This Software is protected by U.S. Patent Numbers 5,794,246; 6,014,670; 6,016,501; 6,029,178; 6,032,158; 6,035,307; 6,044,374; 6,092,086; 6,208,990; 6,339,775; 6,640,226; 6,789,096; 6,820,077; 6,823,373; 6,850,947; 6,895,471; 7,117,215; 7,162,643; 7,254,590; 7,281,001; 7,421,458; 7,496,588; 7,523,121; 7,584,422; 7,720,842; 7,721,270; and 7,774,791, international Patents and other Patents Pending.

DISCLAIMER: Informatica Corporation provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of non-infringement, merchantability, or use for a particular purpose. Informatica Corporation does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.

2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

Part Number: IN-PGS-91000-0001
Table of Contents

Preface ........................................................................... iv
Informatica Resources. ................................................... iv
  Informatica Customer Portal. ........................................... iv
  Informatica Documentation. ............................................ iv
  Informatica Web Site. .................................................... iv
  Informatica How-To Library. ........................................... iv
  Informatica Knowledge Base. ......................................... v
  Informatica Multimedia Knowledge Base. ....................... v
  Informatica Global Customer Support. ............................ v

Chapter 1: Getting Started Overview. ............................... 1
Informatica Domain Overview. ........................................ 1
  Feature Availability. ..................................................... 3
Introducing Informatica Analyst. ....................................... 4
Informatica Developer Overview. .................................... 5
  Informatica Developer Welcome Page. ............................ 5
  Cheat Sheets. ............................................................. 6
  Data Quality and Data Explorer Advanced Edition. ............. 6
The Tutorial Story. ........................................................ 7
The Tutorial Structure. .................................................. 7
  Tutorial Prerequisites. .................................................. 7
Informatica Analyst Tutorial. ........................................... 8
Informatica Developer Tool (Data Quality). ......................... 8

Part I: Getting Started with Informatica Analyst. .................. 10

Setting Up Informatica Analyst Overview. ........................ 11
Task 1. Log In to Informatica Analyst. .............................. 12
Task 2. Create a Project. ............................................... 12
Task 3. Create a Folder. ............................................... 12
Setting Up Informatica Analyst Summary. ......................... 13

Chapter 3: Lesson 2. Creating Data Objects. ....................... 14
Creating Data Objects Overview. .................................... 14
Task 1. Create the Flat File Data Objects. ......................... 15
Task 2. Preview the Data. ............................................. 15
Creating Data Objects Summary. .................................... 16
Chapter 4: Lesson 3. Creating Quick Profiles. ......................................................... 17
Creating Quick Profiles Overview. ................................................................. 17
Task 1. Create and Run a Quick Profile. ......................................................... 18
Task 2. View the Profile Results. ................................................................. 18
Creating Quick Profiles Summary. ............................................................... 19

Chapter 5: Lesson 4. Creating Custom Profiles. .................................................. 20
Creating Custom Profiles Overview. ............................................................... 20
Task 1. Create a Custom Profile. ................................................................. 21
Task 2. Run the Profile. ................................................................. 21
Task 3. Drill Down on Profile Results. .......................................................... 22
Creating Custom Profiles Summary. ............................................................... 22

Chapter 6: Lesson 5. Creating Expression Rules. .................................................. 23
Creating Expression Rules Overview. ............................................................... 23
Task 1. Create Expression Rules and Run the Profile. ...................................... 24
Task 2. View the Expression Rule Output. ..................................................... 24
Task 3. Edit the Expression Rules. ................................................................. 25
Creating Expression Rules Summary. ............................................................... 25

Chapter 7: Lesson 6. Creating and Running Scorecards. ....................................... 26
Creating and Running Scorecards Overview. ................................................... 26
Task 1. Create a Scorecard from the Profile Results. ....................................... 27
Task 2. Run the Scorecard. ................................................................. 28
Task 3. View the Scorecard. ................................................................. 28
Task 4. Edit the Scorecard. ................................................................. 28
Task 5. Configure Thresholds. ................................................................. 29
Task 6. View Score Trend Charts. ................................................................. 29
Creating and Running Scorecards Summary. ................................................... 29

Chapter 8: Lesson 7. Creating Reference Tables from Profile Columns. .................. 30
Creating Reference Tables from Profile Columns Overview. ................................ 30
Task 1. Create a Reference Table from Profile Columns. .................................. 31
Task 2. Edit the Reference Table. ................................................................. 32
Creating Reference Tables from Profile Columns Summary. ................................ 32

Chapter 9: Lesson 8. Creating Reference Tables. .................................................. 33
Creating Reference Tables Overview. ............................................................... 33
Task 1. Create a Reference Table. ................................................................. 34
Creating Reference Tables Summary. ............................................................... 34
Part II: Getting Started with Informatica Developer (Data Quality) ................. 35

Chapter 10: Lesson 1. Setting Up Informatica Developer ........................................ 36
Setting Up Informatica Developer Overview ....................................................... 36
Task 1. Start Informatica Developer ............................................................... 37
Task 2. Add a Domain .................................................................................. 37
Task 3. Add a Model Repository ................................................................. 38
Task 4. Create a Project ............................................................................... 38
Task 5. Create a Folder .............................................................................. 38
Task 6. Select a Default Data Integration Service ......................................... 39
Setting Up Informatica Developer Summary .................................................... 39

Chapter 11: Lesson 2. Importing Physical Data Objects ................................. 40
Importing Physical Data Objects Overview .................................................... 40
Task 1. Import the Boston_Customers Flat File Data Object ......................... 41
Task 2. Import the LA_Customers Flat File Data Object ................................ 41
Task 3. Importing the All_Customers Flat File Data Object ............................. 42
Importing Physical Data Objects Summary .................................................... 43

Chapter 12: Lesson 3. Profiling Data ......................................................... 44
Profiling Data Overview ............................................................................... 44
Task 1. Perform a Join Analysis on Two Data Sources .................................. 45
Task 2. View Join Analysis Results .............................................................. 46
Task 3. Run a Profile on a Data Source ........................................................ 46
Task 4. View Column Profiling Results ........................................................ 47
Profiling Data Summary ............................................................................... 47

Appendix A: Frequently Asked Questions ................................................... 48
Informatica Administrator Frequently Asked Questions ................................... 48
Informatica Analyst FAQ .............................................................................. 49
Informatica Developer Frequently Asked Questions ...................................... 49
Preface

The Informatica Data Explorer Getting Started Guide is written for data quality and data services developers and analysts. It provides a tutorial to help first-time users learn how to use Informatica Developer and Informatica Analyst. This guide assumes that you have an understanding of data quality concepts, flat file and relational database concepts, and the database engines in your environment.

Informatica Resources

Informatica Customer Portal

As an Informatica customer, you can access the Informatica Customer Portal site at http://mysupport.informatica.com. The site contains product information, user group information, newsletters, access to the Informatica customer support case management system (ATLAS), the Informatica How-To Library, the Informatica Knowledge Base, the Informatica Multimedia Knowledge Base, Informatica Product Documentation, and access to the Informatica user community.

Informatica Documentation

The Informatica Documentation team takes every effort to create accurate, usable documentation. If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com. We will use your feedback to improve our documentation. Let us know if we can contact you regarding your comments.

The Documentation team updates documentation as needed. To get the latest documentation for your product, navigate to Product Documentation from http://mysupport.informatica.com.

Informatica Web Site

You can access the Informatica corporate web site at http://www.informatica.com. The site contains information about Informatica, its background, upcoming events, and sales offices. You will also find product and partner information. The services area of the site includes important information about technical support, training and education, and implementation services.

Informatica How-To Library

As an Informatica customer, you can access the Informatica How-To Library at http://mysupport.informatica.com. The How-To Library is a collection of resources to help you learn more about Informatica products and features. It includes articles and interactive demonstrations that provide solutions to common problems, compare features and behaviors, and guide you through performing specific real-world tasks.
Informatica Knowledge Base

As an Informatica customer, you can access the Informatica Knowledge Base at http://mysupport.informatica.com. Use the Knowledge Base to search for documented solutions to known technical issues about Informatica products. You can also find answers to frequently asked questions, technical white papers, and technical tips. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team through email at KB_Feedback@informatica.com.

Informatica Multimedia Knowledge Base

As an Informatica customer, you can access the Informatica Multimedia Knowledge Base at http://mysupport.informatica.com. The Multimedia Knowledge Base is a collection of instructional multimedia files that help you learn about common concepts and guide you through performing specific tasks. If you have questions, comments, or ideas about the Multimedia Knowledge Base, contact the Informatica Knowledge Base team through email at KB_Feedback@informatica.com.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or through the Online Support. Online Support requires a user name and password. You can request a user name and password at http://mysupport.informatica.com.

Use the following telephone numbers to contact Informatica Global Customer Support:

<table>
<thead>
<tr>
<th>North America / South America</th>
<th>Europe / Middle East / Africa</th>
<th>Asia / Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Toll Free</strong></td>
<td><strong>Toll Free</strong></td>
<td><strong>Toll Free</strong></td>
</tr>
<tr>
<td>Brazil: 0800 891 0202</td>
<td>France: 00800 4632 4357</td>
<td>Australia: 1 800 151 830</td>
</tr>
<tr>
<td>Mexico: 001 888 209 8853</td>
<td>Germany: 00800 4632 4357</td>
<td>New Zealand: 1 800 151 830</td>
</tr>
<tr>
<td>North America: +1 877 463 2435</td>
<td>Israel: 00800 4632 4357</td>
<td>Singapore: 001 800 4632 4357</td>
</tr>
<tr>
<td><strong>Standard Rate</strong></td>
<td><strong>Standard Rate</strong></td>
<td><strong>Standard Rate</strong></td>
</tr>
<tr>
<td>North America: +1 650 653 6332</td>
<td>Standard Rate France: 0805 804632</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>India: +91 80 4112 5738</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Rate Germany: 01805 702702</td>
<td>030 6022 797</td>
<td></td>
</tr>
<tr>
<td>Netherlands: 00800 4632 4357 or 0800 463 200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland: 00800 4632 4357 or 0800 463 200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom: 00800 4632 4357 or 0800 463 200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 1

Getting Started Overview

This chapter includes the following topics:

- Informatica Domain Overview, 1
- Introducing Informatica Analyst, 4
- Informatica Developer Overview, 5
- The Tutorial Story, 7
- The Tutorial Structure, 7

Informatica Domain Overview

Informatica has a service-oriented architecture that provides the ability to scale services and to share resources across multiple machines. The Informatica domain is the primary unit for management and administration of services.

Informatica contains the following components:

- Application clients. A group of clients that you use to access underlying Informatica functionality. Application clients make requests to the Service Manager or application services.
- Application services. A group of services that represent server-based functionality. An Informatica domain can contain a subset of application services. You configure the application services that are required by the application clients that you use.
- Repositories. A group of relational databases that store metadata about objects and processes required to handle user requests from application clients.
- Service Manager. A service that is built in to the domain to manage all domain operations. The Service Manager runs the application services and performs domain functions including authentication, authorization, and logging.

You can log in to Informatica Administrator (the Administrator tool) after you install Informatica. You use the Administrator tool to manage the domain and configure the required application services before you can access the remaining application clients.
The following figure shows the application services and the repositories that each application client uses in an Informatica domain:

The following table lists the application clients, not including the Administrator tool, and the application services and the repositories that the client requires:

<table>
<thead>
<tr>
<th>Application Client</th>
<th>Application Services</th>
<th>Repositories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Analyzer</td>
<td>Reporting Service</td>
<td>Data Analyzer repository</td>
</tr>
<tr>
<td>Informatica Analyst</td>
<td>- Analyst Service</td>
<td>Model repository</td>
</tr>
<tr>
<td></td>
<td>- Data Integration Service</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Model Repository Service</td>
<td></td>
</tr>
<tr>
<td>Informatica Developer</td>
<td>- Analyst Service</td>
<td>Model repository</td>
</tr>
<tr>
<td></td>
<td>- Content Management Service</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Data Integration Service</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Model Repository Service</td>
<td></td>
</tr>
<tr>
<td>Metadata Manager</td>
<td>- Metadata Manager Service</td>
<td>Metadata Manager repository</td>
</tr>
<tr>
<td></td>
<td>- PowerCenter Integration Service</td>
<td>PowerCenter repository</td>
</tr>
<tr>
<td></td>
<td>- PowerCenter Repository Service</td>
<td></td>
</tr>
</tbody>
</table>
The following application services are not accessed by an Informatica application client:

- **PowerExchange Listener Service.** Manages the PowerExchange Listener for bulk data movement and change data capture. The PowerCenter Integration Service connects to the PowerExchange Listener through the Listener Service.

- **PowerExchange Logger Service.** Manages the PowerExchange Logger for Linux, UNIX, and Windows to capture change data and write it to the PowerExchange Logger Log files. Change data can originate from DB2 recovery logs, Oracle redo logs, a Microsoft SQL Server distribution database, or data sources on an i5/OS or z/OS system.

- **SAP BW Service.** Listens for RFC requests from SAP BI and requests that the PowerCenter Integration Service run workflows to extract from or load to SAP BI.

### Feature Availability

Informatica 9.1.0 products use a common set of applications. The product features you can use depend on your product license.

The following table describes the licensing options and the application features available with each option:

<table>
<thead>
<tr>
<th>Licensing Option</th>
<th>Informatica Developer Features</th>
<th>Informatica Analyst Features</th>
</tr>
</thead>
</table>
| Data Explorer Advanced Edition | - Profiling  
- Scorecarding                                  | - Profiling  
- Scorecarding  
- Create and run profiling rules  
- Reference table management |
| Data Quality                | - Create and run mappings with all transformations  
- Create and run rules  
- Profiling  
- Scorecarding  
- Export objects to PowerCenter | - Profiling  
- Scorecarding  
- Reference table management  
- Create profiling rules  
- Run rules in profiles  
- Bad and duplicate record management |
<table>
<thead>
<tr>
<th>Licensing Option</th>
<th>Informatica Developer Features</th>
<th>Informatica Analyst Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Services</td>
<td>- Create logical data object models&lt;br&gt;- Create and run mappings with Data Services transformations&lt;br&gt;- Create SQL data services&lt;br&gt;- Create web services&lt;br&gt;- Export objects to PowerCenter</td>
<td>- Reference table management</td>
</tr>
<tr>
<td>Data Services and Profiling Option</td>
<td>- Create logical data object models&lt;br&gt;- Create and run mappings with Data Services transformations&lt;br&gt;- Create SQL data services&lt;br&gt;- Create web services&lt;br&gt;- Export objects to PowerCenter&lt;br&gt;- Create and run rules with Data Services transformations&lt;br&gt;- Profiling</td>
<td>- Reference table management</td>
</tr>
</tbody>
</table>

**Note:** Informatica Data Explorer Advanced Edition functionality is a subset of Informatica Data Quality functionality.

## Introducing Informatica Analyst

Informatica Analyst is a web-based application client that analysts can use to analyze, cleanse, standardize, profile, and score data in an enterprise.

Business analysts and developers use Informatica Analyst for data-driven collaboration. You can perform column and rule profiling, scorecarding, and bad record and duplicate record management. You can also manage reference data and provide the data to developers in a data quality solution.

Use Informatica Analyst to accomplish the following tasks:

- **Profile data.** Create and run a profile to analyze the structure and content of enterprise data and identify strengths and weaknesses. After you run a profile, you can selectively drill down to see the underlying rows from the profile results. You can also add columns to scorecards and add column values to reference tables.
- **Create rules in profiles.** Create and apply rules within profiles. A rule is reusable business logic that defines conditions applied to data when you run a profile. Use rules to further validate the data in a profile and to measure data quality progress.
- **Score data.** Create scorecards to score the valid values for any column or the output of rules. Scorecards display the value frequency for columns in a profile as scores. Use scorecards to measure and visually represent data quality progress. You can also view trend charts to view the history of scores over time.
- **Manage reference data.** Create and update reference tables for use by analysts and developers to use in data quality standardization and validation rules. Create, edit, and import data quality dictionary files as reference tables. Create reference tables to establish relationships between source data and valid and standard values. Developers use reference tables in standardization and lookup transformations in Informatica Developer.
- **Manage bad records and duplicate records.** Fix bad records and consolidate duplicate records.
Informatica Developer Overview

Informatica Developer is an application client that developers use to design and implement data quality and data services solutions.

The following figure shows the Developer tool:

![Developer Tool Screenshot]

The Developer tool includes an editor, in which you can edit objects. In this example, the editor shows the Customer_Objects logical data object model. Depending on the object in the editor, the Developer tool displays views, such as the default view.

The Developer tool also includes the following views that appear independently of the objects in the editor:

- **Object Explorer.** Shows projects, folders, and the objects they contain.
- **Outline.** Shows dependent objects in an object.
- **Properties.** Shows object properties.
- **Data Viewer.** Shows the results of a mapping, data preview, or an SQL query.
- **Validation Log.** Shows object validation errors.
- **Cheat Sheets.** Shows cheat sheets.

You can hide any view and move any view to another location in the Developer tool. You can also display other views, such as the Search view. Click **Window > Show View** to select the views you want to display.

Informatica Developer Welcome Page

The first time you open the Developer tool, the Welcome page appears. Use the Welcome page to learn more about the Developer tool, set up the Developer tool, and to start working in the Developer tool.

The Welcome page displays the following options:

- **Overview.** Click the Overview button to get an overview of data quality and data services solutions.
- **First Steps.** Click the First Steps button to learn more about setting up the Developer tool and accessing Informatica Data Quality and Informatica Data Services lessons.
- **Tutorials.** Click the Tutorials button to see tutorial lessons for data quality and data services solutions.
Web Resources. Click the Web Resources button for a link to mysupport.informatica.com. You can access the Informatica How-To Library. The Informatica How-To Library contains articles about Informatica Data Quality, Informatica Data Services, and other Informatica products.

Workbench. Click the Workbench button to start working in the Developer tool.

Cheat Sheets

The Developer tool includes cheat sheets as part of the online help. A cheat sheet is a step-by-step guide that helps you complete one or more tasks in the Developer tool.

After you complete a cheat sheet, you complete the tasks and see the results. For example, after you complete a cheat sheet to import and preview a relational data object, you have imported a relational database table and previewed the data in the Developer tool.

To access cheat sheets, click Help > Cheat Sheets.

Data Quality and Data Explorer Advanced Edition

Use the data quality capabilities in the Developer tool to analyze the content and structure of your data and enhance the data in ways that meet your business needs.

Use the Developer tool to design and run processes that achieve the following objectives:

- Profile data. Profiling reveals the content and structure of your data. Profiling is a key step in any data project, as it can identify strengths and weaknesses in your data and help you define your project plan.
- Create scorecards to review data quality. A scorecard is a graphical representation of the quality measurements in a profile.
- Standardize data values. Standardize data to remove errors and inconsistencies that you find when you run a profile. You can standardize variations in punctuation, formatting, and spelling. For example, you can ensure that the city, state, and ZIP code values are consistent.
- Parse records. Parse data records to improve record structure and derive additional information from your data. You can split a single field of freeform data into fields that contain different information types. You can also add information to your records. For example, you can flag customer records as personal or business customers.
- Validate postal addresses. Address validation evaluates and enhances the accuracy and deliverability of your postal address data. Address validation corrects errors in addresses and completes partial addresses by comparing address records against reference data from national postal carriers. Address validation can also add postal information that speeds mail delivery and reduces mail costs.
- Find duplicate records. Duplicate record analysis compares a set of records against each other to find similar or matching values in selected data columns. You set the level of similarity that indicates a good match between field values. You can also set the relative weight given to each column in match calculations. For example, you can prioritize surname information over forename information.
- Create reference data tables. Reference data tables are key elements in data standardization. Informatica provides a comprehensive set of reference data tables. You can create custom reference tables from columns in your source data.
- Create and run data quality rules. Informatica provides pre-built rules that you can run or edit to suit your project objectives. You can create rules in the Developer tool.
- Collaborate with Informatica users. The rules and reference data tables you add to the Model repository are available to users in the Developer tool and the Analyst tool. Users can collaborate on projects, and different users can take ownership of objects at different stages of a project.
- Export mappings to PowerCenter. You can export mappings to PowerCenter to reuse the metadata for physical data integration or to create web services.
Data Quality users can perform all the tasks above.

Data Explorer Advanced Edition users can profile data in the Developer tool and can create scorecards that run in the Analyst tool.

The Tutorial Story

HypoStores Corporation is a national retail organization with headquarters in Boston and stores in several states. It integrates operational data from stores nationwide with the data store at headquarters on a regular basis. It recently opened a store in Los Angeles.

The headquarters includes a central ICC team of administrators, developers, and architects responsible for providing a common data services layer for all composite and BI applications. The BI applications include a CRM system that contains the master customer data files used for billing and marketing.

HypoStores Corporation wants to profile the Boston and Los Angeles data before it integrates the data sets. The profile operations identify data quality issues that HypoStores can fix before the integration.

The Tutorial Structure

The Getting Started Guide contains tutorials that include lessons and tasks.

Lessons

Each lesson introduces concepts that will help you understand the tasks to complete in the lesson. The lesson provides business requirements from the overall story. The objectives for the lesson outline the tasks that you will complete to meet business requirements. Each lesson provides an estimated time for completion. When you complete the tasks in the lesson, you can review the lesson summary.

If the environment within the tool is not configured, the first lesson in each tutorial helps you do so.

Tasks

The tasks provide step-by-step instructions. Complete all tasks in the order listed to complete the lesson.

Tutorial Prerequisites

Before you can begin the tutorial lessons, the Informatica domain must be running with at least one node set up.

The installer includes tutorial files that you will use to complete the lessons. You can find all the files in both the client and server installations:

- You can find the tutorial files in the following location in the Developer tool installation path:
  `<Informatica Installation Directory>\clients\DeveloperClient\Tutorials`
- You can find the tutorial files in the following location in the services installation path:
  `<Informatica Installation Directory>\server\Tutorials`

You need the following files for the tutorial lessons:

- All_Customers.csv
- Boston_Customers.csv
Informatica Analyst Tutorial

During this tutorial, an analyst logs into the Analyst tool, creates projects and folders, creates profiles and rules, scores data, and creates reference tables.

The lessons you can perform depend on whether you have the Informatica Data Quality, Informatica Data Explorer, Informatica Data Services, or PowerCenter products.

The following table describes the lessons you can perform, depending on your product.

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Description</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1. Setting up Informatica Analyst</td>
<td>Log in to the Analyst tool and create a project and folder for the tutorial lessons.</td>
<td>All</td>
</tr>
<tr>
<td>Lesson 2. Creating Data Objects</td>
<td>Import a flat file as a data object and preview the data.</td>
<td>Data Quality, Data Explorer</td>
</tr>
<tr>
<td>Lesson 3. Creating Quick Profiles</td>
<td>Creating a quick profile to quickly get an idea of data quality.</td>
<td>Data Quality, Data Explorer</td>
</tr>
<tr>
<td>Lesson 4. Creating Custom Profiles</td>
<td>Create a custom profile to configure columns, and sampling and drilldown options.</td>
<td>Data Quality, Data Explorer</td>
</tr>
<tr>
<td>Lesson 5. Creating Expression Rules</td>
<td>Create expression rules to modify and profile column values.</td>
<td>Data Quality</td>
</tr>
<tr>
<td>Lesson 6. Creating and Running Scorecards</td>
<td>Create and run a scorecard to measure data quality progress over time.</td>
<td>Data Quality, Data Explorer</td>
</tr>
<tr>
<td>Lesson 7. Creating Reference Tables from Profile Results</td>
<td>Create a reference table that you can use to standardize source data.</td>
<td>Data Quality, Data Explorer, Data Services</td>
</tr>
<tr>
<td>Lesson 8. Creating Reference Tables</td>
<td>Create a reference table to establish relationships between source data and valid and standard values.</td>
<td>All</td>
</tr>
</tbody>
</table>

Note: This tutorial does not include lessons on bad record and consolidation record management.

Informatica Developer Tool (Data Quality)

In this tutorial, you use the Developer tool to perform several data quality operations.

Informatica Data Quality and Informatica Data Explorer users use the Developer tool to create and run profiles that analyze the content and structure of data.

Informatica Data Quality users use the Developer tool to design and run processes that enhance data quality.

Complete the following lessons in the data quality tutorial:
Lesson 1. Setting Up Informatica Developer
Create a connection to a Model repository that is managed by a Model Repository Service in a domain. Create a project and folder to store work for the lessons in the tutorial. Select a default Data Integration Service.

Lesson 2. Importing Physical Data Objects
You will define data quality processes for the customer data files associated with these objects.

Lesson 3. Profiling Data
Profiling reveals the content and structure of your data.
Profiling includes join analysis, a form of analysis that determines if a valid join is possible between two data columns.

Lesson 4. Parsing Data
Parsing enriches your data records and improves record structure. It can find useful information in your data and also derive new information from current data.

Lesson 5. Standardizing Data
Standardization removes data errors and inconsistencies found during profiling.

Lesson 6. Validating Address Data
Address validation evaluates the accuracy and deliverability of your postal addresses and fixes address errors and omissions in addresses.
Part I: Getting Started with Informatica Analyst

This part contains the following chapters:

- Lesson 1. Setting Up Informatica Analyst, 11
- Lesson 2. Creating Data Objects, 14
- Lesson 3. Creating Quick Profiles, 17
- Lesson 4. Creating Custom Profiles, 20
- Lesson 5. Creating Expression Rules, 23
- Lesson 6. Creating and Running Scorecards, 26
- Lesson 7. Creating Reference Tables from Profile Columns, 30
- Lesson 8. Creating Reference Tables, 33
Lesson 1. Setting Up Informatica Analyst

This chapter includes the following topics:
- Setting Up Informatica Analyst Overview, 11
- Task 1. Log In to Informatica Analyst, 12
- Task 2. Create a Project, 12
- Task 3. Create a Folder, 12
- Setting Up Informatica Analyst Summary, 13

Setting Up Informatica Analyst Overview

Before you start the lessons in this tutorial, you must set up the Analyst tool. To set up the Analyst tool, log in to the Analyst tool and create a project and a folder to store your work.

The Informatica domain is a collection of nodes and services that define the Informatica environment. Services in the domain include the Analyst Service and the Model Repository Service. The Analyst Service runs the Analyst tool, and the Model Repository Service manages the Model repository. When you work in the Analyst tool, the Analyst tool stores the objects that you create in the Model repository.

You must create a project before you can create objects in the Analyst tool. A project contains objects in the Analyst tool. A project can also contain folders that store related objects, such as objects that are part of the same business requirement.

Objectives
In this lesson, you complete the following tasks:
- Log in to the Analyst tool.
- Create a project to store the objects that you create in the Analyst tool.
- Create a folder in the project that can store related objects.

Prerequisites
Before you start this lesson, verify the following prerequisites:
- An administrator has configured a Model Repository Service and an Analyst Service in the Administrator tool.
- You have the host name and port number for the Analyst tool.
You have a user name and password to access the Analyst Service. You can get this information from an administrator.

Timing
Set aside 5 to 10 minutes to complete this lesson.

Task 1. Log In to Informatica Analyst

Log in to the Analyst tool to begin the tutorial.

1. Start a Microsoft Internet Explorer or Mozilla Firefox browser.
2. In the Address field, enter the URL for Informatica Analyst:
   
   http[s]://<host name>:<port number>/AnalystTool

3. On the login page, enter the user name and password.
4. Select Native or the name of a specific security domain.
   The Security Domain field appears when the Informatica domain contains an LDAP security domain. If you do not know the security domain that your user account belongs to, contact the Informatica domain administrator.
5. Click Login.
   The welcome screen appears.
6. Click Close to exit the welcome screen and access the Analyst tool.

Task 2. Create a Project

In this task, you create a project to contain the objects that you create in the Analyst tool. Create a tutorial project to contain the folder for the data quality project.

1. In the Analyst tool, select the Projects folder in the Navigator.
   The Navigator is the left pane in the Analyst interface.
2. Click Actions > New Project in the Navigator.
   The New Project window appears.
3. Enter your name prefixed by “Tutorial_” as the name of the project.
4. Verify that Unshared is selected.
5. Click OK.

Task 3. Create a Folder

In this task, you create a folder to store related objects. You can create a folder in a project or another folder. Create a folder named Customers to store the objects related to the data quality project.

1. In the Navigator, select the tutorial project.
Setting Up Informatica Analyst Summary

In this lesson, you learned that the Analyst tool stores objects in projects and folders. A Model repository contains the projects and folders. The Analyst Service runs the Analyst tool. The Model Repository Service manages the Model repository. The Analyst Service and the Model Repository Service are application services in the Informatica domain.

You logged in to the Analyst tool and created a project and a folder.

Now, you can use the Analyst tool to complete other lessons in this tutorial.
Creating Data Objects Overview

In the Analyst tool, a data object is a representation of data based on a flat file or relational database table. You create a flat file or table object and then run a profile against the data in the flat file or relational database table. When you create a flat file data object in the Analyst tool, you can upload the file to the flat file cache on the machine that runs the Analyst tool or you can specify the network location where the flat file is stored.

Story

HypoStores keeps the Los Angeles customer data in flat files. HypoStores needs to profile and analyze the data and perform data quality tasks.

Objectives

In this lesson, you complete the following tasks:

1. Upload the flat file to the flat file cache location and create a data object.
2. Preview the data for the flat file data object.

Prerequisites

Before you start this lesson, verify the following prerequisites:

- You have completed lesson 1 in this tutorial.
- You have the LA_Customers.csv flat file. You can download the file here (requires a my.informatica.com account).

Timing

Set aside 5 to 10 minutes to complete this task.
Task 1. Create the Flat File Data Objects

In this task, you use the Add Flat File wizard to create a flat file data objects from the LA_Customers, Customers, Accounts and Account_Customers data files.

1. In the Navigator, select the Customers folder in your tutorial project.
   
   **Note:** You must select the project or folder where you want to create the flat file data object before you can create it.

2. Click **Actions > New > New Flat File**.
   
   The Add Flat File wizard appears.

3. Select **Browse and Upload**, and click **Browse**.

4. Browse to the location of LA_Customers.csv, and click **Open**.

5. Click **Next**.

6. Under **Specify lines to import**, select **Import from first line** to import column names from the first non-blank line.

7. Click **Show**.
   
   The details panel updates to show the column headings from the first row.

8. Click **Next**.

   The Column Attributes panel shows the datatype, precision, scale, and format for each column.

9. For the CreateDate and MiscDate columns, click the Data Type cell and change the datatype to datetime.

10. Click **Next**.

    The Name field displays LA_Customers.

11. Optionally, change the name of the file and add a description.

    The Customers folder is selected by default on the bottom, left pane.

12. Click **Finish**.

    The data object appears in the folder contents for the Customers folder.

13. Repeat steps 2 through 12 to create flat file data objects for the Customers, Accounts, and Account_Customers data files.

Task 2. Preview the Data

In this task, you preview the data for the flat file data object to review the structure and content of the data.

1. In the Navigator, select the Customers folder in your tutorial project.

   The contents of the folder appear in the Content panel.

2. Click the LA_Customers data object.

   The data object opens in a tab. The Analyst tool displays the first 100 rows of the flat file data object in the Data preview view.

3. Click the Properties view for the flat file data object.

   The Properties view displays the name, description, and location of the data object. It also displays the columns and column properties for the data object.
Creating Data Objects Summary

In this lesson, you learned that data objects are representations of data based on a flat file or a relational database source. You learned that you can create a flat file data object and preview the data in it.

You uploaded a flat file and created a flat file data object, previewed the data for the data object, and viewed the properties for the data object.

After you create a data object, you create a quick profile for the data object in Lesson 3, and you create a custom profile for the data object in Lesson 4.
Lesson 3. Creating Quick Profiles

This chapter includes the following topics:
- Creating Quick Profiles Overview, 17
- Task 1. Create and Run a Quick Profile, 18
- Task 2. View the Profile Results, 18
- Creating Quick Profiles Summary, 19

Creating Quick Profiles Overview

A profile is the analysis of data quality based on the content and structure of data. A quick profile is a profile that you create with default options. Use a quick profile to get profile results without configuring all columns and options for a profile.

Create and run a quick profile to analyze the quality of the data when you start a data quality project. When you create a quick profile object, you select the data object and the data object columns that you want to analyze. A quick profile skips the profile column and option configuration. The Analyst tool performs profiling on the staged flat file for the flat file data object.

Story

HypoStores wants to incorporate data from the newly-acquired Los Angeles office into its data warehouse. Before the data can be incorporated into the data warehouse, it needs to be cleansed. You are the analyst who is responsible for assessing the quality of the data and passing the information on to the developer who is responsible for cleansing the data. You want to view the profile results quickly and get a basic idea of the data quality.

Objectives

In this lesson, you complete the following tasks:
1. Create and run a quick profile for the Customers_LA flat file data object.
2. View the profile results.

Prerequisites

Before you start this lesson, verify the following prerequisite:
- You have completed lessons 1 and 2 in this tutorial.

Timing

Set aside 5 to 10 minutes to complete this lesson.
Task 1. Create and Run a Quick Profile

In this task, you create a quick profile for all columns in the data object and use default sampling and drilldown options.

1. In the Navigator, select the Customers folder in your tutorial project.
2. In the Contents panel, click to the right of the link for the Customers_LA data object.
   Do not click the link for the object.
3. Click Actions > New > New Profile.
   The New Profile wizard appears.
4. Click Save and Run to create and run the profile.
   The Analyst tool creates the profile in the same project and folder as the data object.
   The profile results for the quick profile appear in a new tab after you save and run the profile.

Task 2. View the Profile Results

In this task, you use Column Profiling view for the LA_Customers profile to get a quick overview of the profile results.

The following table describes the information that appears for each column in a profile:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name of the column in the profile.</td>
</tr>
<tr>
<td>Unique Values</td>
<td>Number of unique values in the column</td>
</tr>
<tr>
<td>Unique %</td>
<td>Percentage of unique values in the column.</td>
</tr>
<tr>
<td>Null</td>
<td>Number of null values in the column.</td>
</tr>
<tr>
<td>Null %</td>
<td>Percentage of column values that are null.</td>
</tr>
<tr>
<td>Datatype</td>
<td>Data type derived from the values in the column. The Analyst tool can derive the following datatypes from the column values:</td>
</tr>
<tr>
<td></td>
<td>String</td>
</tr>
<tr>
<td></td>
<td>Varchar</td>
</tr>
<tr>
<td></td>
<td>Decimal</td>
</tr>
<tr>
<td></td>
<td>Integer</td>
</tr>
<tr>
<td></td>
<td>Null [-]</td>
</tr>
<tr>
<td>Inferred %</td>
<td>Percentage of values that match the data type inferred by the Analyst tool.</td>
</tr>
<tr>
<td>Documented Datatype</td>
<td>Data type declared for the column in the profiled object.</td>
</tr>
<tr>
<td>Max Value</td>
<td>Maximum value in the column.</td>
</tr>
<tr>
<td>Min Value</td>
<td>Minimum value in the column.</td>
</tr>
</tbody>
</table>
1. Click the header for the Null Values column to sort the values.
   Notice that the Address2, Address3, City2, CreateDate, and MiscDate columns have 100% null values.
   In Lesson 4, you create a custom profile to exclude these columns.

2. Click the Full Name column. The values for the column appear in the Values view.
   Notice that the first and last names do not appear in separate columns.
   In Lesson 5, you create a rule to separate the first and last names into separate columns.

3. Click the CustomerTier column.
   Notice that the values for the CustomerTier are inconsistent.
   In Lesson 6, you create a scorecard to score the CustomerTier values. In Lesson 7, you create a reference table that a developer can use to standardize the CustomerTier values.

4. Click the State column and then click the Patterns view.
   Notice that 483 columns have a pattern of XX, which indicate valid values. Seventeen values are not valid because they do not match the valid pattern.
   In Lesson 6, you create a scorecard to score the State values.

Creating Quick Profiles Summary

In this lesson, you learned that a quick profile shows profile results without configuring all columns and row sampling options for a profile. You learned that you create and run a quick profile to analyze the quality of the data when you start a data quality project. You also learned that the Analyst tool performs profiling on the staged flat file for the flat file data object.

You created a quick profile and analyzed the profile results. You got more information about the columns in the profile, including null values and datatypes. You also used the column values and patterns to identify data quality issues.

After you analyze the results of a quick profile, you can complete the following tasks:

- Create a custom profile to exclude columns from the profile and only include the columns you are interested in.
- Create an expression rule to create virtual columns and profile them.
- Create a reference table to include valid values for a column.
Lesson 4. Creating Custom Profiles

Creating Custom Profiles Overview

A profile is the analysis of data quality based on the content and structure of data. A custom profile is a profile that you create when you want to configure the columns, sampling options, and drilldown options for faster profiling. Configure sampling options to select the sample rows in the flat file. Configure drilldown options to drill down to records in the profile results and drill down to data rows in the source data or staged data.

You create and run a profile to analyze the quality of the data when you start a data quality project. When you create a profile object, you select the data object and the data object columns that you want to profile, configure the sampling options, and configure the drilldown options.

Story

HypoStores needs to incorporate data from the newly-acquired Los Angeles office into its data warehouse. HypoStores wants to access the quality of the customer tier data in the LA customer data file. You are the analyst who is responsible for assessing the quality of the data and passing the information on to the developer who is responsible for cleansing the data.

Objectives

In this lesson, you complete the following tasks:

1. Create a custom profile for the flat file data object and exclude the columns with null values.
2. Run the profile to analyze the content and structure of the CustomerTier column.
3. Drill down into the rows for the profile results.

Prerequisites

Before you start this lesson, verify the following prerequisite:

- You have completed lessons 1, 2, and 3 in this tutorial.

Timing

Set aside 5 to 10 minutes to complete this lesson.
Task 1. Create a Custom Profile

In this task, you use the New Profile wizard to create a custom profile. When you create a profile, you select the data object and the columns that you want to profile. You also configure the sampling and drill down options.

1. In the Navigator, select the Customers folder in your tutorial project.
2. Click Actions > New > New Profile.
   The New Profile wizard appears.
3. In the Sources panel, select the LA_Customers data object.
   The Columns panel shows the columns for the data object.
4. Click Next.
5. Enter Profile_LA_Customers_Custom for the name.
6. Verify the location in the Folders panel. The location shows the tutorial project and the Customers folder.
   The Profiles panel shows Profile.LA_Customers.
7. Click Next.
8. In the Columns panel, clear the Address2, Address3, City2, CreateDate, and MiscDate columns.
9. In the Sampling Options panel, select the All Rows option.
10. In the Drilldown Options panel, verify that Enable Row Drilldown is selected and select on staged data for the Drilldown option.
11. Click Next.
12. Optionally, define a filter for the profile.
13. Click Save.
   The Analyst tool creates the profile and displays the profile in another tab.

Task 2. Run the Profile

In this task, you run a profile to perform profiling on the data object and display the profile results. The Analyst tool performs profiling on the staged flat file for the flat file data object.

1. In the Navigator, select the Customers folder in your tutorial project.
2. In the contents panel, click the Profile_LA_Customers_Custom link.
   The profile appears in a tab.
3. Click Actions > Run Profile.
   The Column Profile window appears.
4. In the Columns panel, select Name to select all columns to profile.
5. In the Sampling Options panel, choose to include the default options.
6. In the Drilldown Options panel, choose to include the default options.
7. Click Run.
   The Analyst tool performs profiling on the data object and displays the profile results.
Task 3. Drill Down on Profile Results

In this task, you drill down on the CustomerTier column values to see the underlying rows in the data object for the profile.

1. In the Navigator, select the Customers folder in your tutorial project.
2. Click the Profile_LA_Customers_Custom profile.
   The profile opens in a tab.
3. In the Column Profiling view, select the CustomerTier column.
   The values for the column appear in the Values view.
4. Use the shift key to select the Diamond, Ruby, Emerald, and Bronze values.
5. Right-click and select Drilldown.
   The rows for the columns with a value of Diamond, Ruby, Emerald, and Bronze appear in the Drilldown panel. Only the selected columns appear in the Drilldown panel.
6. In the Column Profiling view, enable the preview option for the CustomerID column and select the Diamond, Ruby, Emerald, and Bronze values in the Values view.
   The underlying rows in the Drilldown panel now include the CustomerID column. The title bar for the Drilldown panel shows the logic used for the underlying columns.

Creating Custom Profiles Summary

In this lesson, you learned that you can configure the columns that get profiled and that you can configure the sampling and drilldown options. You learned that you can drill down to see the underlying rows for column values and that you can configure the columns that are included when you view the column values.

You created a custom profile that included the CustomerTier column, ran the profile, and drilled down to the underlying rows for the CustomerTier column in the results.

Use the custom profile object to create an expression rule in lesson 5. If you have Data Quality or Data Explorer, you can create a scorecard in lesson 6.
Lesson 5. Creating Expression Rules

This chapter includes the following topics:

- Creating Expression Rules Overview, 23
- Task 1. Create Expression Rules and Run the Profile, 24
- Task 2. View the Expression Rule Output, 24
- Task 3. Edit the Expression Rules, 25
- Creating Expression Rules Summary, 25

Creating Expression Rules Overview

Expression rules use expression functions and source columns to define rule logic. You can create expression rules and add them to a profile in the Analyst tool. An expression rule can be associated with one or more profiles.

The output of an expression rule is a virtual column in the profile. The Analyst tool profiles the virtual column when you run the profile.

You can use expression rules to validate source columns or create additional source columns based on the value of the source columns.

Story

HypoStores wants to incorporate data from the newly-acquired Los Angeles office into its data warehouse. HypoStores wants to analyze the customer names and separate customer names into first name and last name. HypoStores wants to use expression rules to parse a column that contains first and last names into separate virtual columns and then profile the columns. HypoStores also wants to make the rules available to other analysts who need to analyze the output of these rules.

Objectives

In this lesson, you complete the following tasks:

1. Create expression rules to separate the FullName column into first name and last name columns. You create a rule that separates the first name from the full name. You create another rule that separates the last name from the first name. You create these rules for the Profile_LA_Customers_Custom profile.
2. Run the profile and view the output of the rules in the profile.
3. Edit the rules to make them usable for other Analyst tool users.
Prerequisites
Before you start this lesson, verify the following prerequisite:

- You have completed Lessons 1, 2, 3, and 4.

Timing
Set aside 10 to 15 minutes to complete this lesson.

Task 1. Create Expression Rules and Run the Profile

In this task, you create two expression rules to parse the FullName column into two virtual columns named FirstName and LastName. The FirstName and LastName columns are the rule names.

1. In the contents panel, click the Profile_LA_Customers_Custom profile to open it.
   The profile appears in a tab.
2. Click Actions > Add Rule.
   The New Rule window appears.
3. Select Create a rule.
4. Click Next.
5. Enter FirstName for the rule name.
6. In the Expression panel, enter the following expression to separate the first name from the Name column:
   \[\text{SUBSTR(FullName,1,INSTR(FullName, ',1,1) - 1)}\]
7. Click Validate.
8. Click Next.
9. Optionally, configure the column, sampling, and drilldown options.
10. Click Save.
    The Analyst tool creates the rule and displays it in the Column Profiling view.
11. Repeat steps 2 through 10 and create a rule named LastName and enter the following expression to separate the last name from the Name column:
    \[\text{SUBSTR(FullName,INSTR(FullName, ',,-1,1),LENGTH(FullName))}\]

Task 2. View the Expression Rule Output

In this task, you view the output of expression rules that separated first and last names after running a profile.

1. In the contents panel, click Actions > Run Profile.
2. In the Column Profiling view, click Preview in the toolbar to clear all columns.
3. Select the FullName column and the FirstName and LastName rules.
4. Click Run.
5. Click the FirstName rule.
   The values appear in the Values view.
6. Select any value in the Values view.
7. Right-click and select Drilldown.

The values for the FullName column and the FirstName and LastName rules appear in the Drilldown panel. Notice that the FullName column is now separated into first and last names.

Task 3. Edit the Expression Rules

In this task, you make the expression rules reusable and available to all Analyst tool users.
1. In the Column Profiling view, select the FirstName rule.
2. Click Actions > Edit Rule.
   The Edit Rule window appears.
3. Select Save as a reusable rule in.
   By default, the Analyst tool saves the rule in the current profile and folder.
4. Click Save.
5. Repeat steps 1 through 4 for the LastName rule.

The FirstName and LastName rules can now be used by any Analyst tool user to split a column with first and last names into separate columns.

Creating Expression Rules Summary

In this lesson, you learned that expression rules use expression functions and source columns to define rule logic. You learned that the output of an expression rule is a virtual column in the profile. The Analyst tool includes the virtual column when you run the profile.

You created two expression rules, added them to a profile, and ran the profile. You viewed the output of the rules and made them available to all Analyst tool users.
Lesson 6. Creating and Running Scorecards

Creating and Running Scorecards Overview

A scorecard is the graphical representation of valid values for a column or the output of a rule in profile results. Use scorecards to measure and monitor data quality progress over time.

To create a scorecard, you add columns from the profile to a scorecard and configure the score thresholds. To run a scorecard, you select the valid values for the column and run the scorecard to see the scores for the columns.

Scorecards display the value frequency for columns in a profile as scores. Scores reflect the percentage of valid values for a column.

Story

HypoStores wants to incorporate data from the newly-acquired Los Angeles office into its data warehouse. Before they merge the data they want to make sure that the data in different customer tiers and states is analyzed for data quality. You are the analyst who is responsible for monitoring the progress of performing the data quality analysis. You want to create a scorecard from the customer tier and state profile columns, configure thresholds for data quality, and view the score trend charts to determine how the scores improve over time.

Objectives

In this lesson, you will complete the following tasks:

1. Create a scorecard from the results of the Profile_LA_Customers_Custom profile to view the scores for the CustomerTier and State columns.
2. Run the scorecard to generate the scores for the CustomerTier and State columns.
3. View the scorecard to see the scores for each column.
4. Edit the scorecard to specify different valid values for the scores.
5. Configure score thresholds and run the scorecard.
6. View score trend charts to determine how scores improve over time.

**Prerequisites**

Before you start this lesson, verify the following prerequisite:

- You have completed lessons 1 through 5 in this tutorial.

**Timing**

Set aside 15 minutes to complete the tasks in this lesson.

---

**Task 1. Create a Scorecard from the Profile Results**

In this task, you create a scorecard from the Profile_LA_Customers_Custom profile to score the CustomerTier and State column values.

1. Open the Profile_LA_Customers_Custom profile.
2. Click **Actions > Add to Scorecard**.
   
   The **Add to Scorecard** wizard appears.
3. Select the CustomerTier and the State columns to add to the scorecard.
4. Click **Next**.
5. Click **New** to create a scorecard.
   
   The **New Scorecard** window appears.
6. Enter sc_LA_Customer for the scorecard name, and navigate to the Customers folder for the scorecard location.
7. Click **OK** and click **Next**.
8. Select the CustomerTier score in the **Scores** panel and select the **Is Valid** column for all values in the **Score using: Values** panel.
9. Select the State score in the **Scores** panel and select the **Is Valid** column for those values that have two letter state codes in the **Score using: Values** panel.
10. For each score in the **Scores** panel, accept the default settings for the score thresholds in **Score Settings** panel.
11. Click **Finish**.
Task 2. Run the Scorecard

In this task, you run the sc_LA_Customer scorecard to generate the scores for the CustomerTier and State columns.

1. Click the sc_LA_Customer scorecard to open it.
   The scorecard appears in a tab.
2. Click Actions > Run Scorecard.
   The Scorecard view displays the scores for the CustomerTier and State columns.

Task 3. View the Scorecard

In this task, you view the sc_LA_Customer scorecard to see the scores for the CustomerTier and State columns.

1. Select the State column that contains the State score you want to view.
2. Click Actions > Show Rows.
   The valid scores for the State column appear in the Valid view. Click Invalid to view the invalid scores for the State column. In the Scores panel, you can view the score name and score percentage. You can view the score displayed as a bar, the data object of the score, and the source and source type of the score.
3. Repeat steps 1 through 2 for the CustomerTier column.
   All scores for the CustomerTier column are valid.

Task 4. Edit the Scorecard

In this task, you will edit the sc_LA_Customer scorecard to specify the Ruby value as not valid for the CustomerTier score.

1. Click Actions > Edit.
   The Edit Scorecard window appears.
2. Select the CustomerTier score in the Scores panel.
3. In the Score using: Values panel, clear Ruby from the Is Valid column.
   Accept the default settings in the Score Settings panel.
4. Click Save to save the changes to the scorecard and run it.
5. View the CustomerTier score again.
Task 5. Configure Thresholds

In this task, you configure thresholds for the State score in the sc_LA_Customer scorecard to determine the acceptable ranges for the data in the State column. Values with a two letter code, such as CA are acceptable, and codes with more than two letters such as Calif are not acceptable.

1. In the Edit Scorecard window, select the State score in the Scores panel.
2. In the Score Settings panel, enter the following ranges for the Good and Unacceptable scores in Set Custom Thresholds for this Score: 90 to 100% Good; 0 to 50% Unacceptable. 51% to 89% are Acceptable. The thresholds represent the lower bounds of the acceptable and good ranges.
3. Click Save to save the changes to the scorecard and run it.

In the Scores panel, view the changes to the score percentage and the score displayed as a bar for the State score.

Task 6. View Score Trend Charts

In this task, you view the trend chart for the State score. You can view trend charts to monitor scores over time.

1. In the Navigator, select the Customers folder in your tutorial project.
2. Click the sc_LA_Customer scorecard to open it.

   The scorecard appears in a tab.
3. In the Scorecard view, select the State score.
4. Click Actions > Show Trend Chart.

   The Trend Chart Detail window appears. You can view the Good, Acceptable, and Unacceptable thresholds for the score. The thresholds change each time you run the scorecard after editing the values for scores in the scorecard.

Creating and Running Scorecards Summary

In this lesson, you learned that you can create a scorecard from the results of a profile. A scorecard contains the columns from a profile. You learned that you can run a scorecard to generate scores for columns. You edited a scorecard to configure valid values and set thresholds for scores. You also learned how to view the score trend chart.

You created a scorecard from the CustomerTier and State columns in a profile to analyze data quality for the customer tier and state columns. You ran the scorecard to generate scores for each column. You edited the scorecard to specify different valid values for scores. You configured thresholds for a score and viewed the score trend chart.
Lesson 7. Creating Reference Tables from Profile Columns

Creating Reference Tables from Profile Columns Overview

A reference table contains reference data that you can use to standardize source data. Reference data can include valid and standard values. Create reference tables to establish relationships between source data values and the valid and standard values.

You can create a reference table from the results of a profile. After you create a reference table, you can edit the reference table to add columns or rows and add or edit standard and valid values. You can view the changes made to a reference table in an audit trail.

Story

HypoStores wants to profile the data to uncover anomalies and standardize the data with valid values. You are the analyst who is responsible for standardizing the valid values in the data. You want to create a reference table based on valid values from profile columns.

Objectives

In this lesson, you complete the following tasks:

1. Create a reference table from the CustomerTier column in the Profile_LA_Customers_Custom profile by selecting valid values for columns.
2. Edit the reference table to configure different valid values for columns.

Prerequisites

Before you start this lesson, verify the following prerequisite:

- You have completed lessons 1 through 6 in this tutorial.
Timing
Set aside 15 minutes to complete the tasks in this lesson.

Task 1. Create a Reference Table from Profile Columns

In this task, you create a reference table and add the CustomerTier column from the Profile_LA_Customers_Custom profile to the reference table.

1. Click the Profile_LA_Customers_Custom profile.
   The profile appears in a tab.

2. In the Column Profiling view, select the CustomerTier column that you want to add to the reference table.
   You can drill down on the value and pattern frequencies for the CustomerTier column to inspect records that have non-standard customer category values.

3. In the Values view, select the valid customer tier values you want to add. Use the CONTROL or SHIFT keys to select the following multiple values: Diamond, Gold, Silver, Bronze, Emerald.

4. Click Actions > Add to Reference Table.
   The New Reference Table wizard appears.

5. Select the option to Create a new reference table.

6. Click Next.

7. Enter Reftab_CustTier_HypoStores as the table name.

8. Enter a description and set 0 as the default value.
   The Analyst tool uses the default value for any table record that does not contain a value.

9. Click Next.

10. In the Column Attributes panel, configure the following column properties for the CustomerTier column:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>CustomerTier</td>
</tr>
<tr>
<td>Datatype</td>
<td>String</td>
</tr>
<tr>
<td>Precision</td>
<td>10</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
</tr>
<tr>
<td>Description</td>
<td>Reference customer tier values</td>
</tr>
</tbody>
</table>

11. Optionally, choose to create a description column for rows in the reference table. Enter the name and precision for the column.

12. Preview the CustomerTier column values in the Preview panel.

13. Click Next.
   The Reftab_CustomerTier_HypoStores reference table name appears. You can enter an optional description.

14. In the Save in panel, select your tutorial project where you want to create the reference table.
   The Reference Tables: panel lists the reference tables in the location you select.
15. Enter an optional audit note.
16. Click Finish.

Task 2. Edit the Reference Table

In this task, you edit the Reftab_CustomerTier_HypoStores table to add alternate values for the customer tiers.

1. In the Navigator, select the Customers folder in your tutorial project.
2. Click the Reftab_CustomerTier_HypoStores reference table.
   The reference table opens in a tab.
3. To edit a row, select the row and click Actions > Editor click the Edit icon.
   The Edit Row window appears. Optionally, select multiple rows to add the same alternate value to each row.
4. Enter the following alternate values for the Diamond, Emerald, Gold, Silver, and Bronze rows: 1, 2, 3, 4, 5.
   Enter an optional audit note.
5. Click Apply to apply the changes.

Creating Reference Tables from Profile Columns

Summary

In this lesson, you learned how to create reference tables from the results of a profile to configure valid values for source data.

You created a reference table from a profile column by selecting valid values for columns. You edited the reference table to configure different valid values for columns.
This chapter includes the following topics:

- Creating Reference Tables Overview, 33
- Task 1. Create a Reference Table, 34
- Creating Reference Tables Summary, 34

Creating Reference Tables Overview

A reference table contains reference data that you can use to standardize source data. Reference data can include valid and standard values. Create reference tables to establish relationships between the source data values and the valid and standard values.

You can manually create a reference table using the reference table editor. Use the reference table to define and standardize the source data. You can share the reference table with a developer to use in Standardizer and Lookup transformations in the Developer tool.

Story

HypoStores wants to standardize data with valid values. You are the analyst who is responsible for standardizing the valid values in the data. You want to create a reference table to define standard customer tier codes that reference the LA customer data. You can then share the reference table with a developer.

Objectives

In this lesson, you complete the following task:

- Create a reference table using the reference table editor to define standard customer tier codes that reference the LA customer data.

Prerequisites

Before you start this lesson, verify the following prerequisite:

- You have completed lessons 1 and 2 in this tutorial.

Timing

Set aside 10 minutes to complete the task in this lesson.
Task 1. Create a Reference Table

In this task, you will create the Reftab_CustomerTier_Codes reference table to standardize the valid values for the customer tier data.

1. In the Navigator, select the Customer folder in your tutorial project where you want to create the reference table.
2. Click Actions > New Reference Table.
   The New Reference Table wizard appears.
3. Select the option to Use the reference table editor.
4. Click Next.
5. Enter the Reftab_CustomerTier_Codes as the table name and enter an optional description and set the default value of 0.
The Analyst tool uses the default value for any table record that does not contain a value.
6. For each column you want to include in the reference table, click the Add New Column icon and configure the column properties for each column.
   Add the following column names: CustomerID, CustomerTier, and Status. You can reorder the columns or delete columns.
7. Click Finish.
8. Open the Reftab_CustomerTier_Codes reference table and click Actions > Add Row to populate each reference table column with four values.
   CustomerID = LA1, LA2, LA3, LA4
   CustomerTier = 1, 2, 3, 4, 5.
   Status= Active, Inactive

Creating Reference Tables Summary

In this lesson, you learned how to create reference tables using the reference table editor to create standard valid values to use with source data.

You created a reference table using the reference table editor to standardize the customer tier values for the LA customer data.
Part II: Getting Started with Informatica Developer (Data Quality)

This part contains the following chapters:

- Lesson 1. Setting Up Informatica Developer, 36
- Lesson 2. Importing Physical Data Objects, 40
- Lesson 3. Profiling Data, 44
Lesson 1. Setting Up Informatica Developer

This chapter includes the following topics:
- Setting Up Informatica Developer Overview, 36
- Task 1. Start Informatica Developer, 37
- Task 2. Add a Domain, 37
- Task 3. Add a Model Repository, 38
- Task 4. Create a Project, 38
- Task 5. Create a Folder, 38
- Task 6. Select a Default Data Integration Service, 39
- Setting Up Informatica Developer Summary, 39

Setting Up Informatica Developer Overview

Before you start the lessons in this tutorial, you must start and set up the Developer tool. To set up the Developer tool, you add a domain. You add a Model repository that is in the domain, and you create a project and folder to store your work. You also select a default Data Integration Service.

The Informatica domain is a collection of nodes and services that define the Informatica environment. Services in the domain include the Model Repository Service and the Data Integration Service.

The Model Repository Service manages the Model repository. The Model repository is a relational database that stores the metadata for projects that you create in the Developer tool. A project stores objects that you create in the Developer tool. A project can also contain folders that store related objects, such as objects that are part of the same business requirement.

The Data Integration Service performs data integration tasks in the Developer tool.

Objectives

In this lesson, you complete the following tasks:
- Start the Developer tool and go to the Developer tool workbench.
- Add a domain in the Developer tool.
- Add a Model repository so that you can create a project.
- Create a project to store the objects that you create in the Developer tool.
Create a folder in the project that can store related objects.
Select a default Data Integration Service to perform data integration tasks.

**Prerequisites**
Before you start this lesson, verify the following prerequisites:

- You have installed the Developer tool.
- You have a domain name, host name, and port number to connect to a domain. You can get this information from a domain administrator.
- A domain administrator has configured a Model Repository Service in the Administrator tool.
- You have a user name and password to access the Model Repository Service. You can get this information from a domain administrator.
- A domain administrator has configured a Data Integration Service.
- The Data Integration Service is running.

**Timing**
Set aside 5 to 10 minutes to complete the tasks in this lesson.

---

**Task 1. Start Informatica Developer**

Start the Developer tool to begin the tutorial.

1. Start the Developer tool.
   - The Welcome page of the Developer tool appears.
2. Click the Workbench button.
   - The Developer tool workbench appears.

**Task 2. Add a Domain**

In this task, you add a domain in the Developer tool to access a Model repository.

1. Click Window > Preferences.
   - The Preferences dialog box appears.
2. Select Informatica > Domains.
3. Click Add.
   - The New Domain dialog box appears.
4. Enter the domain name, host name, and port number.
5. Click Finish.
6. Click OK.
Task 3. Add a Model Repository

In this task, you add the Model repository that you want to use to store projects and folders.

1. Click File > Connect to Repository.
   The Connect to Repository dialog box appears.
2. Click Browse to select a Model Repository Service.
3. Click OK.
4. Click Next.
5. Enter your user name and password.
6. Click Finish.
   The Model repository appears in the Object Explorer view.

Task 4. Create a Project

In this task, you create a project to store objects that you create in the Developer tool. You can create one project for all tutorials in this guide.

1. In the Object Explorer view, select a Model Repository Service.
2. Click File > New > Project.
   The New Project dialog box appears.
3. Enter your name prefixed by "Tutorial_" as the name of the project.
4. Click Finish.
   The project appears under the Model Repository Service in the Object Explorer view.

Task 5. Create a Folder

In this task, you create a folder to store related objects. You can create one folder for all tutorials in this guide.

1. In the Object Explorer view, select the project that you want to add the folder to.
2. Click File > New > Folder.
3. Enter a name for the folder.
4. Click Finish.
   The Developer tool adds the folder under the project in the Object Explorer view. Expand the project to see the folder.
Task 6. Select a Default Data Integration Service

In this task, you select a default Data Integration Service so you can run mappings and preview data.

1. Click **Window > Preferences**.
   The **Preferences** dialog box appears.
2. Select **Informatica > Data Integration Services**.
3. Expand the domain.
4. Select a Data Integration Service.
5. Click **Set as Default**.
6. Click **OK**.

Setting Up Informatica Developer Summary

In this lesson, you learned that the Informatica domain includes the Model Repository Service and Data Integration Service. The Model Repository Service manages the Model repository. A Model repository contains projects and folders. The Data Integration Service performs data integration tasks.

You started the Developer tool and set up the Developer tool. You added a domain to the Developer tool, added a Model repository, and created a project and folder. You also selected a default Data Integration Service.

Now, you can use the Developer tool to complete other lessons in this tutorial.
Lesson 2. Importing Physical Data Objects

This chapter includes the following topics:

- Importing Physical Data Objects Overview, 40
- Task 1. Import the Boston_Customers Flat File Data Object, 41
- Task 2. Import the LA_Customers Flat File Data Object, 41
- Task 3. Importing the All_Customers Flat File Data Object, 42
- Importing Physical Data Objects Summary, 43

Importing Physical Data Objects Overview

A physical data object is a representation of data based on a flat file or relational database table. You can import a flat file or relational database table as a physical data object to use as a source or target in a mapping.

Story

HypoStores Corporation stores customer data from the Los Angeles office and Boston office in flat files. You want to work with this customer data in the Developer tool. To do this, you need to import each flat file as a physical data object.

Objectives

In this lesson, you import flat files as physical data objects. You also set the source file directory so that the Data Integration Service can read the source data from the correct directory.

Prerequisites

Before you start this lesson, verify the following prerequisite:

- You have completed lesson 1 in this tutorial.

Timing

Set aside 10 to 15 minutes to complete the tasks in this lesson.
Task 1. Import the Boston_Customers Flat File Data Object

In this task, you import a physical data object from a file that contains customer data from the Boston office.

1. In the Object Explorer view, select the tutorial project.
2. Click File > New > Data Object.
   The New dialog box appears.
3. Select Physical Data Objects > Flat File Object and click Next.
   The New Flat File Object dialog box appears.
4. Select Create from an Existing Flat File.
5. Click Browse and navigate to Boston_Customers.csv in the following directory: <Informatica Installation Directory>\clients\DeveloperClient\Tutorials
6. Click Open.
   The wizard names the data object Boston_Customers.
7. Click Next.
8. Verify that the code page is MS Windows Latin 1 (ANSI), superset of Latin 1.
9. Verify that the format is delimited.
10. Click Next.
11. Verify that the delimiter is set to comma.
12. Select Import column names from first line.
13. Click Finish.
   The Boston_Customers physical data object appears under Physical Data Objects in the tutorial project.
14. Click the Read view and select the Output transformation.
15. Click the Runtime tab on the Properties view.
16. Set the Source File Directory to the following directory on the Data Integration Service machine: <Informatica Installation Directory>\server\Tutorials
17. Click File > Save.

Task 2. Import the LA_Customers Flat File Data Object

In this task, you import a physical data object from a flat file that contains customer data from the Los Angeles office.

1. In the Object Explorer view, select the tutorial project.
2. Click File > New > Data Object.
   The New dialog box appears.
3. Select Physical Data Objects > Flat File Data Object and click Next.
   The New Flat File Data Object dialog box appears.
4. Select Create from an Existing Flat File.
5. Click **Browse** and navigate to LA_Customers.csv in the following directory: `<Informatica Installation Directory>\clients\DeveloperClient\Tutorials`

6. Click **Open**.
   The wizard names the data object LA_Customers.

7. Click **Next**.

8. Verify that the code page is MS Windows Latin 1 (ANSI), superset of Latin 1.

9. Verify that the format is delimited.

10. Click **Next**.

11. Verify that the delimiter is set to comma.

12. Select **Import column names from first line**.

13. Click **Finish**.
   The LA_Customers physical data object appears under **Physical Data Objects** in the tutorial project.

14. Click the **Read** view and select the Output transformation.

15. Click the **Runtime** tab on the **Properties** view.

16. Set the Source File Directory to the following directory on the Data Integration Service machine: `<Informatica Installation Directory>\server\Tutorials`

17. Click **File > Save**.

---

**Task 3. Importing the All_Customers Flat File Data Object**

In this task, you import a physical data object from a flat file that combines the customer order data from the Los Angeles and Boston offices.

1. In the **Object Explorer** view, select the tutorial project.

2. Click **File > New > Data Object**.
   The **New** dialog box appears.

3. Select **Physical Data Objects > Flat File Data Object** and click **Next**.
   The **New Flat File Data Source** dialog box appears.

4. Select **Create from an Existing Flat File**.

5. Click **Browse** and navigate to All_Customers.csv in the following directory: `<Informatica Installation Directory>\clients\DeveloperClient\Tutorials`.

6. Click **Open**.
   The wizard names the data object All_Customers.

7. Click **Next**.

8. Verify that the code page is MS Windows Latin 1 (ANSI), superset of Latin 1.

9. Verify that the format is delimited.

10. Click **Next**.

11. Verify that the delimiter is set to comma.

12. Select **Import column names from first line**.
13. Click **Finish**.
   The All_Customers physical data object appears under **Physical Data Objects** in the tutorial project.

14. Click the **Read** view and select the Output transformation.

15. Click the **Runtime** tab on the **Properties** view.

16. Set the Source File Directory to the following directory on the Data Integration Service machine: `<Informatica Installation Directory>\server\Tutorials`

17. Click **File > Save**.

### Importing Physical Data Objects Summary

In this lesson, you learned that physical data objects are representations of data based on a flat file or a relational database table.

You created physical data objects from flat files. You also set the source file directory so that the Data Integration Service can read the source data from the correct directory.

You use the data objects as mapping sources in the data quality lessons.
Lesson 3. Profiling Data

This chapter includes the following topics:
- Profiling Data Overview, 44
- Task 1. Perform a Join Analysis on Two Data Sources, 45
- Task 2. View Join Analysis Results, 46
- Task 3. Run a Profile on a Data Source, 46
- Task 4. View Column Profiling Results, 47
- Profiling Data Summary, 47

Profiling Data Overview

A profile is a set of metadata that describes the content and structure of a dataset.

Data profiling is often the first step in a project. You can run a profile to evaluate the structure of data and verify that data columns are populated with the types of information you expect. If a profile reveals problems in data, you can define steps in your project to fix those problems. For example, if a profile reveals that a column contains values of greater than expected length, you can design data quality processes to remove or fix the problem values.

A profile that analyzes the data quality of selected columns is called a column profile.

**Note:** You can also use the Developer tool to discover primary key, foreign key, and functional dependency relationships, and to analyze join conditions on data columns.

A column profile provides the following facts about data:
- The number of unique and null values in each column, expressed as a number and a percentage.
- The patterns of data in each column, and the frequencies with which these values occur.
- Statistics about the column values, such as the maximum and minimum lengths of values and the first and last values in each column.
- For join analysis profiles, the degree of overlap between two data columns, displayed as a Venn diagram and as a percentage value. Use join analysis profiles to identify possible problems with column join conditions.

You can run a column profile at any stage in a project to measure data quality and to verify that changes to the data meet your project objectives. You can run a column profile on a transformation in a mapping to indicate the effect that the transformation will have on data.
**Story**

HypoStores wants to verify that customer data is free from errors, inconsistencies, and duplicate information. Before HypoStores designs the processes to deliver the data quality objectives, it needs to measure the quality of its source data files and confirm that the data is ready to process.

**Objectives**

In this lesson, you complete the following tasks:

- Perform a join analysis on the Boston_Customers data source and the LA_Customers data source.
- View the results of the join analysis to determine whether or not you can successfully merge data from the two offices.
- Run a column profile on the All_Customers data source.
- View the column profiling results to observe the values and patterns contained in the data.

**Prerequisites**

Before you start this lesson, verify the following prerequisite:

- You have completed lessons 1 and 2 in this tutorial.

**Time Required**

- Set aside 20 minutes to complete this lesson.

---

**Task 1. Perform a Join Analysis on Two Data Sources**

In this task, you perform a join analysis on the Boston_Customers and LA_Customers data sources to view the join conditions.

1. In the Object Explorer view, browse to the data objects in your tutorial project.
2. Select the Boston_Customers and LA_Customers data sources.
   - **Tip:** Hold down the Shift key to select multiple data objects.
3. Right-click the selected data objects and select **Profile**.
   - The **New Profile** wizard opens.
4. Select **Profile Model**, and click **Next**.
5. In the **Name** field, enter **Tutorial_Model**.
   - Click **Next**.
6. Verify that Boston_Customers and LA_Customers appear in the **Data Object** column.
7. Click **Finish**.
   - The Tutorial_Model profile model appears in the Object Explorer.
8. Use your mouse to select Boston_Customers and LA_Customers in the modeling canvas.
9. Right-click a data object name and select **Join Profile**.
   - The **New Join Profile** wizard opens.
10. In the **Name** field, enter **JoinAnalysis**.
11. Verify that Boston_Customers and LA_Customers appear as data objects.
   - Click **Next**.
12. Select the CustomerID column in both data sources. Scroll down the wizard pane to view the columns in both data sets. Click Next.

13. Click Add to add join conditions. The Join Condition window opens.

14. In the Columns section, click the New button.

15. Double-click the first row in the left column and select CustomerID.

16. Double-click the first row in the right column and select CustomerID.

17. Click OK, and click Finish. The JoinAnalysis profile opens in the editor and the profile runs.

Note: Do not close the profile. You view the profile results in the next task.

Task 2. View Join Analysis Results

In this task, you view the join analysis results in the Join Results view of the JoinAnalysis profile.

1. Click the JoinAnalysis tab in the modeling canvas.

2. In the Join Profile section, click the first row. The Details section displays a Venn diagram and a key that details the results of the join analysis.

3. Verify that the Join Rows column reports zero as the number of rows that contain a join. This indicates that none of the CustomerID fields are duplicates, suggesting you can successfully merge the two data sources.

4. To view the CustomerID values for the LA_Customers data object, double-click the circle labeled LA_Customers in the Venn diagram.

Tip: Double-click the circles in the Venn diagram to view the data rows described by these items. In cases where circles intersect in the Venn diagram, double-click the intersection to view data values common to both data sets.

The Data Viewer displays the CustomerID values contained in the LA_Customers data object.

Task 3. Run a Profile on a Data Source

In this task, you run a profile on the All_Customers data source to view the content and structure of the data.

1. In the Object Explorer view, browse to the data objects in your tutorial project.

2. Select the All_Customers data source.


4. In the Name field, enter All_Customers.

5. Click Finish. The All_Customers profile opens in the editor and the profile runs.
Task 4. View Column Profiling Results

In this task, you view the column profiling results for the All_Customers data object and examine the values and patterns contained in the data.

1. Click **Window > Show View > Progress** to view the progress of the All_Customers profile.

   The Progress view opens.

2. When the Progress view reports that the All_Customers profile finishes running, click the Results view in the editor.

3. In the **Column Profiling** section, click the **CustomerTier** column.

   The Details section displays all values contained in the CustomerTier column and displays information about how frequently the values occur in the dataset.

4. In the Details section, double-click the value **Ruby**.

   The Data Viewer runs and displays the records where the CustomerTier column contains the value **Ruby**.

5. In the **Column Profiling** section, click the **OrderAmount** column.

6. In the Details section, click the **Show** list and select **Patterns**.

   The Details section shows the patterns found in the OrderAmount column. The string **9(5)** in the Pattern column refers to records that contain five-figure order amounts. The string **9(4)** refers to records containing four-figure amounts.

7. In the **Pattern** column, double-click the string **9(4)**.

   The Data Viewer runs and displays the records where the OrderAmount column contains a four-figure order amount.

8. In the Details section, click the **Show** list and select **Statistics**.

   The Details section shows statistics for the OrderAmount column, including the average value, the standard deviation, maximum and minimum lengths, the five most common values, and the five least common values.

Profiling Data Summary

In this lesson, you learned that a profile provides information about the content and structure of the data.

You learned that you can perform a join analysis on two data objects and view the degree of overlap between the data objects. You also learned that you can run a column profile on a data object and view values, patterns, and statistics that relate to each column in the data object.

You created the **JoinAnalysis** profile to determine whether data from the Boston_Customers data object can merge with the data in the LA_Customers data object. You viewed the results of this profile and determined that all values in the CustomerID column are unique and that you can merge the data objects successfully.

You created the All_Customers profile and ran a column profile on the All_Customers data object. You viewed the results of this profile to discover values, patterns, and statistics for columns in the All_Customers data object. Finally, you ran the Data Viewer to view rows containing values and patterns that you selected, enabling you to verify the quality of the data.
Frequently Asked Questions

This appendix includes the following topics:

- Informatica Administrator Frequently Asked Questions, 48
- Informatica Analyst FAQ, 49
- Informatica Developer Frequently Asked Questions, 49

Informatica Administrator Frequently Asked Questions

Review the frequently asked questions to answer questions you may have about Informatica Administrator.

**What is the difference between Informatica Administrator and the PowerCenter Administration Console?**

The PowerCenter Administration Console is renamed to Informatica Administrator (the Administrator tool).

The Administrator tool has a new interface. Some of the properties and configuration tasks from the PowerCenter Administration Console have been moved to different locations in the Administrator tool. The Administrator tool is also expanded to include new services and objects.

**Can I use one user account to access the Administrator tool and the Developer tool?**

Yes. You can give a user permission to access both tools. You do not need to create separate user accounts for each application client.

**What is the difference between the PowerCenter Repository Service and the Model Repository Service?**

The PowerCenter application services and PowerCenter application clients use the PowerCenter Repository Service. The PowerCenter repository has folder-based security.

The other application services, such as the Data Integration Service, Analyst Service, Developer tool, and Analyst tool, use the Model Repository Service. The Model Repository Service has project-based security.

You can migrate some Model repository objects to the PowerCenter repository.

**Where can I use database connections that I create in the Informatica Administrator?**

You can create, view, and edit database connections in the Administrator tool and the Developer tool. You can create and view database connections in the Analyst tool. You can also configure database connection pooling in the Administrator tool.

**What is the difference between the PowerCenter Integration Service and the Data Integration Service?**

The PowerCenter Integration Service is an application service that runs sessions and workflows. The Data Integration Service is an application service that performs data integration tasks for the Analyst tool, the Developer tool, and external clients. The Analyst tool and the Developer tool send data integration task requests to the Data Integration Service to preview or run data profiles, SQL data services, and mappings.
Commands from the command line or an external client send data integration task requests to the Data Integration Service to run SQL data services or web services.

**Why can I not connect to an SQL data service that is deployed to a Data Integration Service?**

To connect to an SQL data service, the application that contains the SQL data service must be running. You can start an application in the **Applications** view of the Data Integration Service.

---

**Informatica Analyst FAQ**

Review the FAQ to answer questions you may have about Informatica Analyst.

**What combination of product functionality is included in the Analyst tool?**

The Analyst tools contains some functionality that is included in the following products:

- Informatica Data Quality
- Informatica Data Explorer Advanced Edition
- Data Quality Assistant
- PowerCenter Reference Table Manager

**Can I use one user account to access the Administrator tool, the Developer tool, and the Analyst Tool?**

Yes. You can give a user permission to access all three tools. You do not need to create separate user accounts for each client application.

**What happened to the Reference Table Manager? Where is my reference data stored?**

The functionality from the Reference Table Manager is included in the Analyst tool. You can use the Analyst tool to create and share reference data. The reference data is stored in the staging database that you configure when you create an Analyst Service.

---

**Informatica Developer Frequently Asked Questions**

Review the frequently asked questions to answer questions you may have about Informatica Developer.

**What is the difference between a source and target in PowerCenter and a physical data object in the Developer tool?**

In PowerCenter, you create a source definition to include as a mapping source. You create a target definition to include as a mapping target. In the Developer tool, you create a physical data object that you can use as a mapping source or target.

**What is the difference between a mapping in the Developer tool and a mapping in PowerCenter?**

A PowerCenter mapping specifies how to move data between sources and targets. A Developer tool mapping specifies how to move data between the mapping input and output.

A PowerCenter mapping must include one or more source definitions, source qualifiers, and target definitions. A PowerCenter mapping can also include shortcuts, transformations, and mapplets.

A Developer tool mapping must include mapping input and output. A Developer tool mapping can also include transformations and mapplets.
The Developer tool has the following types of mappings:

- Mapping that moves data between sources and targets. This type of mapping differs from a PowerCenter mapping only in that it cannot use shortcuts and does not use a source qualifier.
- Logical data object mapping. A mapping in a logical data object model. A logical data object mapping can contain a logical data object as the mapping input and a data object as the mapping output. Or, it can contain one or more physical data objects as the mapping input and logical data object as the mapping output.
- Virtual table mapping. A mapping in an SQL data service. It contains a data object as the mapping input and a virtual table as the mapping output.
- Virtual stored procedure mapping. Defines a set of business logic in an SQL data service. It contains an Input Parameter transformation or physical data object as the mapping input and an Output Parameter transformation or physical data object as the mapping output.

What is the difference between a mapplet in PowerCenter and a mapplet in the Developer tool?

A mapplet in PowerCenter and in the Developer tool is a reusable object that contains a set of transformations. You can reuse the transformation logic in multiple mappings.

A PowerCenter mapplet can contain source definitions or Input transformations as the mapplet input. It must contain Output transformations as the mapplet output.

A Developer tool mapplet can contain data objects or Input transformations as the mapplet input. It can contain data objects or Output transformations as the mapplet output. A mapping in the Developer tool also includes the following features:

- You can validate a mapplet as a rule. You use a rule in a profile.
- A mapplet can contain other mapplets.

What is the difference between a mapplet and a rule?

You can validate a mapplet as a rule. A rule is business logic that defines conditions applied to source data when you run a profile. You can validate a mapplet as a rule when the mapplet meets the following requirements:

- It contains an Input and Output transformation.
- The mapplet does not contain active transformations.
- It does not specify cardinality between input groups.