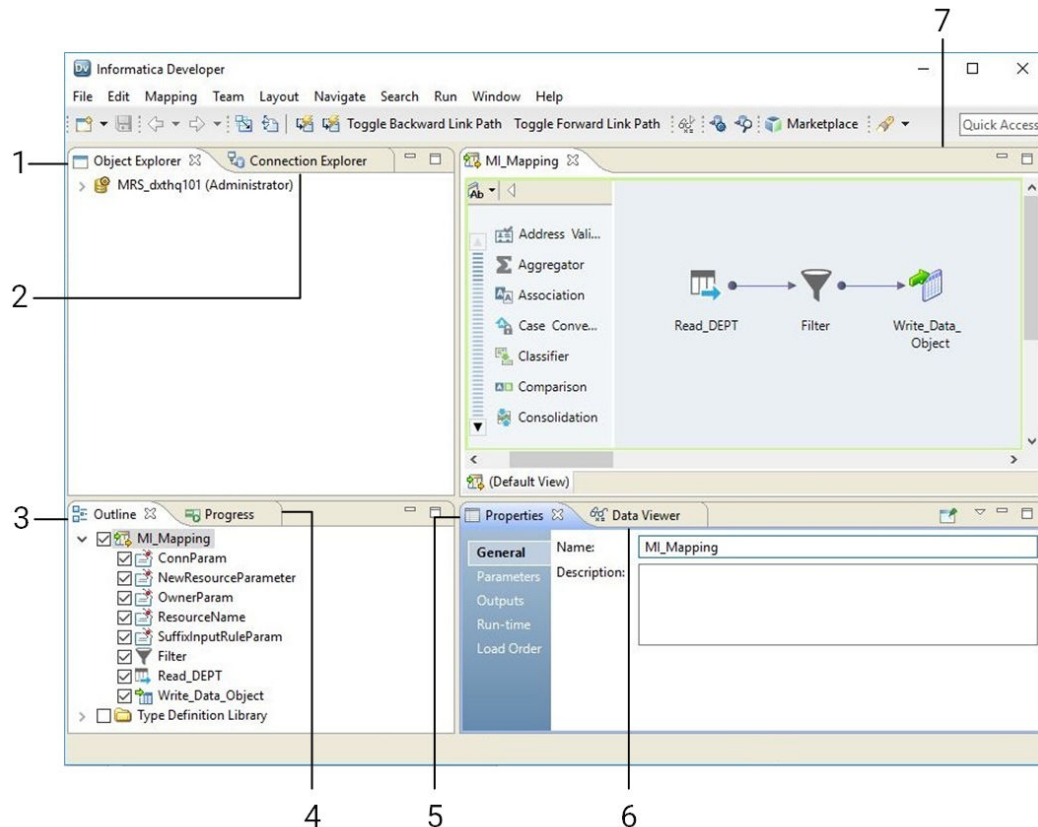
## Scorecards

Öffnen, bearbeiten und führen Sie Scorecards aus, die Sie aus Profilergebnissen erstellt haben. Sie können Metriken hinzufügen, Drilldown für Spalten durchführen, Scorecard-Filter hinzufügen und Trenddiagramme für eine Scorecard anzeigen.

# Informatica Developer-Benutzeroberfläche

Die Benutzeroberfläche des Developer-Tools besteht aus einer Workbench mit mehreren Ansichten, die zur Erstellung der Data Integration und der Data Quality verwendet wird. Datenföderation Lösungen.

Das folgende Bild zeigt die Benutzeroberfläche des Developer-Tools:



1. Ansicht „Objekt-Explorer“
2. Ansicht „Outline“
3. Ansicht „Eigenschaften“
4. Ansicht „Daten-Viewer“
5. Ansicht „Tags“
6. Ansicht „Objektabhängigkeiten“
7. Ansicht „Alarmer“
8. Ansicht „Verbindungs-Explorer“
9. Editor

Das Developer-Tool kann die folgenden Ansichten anzeigen:

## **Object Explorer view**

Displays projects, folders, and the objects within the projects and folders.

## **Connection Explorer view**

Displays connections to relational databases.

## **Outline view**

Displays objects that are dependent on an object selected in the **Object Explorer** view.

#### Cheat Sheets view

Displays the cheat sheet that you open. To open a cheat sheet, click **Help > Cheat Sheets** and select a cheat sheet.

#### Data Viewer view

Displays source data, profile results, and previews the output of a transformation. You can also preview web service messages or run an SQL query from the Data Viewer view.

#### Object Dependencies view

Displays object dependencies when you view, modify, or delete an object.

#### Alerts view

Displays connection status alerts.

#### Validation Log view

Displays object validation errors.

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](https://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**.

## Informatica Developer-Aufgaben

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:

- Ausführen eines Profils auf Quelldaten. 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.
- Erstellen und Durchführen von Datenqualitätsregeln. Informatica bietet vorgefertigte Regeln, die Sie ausführen oder bearbeiten können, um Ihre Projektziele zu erreichen. Sie können im Developer Tool Regeln erstellen.
- Zusammenarbeiten mit Informatica-Anwendern. Die Regeln, die Sie zum Modellrepository hinzufügen, sind für Benutzer im Developer Tool und Analyst Tool verfügbar. Benutzer können in Projekten zusammenarbeiten, und in unterschiedlichen Stadien eines Projekts können unterschiedliche Benutzer den Eigentümerstatus von Objekten übernehmen.

## Die Tutorial-Story

HypoStores Corporation ist eine nationale Einzelhandelsorganisation mit Hauptquartieren in Boston und Geschäften in mehreren Bundesstaaten. Sie integriert regelmäßig Betriebsdaten aus den landesweiten Geschäften in den Datenspeicher im Hauptquartier. Vor kurzem wurde ein Geschäft in Los Angeles eröffnet.

Zum Hauptquartier gehört ein zentrales ICC-Team von Administratoren, Entwicklern und Architekten, die dafür verantwortlich sind, eine gemeinsame Datendienst-Ebene 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 möchte die Daten aus Boston und Los Angeles profilieren, bevor die Datensätze integriert werden. Die Profiloperationen identifizieren Datenqualitätsprobleme, die HypoStores vor der Integration beheben kann.

## 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.

## Voraussetzungen des Tutorials

Bevor Sie mit den Lektionen des Tutorials beginnen, muss die Informatica-Domäne mit mindestens einem eingerichteten Knoten laufen. Sie müssen sicherstellen, dass der Analyst-Dienst, der Modellrepository-Dienst und der Datenintegrationsdienst ausgeführt werden.

Das Installationsprogramm 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
- Customer\_Order.xsd
- LA\_customers.csv
- orders.csv

## Informatica Analyst-Tutorial

In diesem Tutorial meldet sich ein Analyst beim Analyst-Tool an und erstellt Projekte und Ordner, Profile und Regeln, Scorecards für Daten und Referenztabellen.

In der folgenden Tabelle werden die Lektionen beschrieben, die Sie durchführen können:

Lektion	Beschreibung
Lektion 1. Einrichten von Informatica Analyst	Anmelden am Analysetool und Erstellen eines Projekts und Ordners für die Tutoriallektionen.
Lektion 2. Erstellen von Datenobjekten	Importieren einer Einfachdatei als Datenobjekt und Vorschau der Daten.
Lektion 3. Erstellen von Schnellprofilen	Erstellen eines Schnellprofils, um schnell eine Vorstellung von Data Quality zu erhalten.
Lektion 4. Erstellen von benutzerdefinierten Profilen	Erstellen Sie ein benutzerdefiniertes Profil zum Konfigurieren von Spalten sowie für Stichproben- und Drilldown-Optionen.
Lektion 5. Erstellen von Ausdrucksregeln	Erstellen von Expressionregeln zum Modifizieren und Profilieren von Spaltenwerten
Lektion 6. Erstellen und Ausführen von Scorecards	Erstellen und Ausführen einer Scorecard, um den Fortschritt der Datenqualität im Zeitverlauf zu messen.

Lektion	Beschreibung
Lektion 7. Erstellen von Referenztabellen aus Profilergebnissen	Erstellen einer Referenztable, die Sie zum Standardisieren von Quelldaten verwenden können.
Lektion 8. Erstellen von Referenztabellen	Erstellen einer Referenztable, um Beziehungen zwischen Quelldaten sowie gültigen und Standard-Werten einzurichten.

## Informatica Developer Tool

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

Sie können das Developer Tool zum Erstellen und Ausführen von Profilen verwenden, die den Inhalt und die Struktur der Daten analysieren.

In der folgenden Tabelle werden die Lektionen beschrieben, die Sie durchführen können:

Lektion	Beschreibung
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. Ausführen eines Profils auf Quelldaten	Wenn Sie ein Profil ausführen, machen die Profilergebnisse den Inhalt und die Struktur Ihrer Daten erkennbar. Sie können in einem Profil eine Join-Analyse durchführen, eine Form der Analyse zur Bestimmung, ob ein gültiger Join zwischen zwei Datenspalten möglich ist.



# Teil I: Erste Schritte mit Informatica Analyst

Dieser Teil enthält die folgenden Kapitel:

- [Lektion 1. Einrichten von Informatica Analyst, 18](#)
- [Lektion 2. Erstellen von Datenobjekten, 21](#)
- [Lektion 3. Erstellen von Standardprofilen, 24](#)
- [Lektion 4. Erstellen von Kundenprofilen, 28](#)
- [Lektion 5. Erstellen von Expressionsregeln, 32](#)
- [Lektion 6. Erstellen und Ausführen von Scorecards, 36](#)
- [Lektion 7. Erstellen von Referenztabelle aus Profilspalten , 42](#)
- [Lektion 8. Erstellen von Referenztabelle, 45](#)

## KAPITEL 2

# Lektion 1. Einrichten von Informatica Analyst

Dieses Kapitel umfasst die folgenden Themen:

- [Einrichten von Informatica Analyst - Übersicht, 18](#)
- [Aufgabe 1. Anmelden bei Informatica Analyst, 19](#)
- [Aufgabe 2. Erstellen eines Projekts, 19](#)
- [Aufgabe 3. Erstellen eines Ordners, 20](#)
- [Einrichten von Informatica Analyst - Zusammenfassung, 20](#)

## 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**.  
Das Fenster **Neuer Ordner** wird geöffnet.
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, 21](#)
- [Aufgabe 1. Erstellen des Einfachdatei-Datenobjekts, 22](#)
- [Aufgabe 2. Anzeigen der Datenobjekteigenschaften, 22](#)
- [Erstellen von Datenobjekten - Zusammenfassung, 23](#)

### 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 des Einfachdatei-Datenobjekts

In dieser Aufgabe verwenden Sie den Assistenten **Hinzufügen einer Einfachdatei**, um Einfachdatei-Datenobjekte aus der Datendatei LA\_Customers zu erstellen.

1. Wählen Sie im Navigator **Projekte** den Ordner „Kunden“ in Ihrem Tutorialprojekt aus.  
**Hinweis:** Sie müssen das Projekt oder den Ordner auswählen, in dem Sie das Einfachdatei-Datenobjekt erstellen möchten, bevor Sie es erstellen können.
2. Klicken Sie auf **Neu > Einfachdatei-Datenobjekt**.  
Der Assistent **Einfachdatei hinzufügen** wird im Arbeitsbereich **Discovery** angezeigt.
3. Wählen Sie **Durchsuchen und hochladen** aus, und klicken Sie dann auf **Durchsuchen**.
4. Wechseln Sie zum Speicherort der Datei LA\_Customers.csv und klicken Sie auf **Öffnen**.
5. Klicken Sie auf **Weiter**.  
Auf der Seite **Importtyp wählen** werden die Optionen **Mit Trennzeichen** und **Feste Breite** angezeigt. Die Standardoption ist **Mit Trennzeichen**.
6. Klicken Sie auf **Weiter**.
7. Wählen Sie unter **Geben Sie die in Ihren Dateien verwendeten Delimiter und Textqualifikatoren an** die Option **Doppelte Anführungszeichen** als Textqualifikator.
8. Wählen Sie unter **Zeilen für den Import angeben** die Option **Spaltennamen aus erster Zeile importieren**, um Spaltennamen aus der ersten Zeile zu importieren, die nicht leer ist.
9. Klicken Sie auf **Weiter**.  
Im Abschnitt **Spaltenattribute** werden Name, Datentyp, Gesamtstellenanzahl, Größenordnung und Format für jede Spalte angezeigt.
10. Klicken Sie auf **Weiter**.  
Im Feld **Name** wird **LA\_Customers** angezeigt.
11. Optional können Sie den Namen der Datei ändern und eine Beschreibung hinzufügen.
12. Wählen Sie im Abschnitt **Ordner** den Ordner **Kunden**.
13. Klicken Sie auf **Fertig stellen**.  
Das Datenobjekt „LA\_Customers“ wird im Arbeitsbereich **Discovery** mit der Datenvorschau der Quelldaten angezeigt.

# 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, 24](#)
- [Aufgabe 1. Erstellen und Ausführen eines Standardprofils, 25](#)
- [Aufgabe 2. Anzeigen der Profilergebnisse in der Zusammenfassungsansicht, 26](#)
- [Aufgabe 3. Anzeigen der Profilergebnisse in der Detailansicht, 26](#)
- [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.

## Aufgabe 3. Anzeigen der Profilergebnisse in der Detailansicht

In dieser Aufgabe verwenden Sie die Detailansicht für das Profil LA\_Customers, um einen raschen Überblick über eine Spalte in den Profilergebnissen zu erhalten.

1. Klicken Sie auf einen Spaltennamen in der Zusammenfassungsansicht, um die Profilergebnisse für die Spalte in der Detailansicht anzuzeigen.  
Die Profilergebnisse für die Spalte wird in der Detailansicht angezeigt.
2. Im Bereich **Allgemein** können Sie die Nullwerte, die eindeutigen Werte und die nicht eindeutigen Werte als horizontales Balkendiagramm und als Prozentsätze anzeigen.

3. Im Bereich **Werte** können Sie alle Werte im oberen Bereich anzeigen. Wenn Sie den Schieberegler über die Werte im oberen Bereich ziehen, werden alle Werte innerhalb des Schiebereglers im unteren Bereich angezeigt.
4. Klicken Sie im Bereich **Datenvorschau** auf **Aktionen > Vorschau anzeigen**, um die Quelldaten anzuzeigen.
5. Im Bereich **Muster** können Sie die abgeleiteten Muster in der Spalte als horizontales Balkendiagramm und als Prozentsätze anzeigen.
6. Im Bereich **Datentyp** können Sie die abgeleiteten Datentypen in der Spalte als horizontales Balkendiagramm und als Prozentsätze anzeigen.
7. Im Bereich **Datendomäne** können Sie die abgeleiteten Datendomänen anzeigen.

## 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 Referenztabelle, 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, 28](#)
- [Aufgabe 1. Erstellen eines benutzerspezifischen Profils, 29](#)
- [Aufgabe 2. Ausführen des Profils, 30](#)
- [Aufgabe 3. Drilldown der Profilergebnisse, 30](#)
- [Erstellen von benutzerdefinierten Profilen - Zusammenfassung, 31](#)

## 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

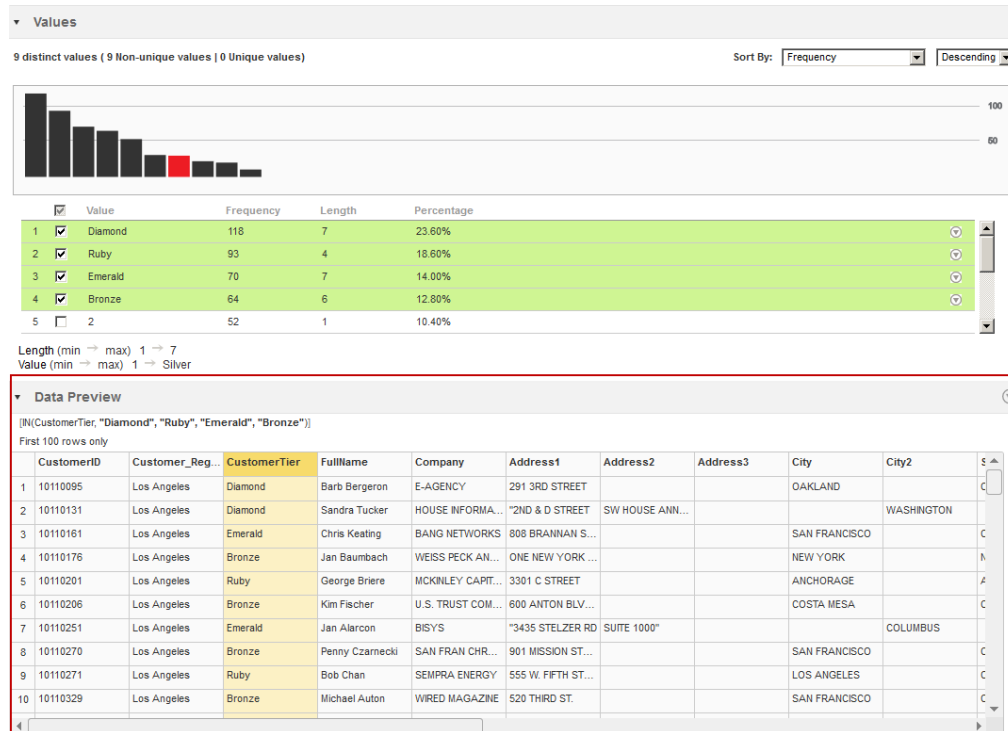
In dieser Aufgabe führen Sie ein Drilldown mit den Werten der Spalte CustomerTier durch, um die Quellzeilen in dem Datenobjekt für das Profil anzuzeigen.

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, 32](#)
- [Aufgabe 1. Erstellen von Expressionsregeln und Ausführen des Profils, 33](#)
- [Aufgabe 2. Anzeigen der Expressionsregel-Ausgabe, 33](#)
- [Aufgabe 3. Bearbeiten der Expressionsregeln, 34](#)
- [Erstellen von Expressionsregeln - Zusammenfassung, 35](#)

### 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, 36](#)
- [Aufgabe 1. Erstellen einer Scorecard aus den Profilergebnissen, 37](#)
- [Aufgabe 2. Ausführen der Scorecard, 38](#)
- [Aufgabe 3. Anzeigen der Scorecard, 38](#)
- [Aufgabe 4. Bearbeiten der Scorecard, 39](#)
- [Aufgabe 5. Konfigurieren von Schwellenwerten, 40](#)
- [Aufgabe 6. Anzeigen von Score-Trenddiagrammen, 40](#)
- [Erstellen und Ausführen von Scorecards - Zusammenfassung, 40](#)

## 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, 42](#)
- [Aufgabe 1. Erstellen einer Referenztable aus Profilspalten, 43](#)
- [Aufgabe 2. Bearbeiten der Referenztable, 44](#)
- [Erstellen von Referenztabellen aus Profilspalten - Zusammenfassung, 44](#)

## 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, 45](#)
- [Aufgabe 1. Erstellen einer Referenztable, 46](#)
- [Erstellen von Referenztabellen - Zusammenfassung, 46](#)

### 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 Referenztabelle

In dieser Aufgabe erstellen Sie die Referenztabelle Reftab\_CustomerTier\_Codes, um die gültigen Werte für die CustomerTier-Daten zu standardisieren.

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

Der Assistent **Neue Referenztabelle** wird eingeblendet.

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

3. Klicken Sie auf **Weiter**.

4. Klicken Sie für jede Spalte, die Sie in der Referenztabelle 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 Referenztabelle wird im Arbeitsbereich **Design** angezeigt.

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

CustomerID = LA1, LA2, LA3, LA4

CustomerTier = 1, 2, 3, 4.

Status = Aktiv, Inaktiv

## Erstellen von Referenztabellen - Zusammenfassung

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

Sie haben eine Referenztabelle unter Verwendung des Referenztabellen-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, 48](#)
- [Lektion 2: Importieren von physischen Datenobjekten., 52](#)
- [Lektion 3. Ausführen eines Profils auf Quelldaten, 61](#)

# KAPITEL 10

## Lektion 1. Einrichten von Informatica Developer

Dieses Kapitel umfasst die folgenden Themen:

- [Einrichten von Informatica Developer - Übersicht, 48](#)
- [Aufgabe 1 Informatica Developer starten, 49](#)
- [Aufgabe 2 Eine Domäne hinzufügen, 49](#)
- [Aufgabe 3 Ein Model Repository hinzufügen, 50](#)
- [Aufgabe 4 Ein Projekt erstellen, 50](#)
- [Aufgabe 5 Einen Ordner erstellen, 50](#)
- [Aufgabe 6. Auswählen eines Standard-Datenintegrationsdiensts, 51](#)
- [Einrichten von Informatica Developer - Zusammenfassung, 51](#)

## 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. Sie müssen auch einen Standard-Datenintegrationsdienst auswählen, wenn die Domäne mehr als einen Dienst enthält.

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 erhalten diese Informationen von einem Domänenadministrator.
- 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. Auswählen eines Standard-Datenintegrationsdiensts

In dieser Aufgabe wählen Sie einen Standard-Datenintegrationsdienst aus, damit Sie Mappings ausführen und die Vorschau von Daten anzeigen können. Dieser Schritt ist erforderlich, wenn in der Domäne mehr als ein Datenintegrationsdienst vorhanden ist. Wenn die Domäne einen Datenintegrationsdienst enthält, wird dieser Dienst als Standard festgelegt.

1. Klicken Sie auf **Fenster > Einstellungen**.  
Das Dialogfeld **Einstellungen** wird eingeblendet
2. Wählen Sie **Informatica > Datenintegrationsdienste**.
3. Erweitern Sie die Domäne.
4. Wählen Sie einen Datenintegrationsdienst.
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, 52](#)
- [Aufgabe 1 Importieren des Flatfile-Datenobjekts Boston\\_Customers, 53](#)
- [Aufgabe 2 Import des Einfachdatei-Datenobjekts LA\\_Customers, 59](#)
- [Aufgabe 3. Importieren des Flatfile-Datenobjekts All\\_Customers, 60](#)
- [Importieren von physischen Datenobjekten - Zusammenfassung, 60](#)

## 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.

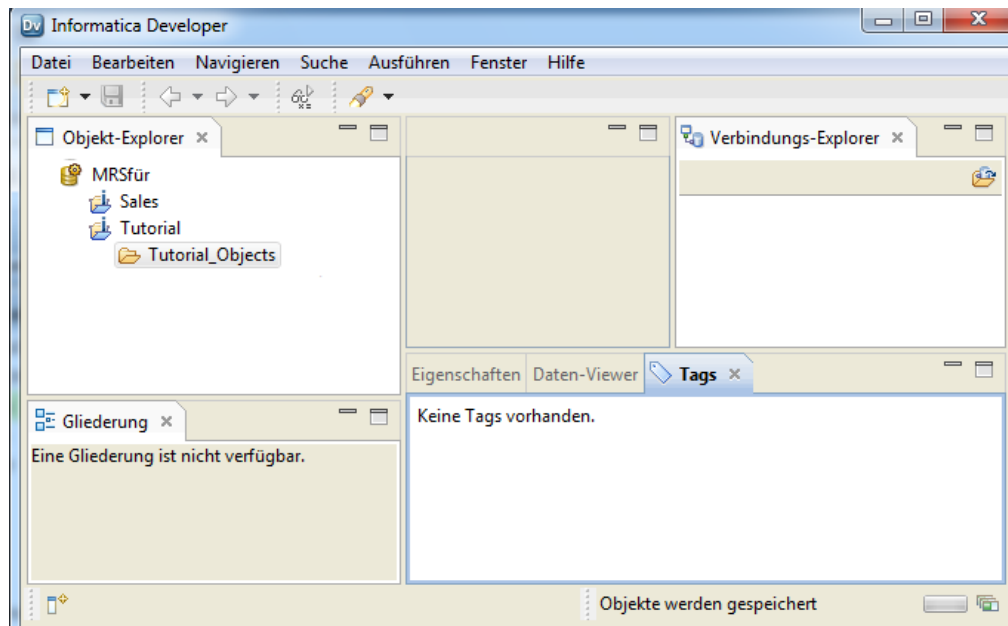
### Zeitplan

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

# Aufgabe 1 Importieren des Flatfile-Datenobjekts Boston\_Customers

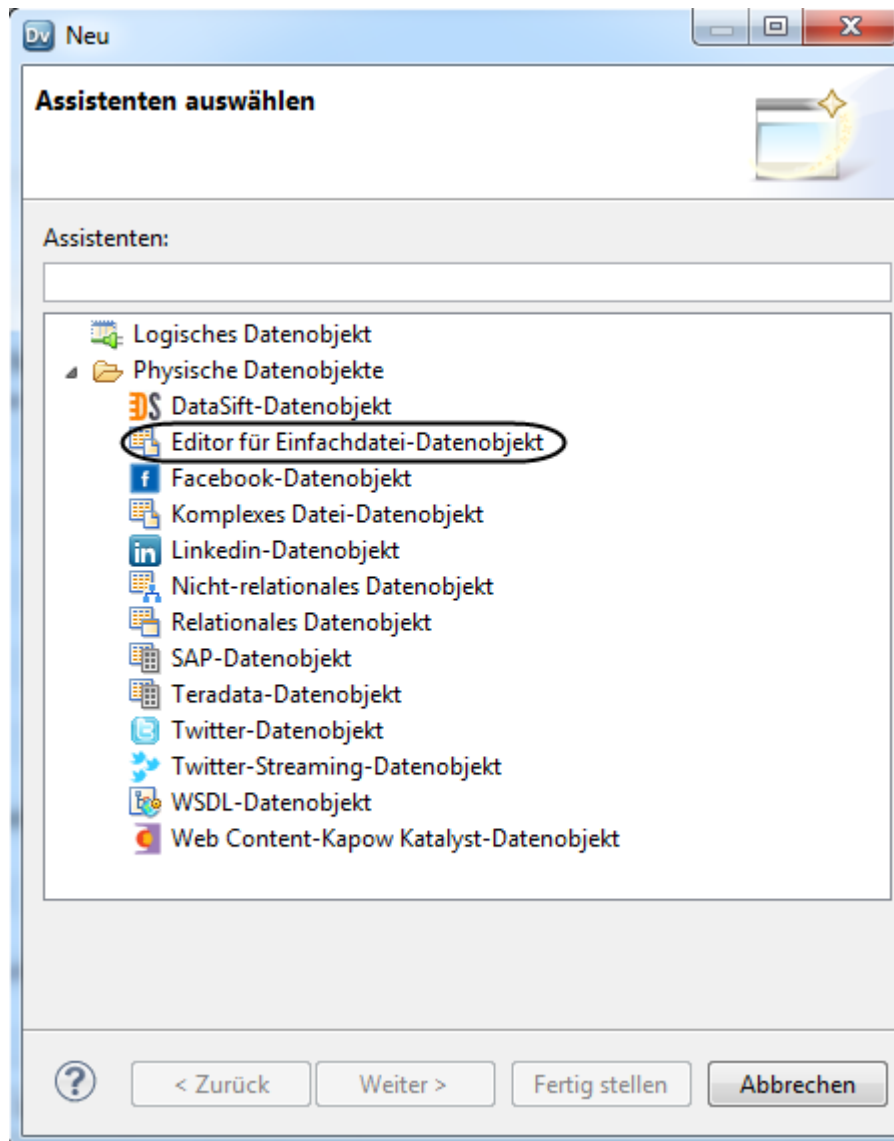
In dieser Aufgabe importieren Sie ein physisches Datenobjekt aus einer Datei, die Kundendaten aus dem Büro in Boston enthält.

1. Wählen Sie in der Ansicht **Objekt-Explorer** den Ordner Tutorial\_Objects.



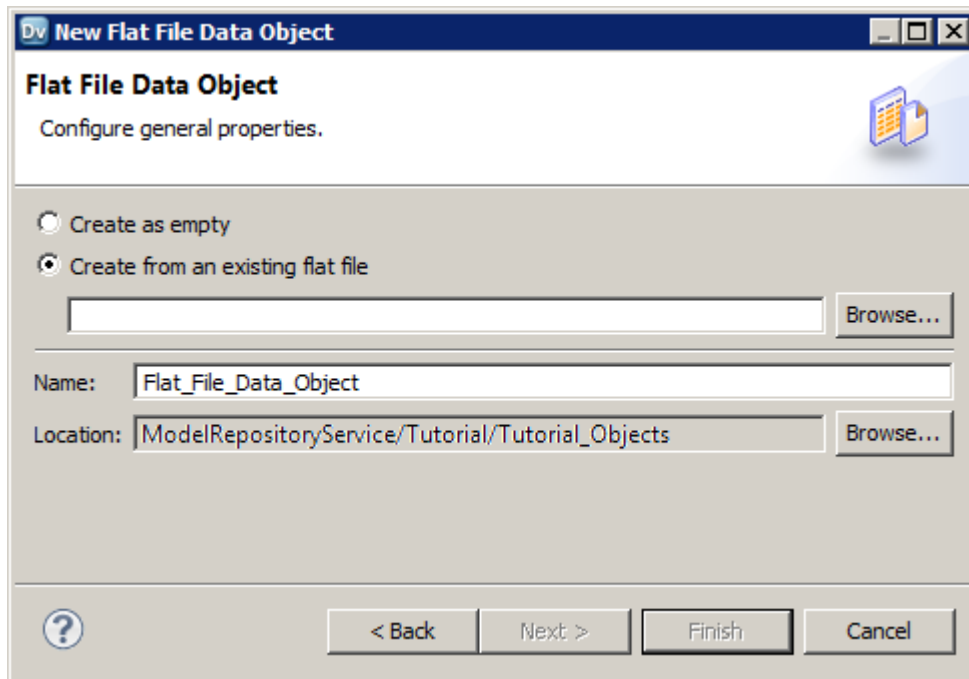
2. Klicken Sie mit der rechten Maustaste auf den Ordner Tutorial\_Objects, und wählen Sie **Neu > Datenobjekt**.

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 die Codepage und das 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	Last Name
2	10110102	Boston	Bronze	Bronze
3	10110105	Boston		Anthony
4	10110106	Boston	Emerald	Bonnie
5	10110107	Boston	Ruby	Cheryl
6	10110109	Boston	Gold	Acacia

? < 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:   
 Zeilen-Delimiter:     ☐ Aufeinander folgende Delimiter wie eins behandeln  
 Escape-Zeichen:     ☐ Escape-Zeichen in Daten beibehalten

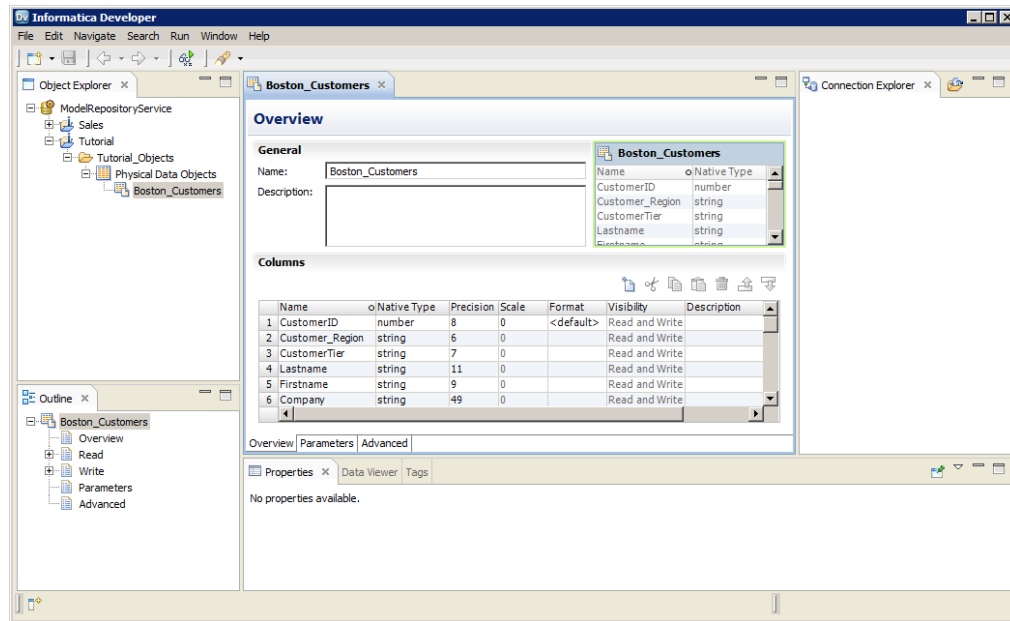
Max. Anzahl Zeilen in der Vorschau:

	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
<b>Runtime : Read</b>	
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, 61](#)
- [Aufgabe 1. Durchführen einer Join-Analyse für zwei Datenquellen, 62](#)
- [Aufgabe 2. Anzeigen der Ergebnisse einer Join-Analyse, 63](#)
- [Aufgabe 3. Ausführen eines Profils für eine Datenquelle, 64](#)
- [Aufgabe 4. Anzeigen der Spalten-Profiling-Ergebnisse, 64](#)
- [Profiling von Daten - Zusammenfassung, 65](#)

### 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.

# INDEX

## B

benutzerdefinierte Profile erstellen  
Übersicht [28](#)

## D

Datenobjekte erstellen  
Übersicht [21](#)

## E

Einrichten des Analyst Tool  
Übersicht [18](#)  
Einrichten des Developer Tool  
Übersicht [48](#)  
Erstellen von Ausdrucksregeln  
Übersicht [32](#)

## P

physische Datenobjekte importieren  
Übersicht [52](#)

Profiling  
Übersicht [8](#)  
Profiling von Daten  
Übersicht [61](#)

## R

Referenztabellen  
Übersicht [45](#)  
Referenztabellen aus Spalten erstellen  
Übersicht [42](#)

## S

Scorecards erstellen  
Übersicht [36](#)  
Standardprofile erstellen  
Übersicht [24](#)