



Informatica® Data Quality
10.1.1

Guía de introducción

© Copyright Informatica LLC 1998, 2018

Este software y esta documentación contienen información privada de Informatica LLC y se facilitan bajo un acuerdo de licencia que incluye restricciones sobre el uso y la divulgación y también están protegidos por leyes de copyright. Está prohibida la ingeniería inversa del software. Ninguna parte de este documento puede ser reproducida o transmitida de cualquier forma o manera (electrónica, fotocopia, grabación o mediante otros métodos) sin el consentimiento previo de Informatica LLC. Este Software puede estar protegido por Patentes de los Estados Unidos y/o patentes internacionales y otras patentes pendientes.

El uso, la duplicación o la divulgación del software por parte del gobierno de los Estados Unidos están sujetos a las restricciones puntualizadas en el acuerdo de licencia de software vigente y según se establece en la reglamentación DFARS 227.7202-1(a) y 227.7702-3(a) (1995), DFARS 252.227-7013(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19 o FAR 52.227-14 (ALT III), según proceda.

La información contenida en este producto o documentación está sujeta a cambios sin previo aviso. Si encuentra algún problema en este producto o documentación, infórmenos por escrito.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging, Informatica Master Data Management y Live Data Map son marcas comerciales o marcas comerciales registradas de Informatica LLC en Estados Unidos y en las jurisdicciones de todo el mundo. Los demás nombres de productos y empresas pueden ser nombres o marcas comerciales de sus respectivos titulares.

Hay fragmentos de este software y/o documentación que están sujetas a copyright perteneciente a terceros, incluido, entre otros: Copyright DataDirect Technologies. Todos los derechos reservados. Copyright © Sun Microsystems. Todos los derechos reservados. Copyright © RSA Security Inc. Todos los derechos reservados. Copyright © Ordinal Technology Corp. Todos los derechos reservados. Copyright © Aandacht c.v. Todos los derechos reservados. Copyright Genivia, Inc. Todos los derechos reservados. Copyright Isomorphic Software. Todos los derechos reservados. Copyright © Meta Integration Technology, Inc. Todos los derechos reservados. Copyright © Intalio. Todos los derechos reservados. Copyright © Oracle. Todos los derechos reservados. Copyright © Adobe Systems Incorporated. Todos los derechos reservados. Copyright © DataArt, Inc. Todos los derechos reservados. Copyright © ComponentSource. Todos los derechos reservados. Copyright © Microsoft Corporation. Todos los derechos reservados. Copyright © Rogue Wave Software, Inc. Todos los derechos reservados. Copyright © Teradata Corporation. Todos los derechos reservados. Copyright © Yahoo! Inc. Todos los derechos reservados. Copyright © Glyph & Cog, LLC. Todos los derechos reservados. Copyright © Thinkmap, Inc. Todos los derechos reservados. Copyright © Clearpace Software Limited. Todos los derechos reservados. Copyright © Information Builders, Inc. Todos los derechos reservados. Copyright © OSS Nokalva, Inc. Todos los derechos reservados. Copyright Edifecs, Inc. Todos los derechos reservados. Copyright Cleo Communications, Inc. Todos los derechos reservados. Copyright © International Organization for Standardization 1986. Todos los derechos reservados. Copyright © ej-technologies GmbH. Todos los derechos reservados. Copyright © Jaspersoft Corporation. Todos los derechos reservados. Copyright © International Business Machines Corporation. Todos los derechos reservados. Copyright © yWorks GmbH. Todos los derechos reservados. Copyright © Lucent Technologies. Todos los derechos reservados. Copyright © University of Toronto. Todos los derechos reservados. Copyright © Daniel Veillard. Todos los derechos reservados. Copyright © Unicode, Inc. Copyright IBM Corp. Todos los derechos reservados. Copyright © MicroQuill Software Publishing, Inc. Todos los derechos reservados. Copyright © PassMark Software Pty Ltd. Todos los derechos reservados. Copyright © LogiXML, Inc. Todos los derechos reservados. Copyright © 2003-2010 Lorenzi Davide. Todos los derechos reservados. Copyright © Red Hat, Inc. Todos los derechos reservados. Copyright © The Board of Trustees of the Leland Stanford Junior University. Todos los derechos reservados. Copyright © EMC Corporation. Todos los derechos reservados. Copyright © Flexera Software. Todos los derechos reservados. Copyright © Jinfonet Software. Todos los derechos reservados. Copyright © Apple Inc. Todos los derechos reservados. Copyright © Telerik Inc. Todos los derechos reservados. Copyright © BEA Systems. Todos los derechos reservados. Copyright © PDFlib GmbH. Todos los derechos reservados. Copyright © Orientation in Objects GmbH. Todos los derechos reservados. Copyright © Tanuki Software, Ltd. Todos los derechos reservados. Copyright © Ricebridge. Todos los derechos reservados. Copyright © Sencha, Inc. Todos los derechos reservados. Copyright © Scalable Systems, Inc. Todos los derechos reservados. Copyright © jQWidgets. Todos los derechos reservados. Copyright © Tableau Software, Inc. Todos los derechos reservados. Copyright © MaxMind, Inc. Todos los derechos reservados. Copyright © TMate Software s.r.o. Todos los derechos reservados. Copyright © MapR Technologies Inc. Todos los derechos reservados. Copyright © Amazon Corporate LLC. Todos los derechos reservados. Copyright © Highsoft. Todos los derechos reservados. Copyright © Python Software Foundation. Todos los derechos reservados. Copyright © BeOpen.com. Todos los derechos reservados. Copyright © CNRI. Todos los derechos reservados.

Este producto incluye software desarrollado por la Apache Software Foundation (<http://www.apache.org/>) y/u otro software protegido por varias versiones de la licencia Apache License ("Licencia"). Puede obtener una copia de estas licencias en <http://www.apache.org/licenses/>. A menos que las leyes aplicables lo requieran o se haya acordado por escrito, el software distribuido bajo estas licencias se distribuye "TAL CUAL", SIN GARANTÍAS NI CONDICIONES DE NINGÚN TIPO, ya sea expresas o implícitas. Consulte las licencias del idioma específico para conocer los permisos y las limitaciones que rigen según las licencias.

Este producto incluye software desarrollado por Mozilla (<http://www.mozilla.org/>), copyright del software de The JBoss Group, LLC, todos los derechos reservados; copyright del software © 1999-2006 de Bruno Lowagie y Paulo Soares y otro software protegido con licencia por el acuerdo GNU Lesser General Public License Agreement, que se puede encontrar en la dirección <http://www.gnu.org/licenses/lgpl.html>. Los materiales se facilitan gratuitamente por parte de Informatica, "tal cual", sin garantía de ningún tipo, ya sea expresa o implícita, incluidas, entre otras, las garantías implícitas de adecuación para un propósito determinado y de validez para el comercio.

El producto incluye software ACE(TM) y TAO(TM) con copyright de Douglas C. Schmidt y su grupo de investigación de la Washington University, University of California, Irvine y Vanderbilt University, Copyright (©) 1993-2006, todos los derechos reservados.

Este producto incluye software desarrollado por el OpenSSL Project para uso en el OpenSSL Toolkit (copyright The OpenSSL Project. Todos los derechos reservados) y la redistribución de este software está sujeta a los términos especificados en <http://www.openssl.org> y <http://www.openssl.org/source/license.html>.

Este producto incluye software Curl con Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. Todos los derechos reservados. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección <http://curl.haxx.se/docs/copyright.html>. La autorización para utilizar, copiar, modificar y distribuir este software para cualquier propósito con o sin tasas se concede por el presente, siempre que el aviso de copyright anterior y este aviso de permiso aparezcan en todas las copias.

El producto incluye copyright de software 2001-2005 (©) MetaStuff, Ltd. Todos los derechos reservados. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección <http://www.dom4j.org/license.html>.

El producto incluye copyright de software © 2004-2007, The Dojo Foundation. Todos los derechos reservados. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección <http://dojotoolkit.org/license>.

Este producto incluye software ICU con copyright de International Business Machines Corporation y otros. Todos los derechos reservados. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

Este producto incluye copyright de software © 1996-2006 Per Bothner. Todos los derechos reservados. Su derecho a utilizar estos materiales está establecido en la licencia que puede encontrarse en la dirección <http://www.gnu.org/software/kawa/Software-License.html>.

Este producto incluye software OSSP UUID con Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Los permisos y las limitaciones relativas a este software están sujetos a los términos disponibles en la dirección <http://www.opensource.org/licenses/mit-license.php>.

Este producto incluye software desarrollado por Boost (<http://www.boost.org/>) o protegido por la licencia de software de Boost. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección http://www.boost.org/LICENSE_1_0.txt.

Este producto incluye copyright de software © 1997-2007 University of Cambridge. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección <http://www.pcre.org/license.txt>.

Este producto incluye copyright de software © 2007 The Eclipse Foundation. Todos los derechos reservados. Los permisos y las limitaciones relativos a este software están sujetos a los términos especificados en <http://www.eclipse.org/org/documents/epl-v10.php> y <http://www.eclipse.org/org/documents/edl-v10.php>.

Este producto incluye software protegido por licencia según los términos que aparecen en <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqldbLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>, <http://antlr.org/license.html>, <http://aopalliance.sourceforge.net/>, <http://www.bouncycastle.org/license.html>, <http://www.jgraph.com/jgraphdownload.html>, <http://www.jcraft.com/jsch/LICENSE.txt>, http://jotm.objectweb.org/bsd_license.html, <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>, <http://www.slf4j.org/license.html>, <http://nanoxml.sourceforge.net/orig/copyright.html>, <http://www.json.org/license.html>, <http://forge.ow2.org/projects/jvaservice/>, <http://www.postgresql.org/about/license.html>, <http://www.sqlite.org/copyright.html>, <http://www.tcl.tk/software/tcltk/license.html>, <http://www.jaxen.org/faq.html>, <http://www.jdom.org/docs/faq.html>, <http://www.slf4j.org/license.html>, <http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>, <http://www.keplerproject.org/md5/license.html>, <http://www.toedter.com/en/jcalendar/license.html>, <http://www.edankert.com/bounce/index.html>, <http://www.net-snmp.org/about/license.html>, <http://www.openmdx.org/#FAQ>, http://www.php.net/license/3_01.txt, <http://srp.stanford.edu/license.txt>, <http://www.schneier.com/blowfish.html>, <http://www.jmock.org/license.html>, <http://xsom.java.net>, <http://benalman.com/about/license/>, <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>, <http://www.h2database.com/html/license.html#summary>, <http://jsoncpp.sourceforge.net/LICENSE>, <http://jdbc.postgresql.org/license.html>, <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>, <https://github.com/rantav/hector/blob/master/LICENSE>, <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>, <http://jibx.sourceforge.net/jibx-license.html>, <https://github.com/yokato/libgeohash/blob/master/LICENSE>, <https://github.com/hjiang/sonxx/blob/master/LICENSE>, <https://code.google.com/p/lz4/>, <https://github.com/jedisct1/libsodium/blob/master/LICENSE>, <http://one-jar.sourceforge.net/index.php?page=documents&file=license>, <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>, <http://www.scala-lang.org/license.html>, <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>, <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>, <https://aws.amazon.com/asl/>, <https://github.com/twbs/bootstrap/blob/master/LICENSE>, <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>, <https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE> y <https://github.com/apache/hbase/blob/master/LICENSE.txt>.

Este producto incluye software desarrollado por la Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), la Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>), la Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), la Sun Binary Code License Agreement Supplemental License Terms, la BSD License (<http://www.opensource.org/licenses/bsd-license.php>), la nueva BSD License (<http://opensource.org/licenses/BSD-3-Clause>), la MIT License (<http://www.opensource.org/licenses/mit-license.php>), la Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) y la Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

Este producto incluye copyright de software © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. Todos los derechos reservados. Los permisos y las limitaciones relativos a este software están sujetos a los términos disponibles en la dirección <http://xstream.codehaus.org/license.html>. Este producto incluye software desarrollado por Indiana University Extreme! Lab. Para obtener más información, visite <http://www.extreme.indiana.edu/>.

Este producto incluye software Copyright © 2013 Frank Balluffi y Markus Moeller. Todos los derechos reservados. Los permisos y las limitaciones relativas a este software están sujetos a los términos de la licencia MIT.

Consulte las patentes en <https://www.informatica.com/legal/patents.html>.

EXENCIÓN DE RESPONSABILIDAD: Informatica LLC proporciona esta documentación "tal cual" sin garantía de ningún tipo, ya sea expresa o implícita, incluidas, entre otras, las garantías implícitas de no incumplimiento, de adecuación para un propósito determinado y de validez para el comercio. Informatica LLC no garantiza que este software o esta documentación estén libres de errores. La información proporcionada en este software o en esta documentación puede contener imprecisiones técnicas o errores tipográficos. La información de este software y esta documentación está sujeta a cambios en cualquier momento sin previo aviso.

AVISOS

Este producto de Informatica (el "Software") incluye ciertos controladores (los "Controladores DataDirect") de DataDirect Technologies, una empresa operativa de Progress Software Corporation ("DataDirect") que están sujetos a los términos y condiciones siguientes:

1. LOS CONTROLADORES DATADIRECT SE PROPORCIONAN "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUIDAS, ENTRE OTRAS, LAS GARANTÍAS IMPLÍCITAS DE NO INCUMPLIMIENTO, DE ADECUACIÓN PARA UN PROPÓSITO DETERMINADO Y DE VALIDEZ PARA EL COMERCIO.
2. EN NINGÚN CASO DATADIRECT NI SUS PROVEEDORES DE TERCEROS SERÁN RESPONSABLES ANTE EL USUARIO FINAL POR NINGÚN DAÑO DIRECTO, INDIRECTO, FORTUITO, ESPECIAL, CONSECUENTE, NI DE NINGÚN OTRO TIPO, RESULTANTE DEL USO DE LOS CONTROLADORES ODBC, INDEPENDIENTEMENTE DE SI SE HA AVISADO O NO DE LOS POSIBLES DAÑOS POR ADELANTADO. ESTAS LIMITACIONES SE APLICAN A TODAS LAS DEMANDAS JUDICIALES, INCLUIDAS, ENTRE OTRAS, AQUELLAS POR INCUMPLIMIENTO DE CONTRATO, INCUMPLIMIENTO DE LA GARANTÍA, NEGLIGENCIA, RESPONSABILIDAD ESTRICTA, TERGIVERSACIÓN Y OTROS AGRAVIOS.

Fecha de publicación: 2018-06-09

Tabla de contenido

Prefacio	8
Documentación de Informatica	8
Informatica Network.	8
Base de conocimiento de Informatica.	8
Documentación de Informatica	8
Matrices de disponibilidad de productos de Informatica.	9
Informatica Velocity.	9
Catálogo de soluciones de Informatica.	9
Servicio internacional de atención al cliente de Informatica.	9
 Capítulo 1: Introducción.....	10
Resumen del dominio de Informatica.	10
Disponibilidad de funciones.	12
Presentación de Informatica Analyst.	12
Introducción a Informatica Developer.	13
Página de bienvenida de Informatica Developer.	14
Referencias rápidas.	14
Data Quality y creación de perfiles.	15
Historial de la guía de autoaprendizaje.	15
Estructura de la guía de autoaprendizaje.	16
Guía de autoaprendizaje de Informatica Analyst.	16
Informatica Developer Tool.	17
Requisitos previos de la guía de autoaprendizaje.	18
 Parte I: Primeros pasos con Informatica Analyst.....	19
 Capítulo 2: Lección 1. Configuración de la herramienta Analyst de Informatica	20
Introducción a la configuración de Informatica Analyst.	20
Tarea 1. Inicio de sesión en Informatica Analyst.	21
Tarea 2. Creación de un proyecto.	21
Tarea 3. Creación de una carpeta.	22
Resumen de la configuración de Informatica Analyst.	22
 Capítulo 3: Lección 2. Cómo crear objetos de datos	23
Introducción a la creación de objetos de datos.	23
Tarea 1. Crear los objetos de datos de archivo sin formato.	24
Tarea 2. Visualización de las propiedades del objeto de datos.	24
Resumen de la creación de objetos de datos	25

Capítulo 4: Lección 3. Creación de perfiles predeterminados.	26
Introducción a la creación de perfiles predeterminados.	26
Tarea 1. Crear y ejecutar un perfil predeterminado.	27
Tarea 2. Ver los resultados del perfil en una vista de resumen.	28
Resumen de la creación de perfiles predeterminados.	28
 Capítulo 5: Lección 4. Creación de perfiles personalizados.	 30
Introducción a la creación de perfiles personalizados.	30
Tarea 1. Crear un perfil personalizado.	31
Tarea 2. Ejecutar el perfil.	32
Tarea 3. Obtener detalles de los resultados del perfil.	32
Resumen de la creación de perfiles personalizados.	33
 Capítulo 6: Lección 5. Creación de reglas de expresión.	 34
Introducción a la creación de reglas de expresión.	34
Tarea 1. Crear reglas de expresión y ejecutar el perfil.	35
Tarea 2. Ver la salida de la regla de expresión.	35
Tarea 3. Editar las reglas de expresión.	36
Resumen de la creación de reglas de expresión.	36
 Capítulo 7: Lección 6. Creación y ejecución de cuadros de mando.	 37
Introducción a la creación y ejecución de cuadros de mando.	37
Tarea 1. Crear un cuadro de mando a partir de los resultados de un perfil.	38
Tarea 2. Ejecutar el cuadro de mandos.	39
Tarea 3. Ver el cuadro de mandos.	40
Tarea 4. Editar el cuadro de mandos.	40
Tarea 5. Configurar umbrales.	41
Tarea 6. Ver gráficos de tendencias de puntuación.	41
Creación y ejecución de cuadros de mando.	42
 Capítulo 8: Lección 7. Creación de tablas de referencia a partir de columnas de perfil.	 43
Introducción a la creación de tablas de referencia a partir de columnas de perfil.	43
Tarea 1. Crear una tabla de referencia a partir de columnas de perfil.	44
Tarea 2. Editar la tabla de referencia.	45
Resumen de la creación de tablas de referencia a partir de columnas de perfil.	45
 Capítulo 9: Lección 8. Creación de tablas de referencia.	 46
Introducción a la creación de tablas de referencia.	46
Tarea 1. Crear una tabla de referencia.	47
Resumen de la creación de tablas de referencia.	47

Parte II: Primeros pasos con Informatica Developer.....	48
Capítulo 10: Lección 1. Configuración de Informatica Developer.	49
Introducción a la configuración de Informatica Developer.	49
Tarea 1. Iniciar Informatica Developer.	50
Tarea 2. Añadir un dominio.	50
Tarea 3. Añadir un repositorio de modelos.	51
Tarea 4. Crear un proyecto.	51
Tarea 5. Crear una carpeta.	51
Tarea 6. Seleccionar un servicio de integración de datos predeterminado..	52
Resumen de la configuración de Informatica Developer.	52
Capítulo 11: Lección 2: Importar objetos de datos físicos.	53
Introducción a la importación de objetos de datos físicos.	53
Tarea 1. Importar el objeto de datos del archivo sin formato Boston_Customers.	54
Tarea 2. Importar el objeto de datos de archivo sin formato LA_Customers.	60
Tarea 3. Importar el objeto de datos del archivo sin formato All_Customers.	61
Resumen de la importación de objetos de datos físicos.	61
Capítulo 12: Lección 3. Ejecutar un perfil con datos de origen.	62
Introducción a la creación de perfiles de datos.	62
Tarea 1. Realizar un análisis de combinación en dos orígenes de datos.	63
Tarea 2. Ver los resultados de un análisis de combinación.	64
Tarea 3. Ejecutar un perfil en un origen de datos.	65
Tarea 4. Ver los resultados de perfil de columna.	65
Resumen de datos de creación de perfiles.	66
Capítulo 13: Lección 4. Análisis de datos.	67
Resumen del análisis de datos.	67
Tarea 1. Crear un objeto de datos de destino.	68
Paso 1. Crear un objeto de datos LA_Customers.	68
Paso 2. Configurar opciones de lectura y escritura.	69
Paso 3. Añadir columnas al objeto de datos.	69
Tarea 2. Crear una asignación para analizar datos.	70
Paso 1. Crear una asignación.	70
Paso 2. Añadir objetos de datos a la asignación.	70
Paso 3. Añadir una transformación de analizador a la asignación.	71
Paso 4. Configurar la transformación de analizador.	71
Tarea 3. Ejecutar un perfil en la transformación de analizador.	72
Tarea 4. Ejecutar la asignación.	72
Tarea 5. Ver la salida de la asignación.	72
Resumen del análisis de datos.	73

Capítulo 14: Lección 5. Estandarización de datos 74

Resumen de la estandarización de datos.	74
Tarea 1. Crear un objeto de datos de destino.	75
Paso 1. Crear un objeto de datos All_Customers_Stdz_tgt.	75
Paso 2. Configurar opciones de lectura y escritura.	76
Tarea 2. Crear una asignación para estandarizar datos.	77
Paso 1. Crear una asignación.	77
Paso 2. Añadir objetos de datos a la asignación.	77
Paso 3. Añadir una transformación de estandarizador a la asignación.	78
Paso 4. Configurar la transformación de estandarizador.	78
Tarea 3. Ejecutar la asignación.	79
Tarea 4. Ver la salida de la asignación.	79
Resumen de la estandarización de datos.	80

Capítulo 15: Lección 6. Validación de datos de dirección. 81

Resumen de la validación de datos de direcciones.	81
Tarea 1. Crear un objeto de datos de destino	82
Paso 1. Crear el objeto de datos All_Customers_av_tgt.	83
Paso 2. Configurar opciones de lectura y escritura.	83
Paso 3. Añadir puertos al objeto de datos.	84
Tarea 2. Crear una asignación para validar direcciones.	84
Paso 1. Crear una asignación.	84
Paso 2. Añadir objetos de datos a la asignación.	85
Paso 3. Añadir una transformación de validador de direcciones a la asignación.	85
Tarea 3. Configurar la transformación de validación de dirección.	85
Paso 1. Establecer el país predeterminado para la validación de dirección.	86
Paso 2. Configurar los puertos de entrada de la transformación del validador de direcciones.	86
Paso 3. Configurar los puertos de salida de la transformación del validador de direcciones.	87
Paso 4. Conectar puertos de origen de datos no utilizados con el destino de datos.	88
Tarea 4. Ejecutar la asignación.	89
Tarea 5. Ver la salida de la asignación.	89
Resumen de la validación de datos de direcciones.	92

Apéndice A: Preguntas frecuentes..... 93

Preguntas frecuentes de Informatica Analyst.	93
Preguntas frecuentes de Informatica Developer.	93

Índice..... 95

Prefacio

La *Guía Primeros pasos* sobre Data Quality está escrita para desarrolladores y analistas de calidad de datos. Ofrece guías de autoaprendizaje para ayudar a usuarios novatos a aprender a usar Informatica Developer e Informatica Analyst. Esta guía da por supuesto que usted entiende los conceptos de integración de datos, archivos planos y bases de datos relacionales, así como los motores de base de datos de su entorno.

Documentación de Informatica

Informatica Network

Informatica Network incluye el servicio internacional de atención al cliente de Informatica, la base de conocimiento de Informatica y otros recursos de producto. Para acceder a Informatica Network, visite <https://network.informatica.com>.

Un miembro puede:

- Acceder a todos sus recursos de Informatica en un solo lugar.
- Busque recursos de producto, como documentación, preguntas frecuentes y mejores prácticas en la base de conocimiento.
- Vea la información de disponibilidad del producto.
- Revisar los casos de asistencia.
- Buscar su red de grupos de usuarios de Informatica locales y colaborar con sus iguales.

Base de conocimiento de Informatica

Utilice la base de conocimiento de Informatica para buscar recursos de producto como documentación, artículos de procedimientos, mejores prácticas y PAM en la red de Informatica.

Para acceder a la base de conocimiento, visite <https://kb.informatica.com>. Si tiene preguntas, comentarios o ideas relacionadas con la base de conocimiento de Informatica, póngase en contacto con el equipo de la base de conocimiento de Informatica en KB_Feedback@informatica.com.

Documentación de Informatica

Para obtener la documentación más reciente del producto, consulte la base de conocimiento de Informatica en https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

Si tiene preguntas, comentarios o ideas relacionadas con esta documentación, póngase en contacto con el equipo de documentación de Informatica enviando un correo electrónico a infa_documentation@informatica.com.

Matrices de disponibilidad de productos de Informatica

Las matrices de disponibilidad de producto (PAM, Product Availability Matrixes) indican las versiones de sistemas operativos, bases de datos y otros tipos de orígenes de datos y destinos admitidos por una versión de un producto. Si es miembro de la red de Informatica, puede acceder a las PAM en <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity es un conjunto de sugerencias y mejores prácticas desarrollado por los servicios profesionales de Informatica. Desarrollado a partir de la experiencia real de cientos de proyectos de administración de datos, Informatica Velocity representa el conocimiento conjunto de nuestros asesores, los cuales han trabajado con organizaciones de todo el mundo para planificar, desarrollar, implementar y mantener con éxito soluciones de administración de datos.

Si es miembro de la red de Informatica, puede acceder a los recursos de Informatica Velocity en <http://velocity.informatica.com>.

Si tiene alguna pregunta, comentario o idea acerca de Informatica Velocity, póngase en contacto con los servicios Profesionales de Informatica en ips@informatica.com.

Catálogo de soluciones de Informatica

El Catálogo de soluciones de Informatica es un foro donde puede buscar soluciones que aumenten, amplíen o mejoren sus implementaciones de Informatica. Al aprovechar cualquiera de los cientos de soluciones de los desarrolladores y los socios de Informatica, puede mejorar la productividad y acelerar el tiempo de implementación en los proyectos. Puede acceder al Catálogo de soluciones de Informatica en <https://marketplace.informatica.com>.

Servicio internacional de atención al cliente de Informatica

Puede ponerse en contacto con un centro de atención global por teléfono o a través del soporte en línea en la red de Informatica.

Para encontrar el número de teléfono local del servicio internacional de atención al cliente de Informatica, visite el sitio web de Informatica en el siguiente vínculo:

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

Si es miembro de la red de Informatica, puede utilizar el soporte en línea en <http://network.informatica.com>.

CAPÍTULO 1

Introducción

Este capítulo incluye los siguientes temas:

- [Resumen del dominio de Informatica, 10](#)
- [Presentación de Informatica Analyst, 12](#)
- [Introducción a Informatica Developer, 13](#)
- [Historial de la guía de autoaprendizaje, 15](#)
- [Estructura de la guía de autoaprendizaje, 16](#)

Resumen del dominio de Informatica

Informatica tiene una arquitectura orientada a los servicios que permite adaptar los servicios y compartir recursos en varios equipos. El dominio de Informatica es la unidad principal para el control y la administración de servicios.

Puede iniciar sesión en Informatica Administrator después de instalar Informatica. Debe usar la Herramienta del administrador para administrar el dominio y configurar los servicios de aplicaciones necesarios para poder acceder a los demás clientes de aplicaciones.

El dominio de Informatica contiene los siguientes componentes:

- Clientes de aplicaciones. Un grupo de clientes usado para acceder a la funcionalidad subyacente de Informatica. Los clientes de aplicaciones envían solicitudes al administrador de servicios o a los servicios de aplicaciones.
- Servicios de aplicaciones. Un grupo de servicios que representan la funcionalidad basada en servidor. Un dominio de Informatica puede contener un subconjunto de servicios de aplicaciones. Debe crear y configurar los servicios de aplicación que requieran los clientes de aplicaciones.

Los servicios de aplicaciones incluyen servicios del sistema que pueden tener una sola instancia en el dominio. Al crear el dominio, los servicios del sistema se crean automáticamente. Puede configurar y habilitar un servicio del sistema para usar la funcionalidad que el servicio proporciona.

- Almacén de perfiles. Una base de datos relacional que utiliza el servicio de integración de datos para almacenar resultados de perfil.
- Almacén de datos de referencia. Una base de datos relacional que almacena los valores de datos de referencia para los objetos de la tabla de referencia en el repositorio de modelos.
- Repositorios. Un grupo de bases de datos relacionales que almacenan los metadatos de los objetos y procesos necesarios para procesar las solicitudes de usuario de los clientes de aplicaciones.

- Administrador de servicios. Un servicio integrado en el dominio para administrar todas las operaciones del dominio. El administrador de servicios ejecuta los servicios de aplicaciones y realiza funciones de dominio, como la autenticación, la autorización y el inicio de sesión.
- Base de datos de flujo de trabajo. Una base de datos relacional que almacena metadatos en tiempo de ejecución para los flujos de trabajo.

En la siguiente tabla, se indican los clientes de aplicaciones (no se incluye la Herramienta del administrador) y los servicios de aplicaciones y repositorios necesarios para el cliente:

Cliente de aplicación	Servicios de aplicación	Repositorios
Informatica Analyst	<ul style="list-style-type: none"> - Servicio del analista - Servicio de administración de contenido - Servicio de integración de datos - Servicio de repositorio de modelos - Servicio de búsqueda 	Repositorio de modelos
Informatica Developer	<ul style="list-style-type: none"> - Servicio del analista - Servicio de administración de contenido - Servicio de integración de datos - Servicio de repositorio de modelos 	Repositorio de modelos
Metadata Manager	<ul style="list-style-type: none"> - Servicio de Metadata Manager - Servicio de integración de PowerCenter - Servicio de repositorio de PowerCenter 	<ul style="list-style-type: none"> - Repositorio de Metadata Manager - Repositorio de PowerCenter
Cliente de PowerCenter	<ul style="list-style-type: none"> - Servicio de integración de PowerCenter - Servicio de repositorio de PowerCenter 	Repositorio de PowerCenter
Consola del concentrador de servicios web	<ul style="list-style-type: none"> - Servicio de integración de PowerCenter - Servicio de repositorio de PowerCenter - Concentrador de servicios web 	Repositorio de PowerCenter

No es posible acceder a los siguientes servicios de aplicaciones mediante un cliente de aplicación de Informatica:

- Servicio de escucha de PowerExchange. Gestiona la escucha de PowerExchange para el movimiento de datos en masa y la captura de datos modificados. El servicio de integración de PowerCenter se conecta a la escucha de PowerExchange a través del servicio de escucha.
- Servicio de registrador de PowerExchange. Administra el registrador PowerExchange para Linux, UNIX y Windows para capturar los datos de cambios y escribirlos en los archivos de registro del registrador PowerExchange. Los datos de cambios se pueden generar en los registros de recuperación de DB2, los registros de puesta al día de Oracle, una base de datos de distribución de Microsoft SQL Server o los orígenes de datos de un sistema i5/OS o z/OS.
- Servicio SAP BW. Escucha las solicitudes RFC de SAP BI y las solicitudes con las que el servicio de integración de PowerCenter ejecuta flujos de trabajo de extracción o carga en SAP BI.

Disponibilidad de funciones

Los productos de Informatica utilizan un conjunto común de aplicaciones. Las funciones de producto que puede utilizar dependen de la licencia del producto.

La siguiente tabla describe las opciones de licencia y las funciones de aplicación disponibles con cada opción:

Opción de concesión de licencias	Características de Informatica Developer	Características de Informatica Analyst
Data Quality	<ul style="list-style-type: none">- Crear y ejecutar asignaciones- Crear y ejecutar reglas- Creación de perfiles que incluye el uso del perfil de detección empresarial y la detección de la clave principal, la clave externa y la dependencia funcional- Conservar los resultados de perfiles deducidos- Creación de cuadros de mando- Exportar objetos a PowerCenter	<ul style="list-style-type: none">- Creación de perfiles, incluida la detección empresarial- Utilizar la búsqueda de detección para ver dónde se encuentran los datos y los metadatos en los repositorios de creación de perfiles- Creación de cuadros de mando- Conservar los resultados de perfiles deducidos- Crear y ejecutar reglas de perfil- Administración de tablas de referencia- Crear especificaciones de regla y compilar especificaciones de regla en mapplets- Administración de registros defectuosos y duplicados
Data Services	<ul style="list-style-type: none">- Crear modelos de objeto de datos lógicos- Crear y ejecutar asignaciones con transformaciones de Data Services- Crear servicios de datos SQL- Crear servicios web- Exportar objetos a PowerCenter	<ul style="list-style-type: none">- Administración de tablas de referencia
Data Services y opciones de perfil	<ul style="list-style-type: none">- Crear modelos de objeto de datos lógicos- Crear y ejecutar asignaciones con transformaciones de Data Services- Crear servicios de datos SQL- Crear servicios web- Exportar objetos a PowerCenter- Crear y ejecutar reglas con transformaciones de Data Services- Creación de perfiles	<ul style="list-style-type: none">- Administración de tablas de referencia

Presentación de Informatica Analyst

Informatica Analyst es un cliente de aplicación basado en web que los analistas pueden utilizar para analizar, limpiar, estandarizar, perfilar y puntuar los datos de una organización.

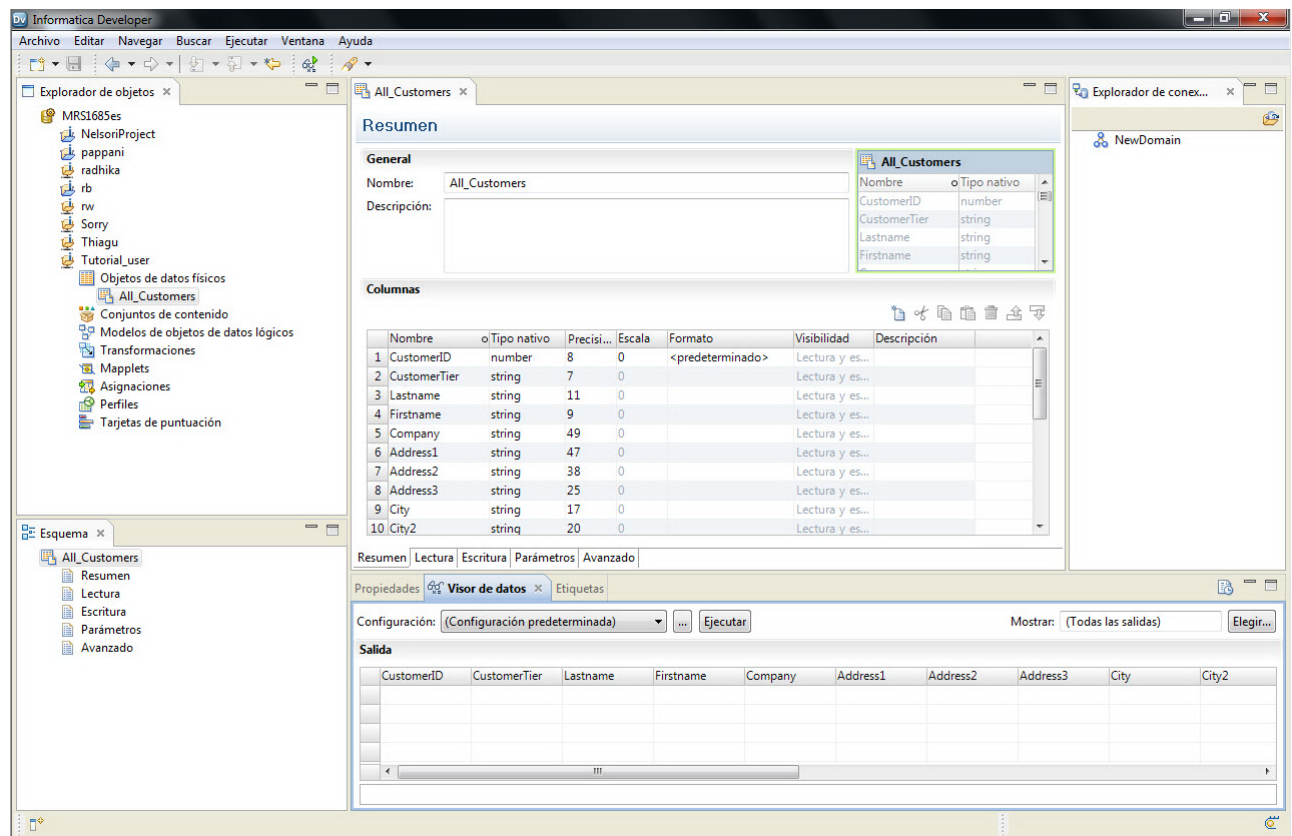
Según su licencia, los desarrolladores y analistas empresariales utilizan Analyst Tool para la colaboración controlada por datos. Puede realizar el perfilado de columnas y reglas, crear cuadros de mando y gestionar

registros incorrectos y duplicados. También puede gestionar datos de referencia y ofrecer los datos a desarrolladores mediante una solución de calidad de datos.

Introducción a Informatica Developer

Informatica Developer es un cliente de aplicación que utilizan los desarrolladores para diseñar e implementar soluciones de datos.

La siguiente figura muestra la herramienta Developer:



La Herramienta del desarrollador puede mostrar las siguientes vistas:

Vista Explorador de objetos

Muestra los proyectos, las carpetas y los objetos que contienen los proyectos y las carpetas.

Vista Explorador de conexiones

Muestra las conexiones con las bases de datos relacionales.

Vista Esquema

Muestra los objetos que dependen de un objeto seleccionado en la vista **Explorador de objetos**.

Vista Referencias rápidas

Muestra la referencia rápida que haya abierto. Para abrir una referencia rápida, haga clic en **Ayuda > Referencias rápidas** y seleccione una referencia rápida.

Vista Visor de datos

Muestra los datos de origen, los resultados de perfil y las vistas previas del resultado de una transformación.

Vista Dependencias de objetos

Muestra las dependencias de objetos al visualizar, modificar o eliminar un objeto.

Vista Alertas

Muestra alertas del estado de conexión.

Vista Registro de validación

Muestra los errores de validación de objetos.

La herramienta Developer puede mostrar también otras vistas. Puede ocultar vistas y mover vistas a otra ubicación en el entorno de trabajo de la Herramienta del desarrollador. Haga clic en **Ventana > Mostrar vista** para seleccionar las vistas que desee mostrar.

Página de bienvenida de Informatica Developer

La primera vez que se abre Developer tool, aparece la página de bienvenida. Utilice la página de bienvenida para conocer mejor Developer tool, configurarla y comenzar a trabajar con ella.

La página de bienvenida muestra las siguientes opciones:

- **Resumen.** Haga clic en el botón Resumen para ver un resumen de las soluciones de calidad de datos y servicios de datos.
- **Primeros pasos.** Haga clic en el botón Primeros pasos para saber más sobre cómo configurar Developer tool y acceder a las lecciones sobre Informatica Data Quality e Informatica Data Services.
- **Guías de autoaprendizaje.** Haga clic en el botón Guías de autoaprendizaje para ver las lecciones de la guía de autoaprendizaje sobre soluciones de calidad de datos y servicios de datos.
- **Recursos web.** Haga clic en el botón Recursos web para obtener un vínculo a mysupport.informatica.com, donde puede acceder a la biblioteca de instrucciones de Informatica. La biblioteca de instrucciones de Informatica contiene artículos sobre Informatica Data Quality, Informatica Data Services y otros productos de Informatica.
- **Novedades** Haga clic en el botón Novedades para ver las últimas características de Developer tool.

Referencias rápidas

La herramienta Developer incluye referencias rápidas como parte de la ayuda en línea. Una referencia rápida es una guía paso a paso que ayuda a completar una o más tareas en la herramienta Developer.

Cuando se sigue una referencia rápida, se realizan las tareas y se ven los resultados. Por ejemplo, después de completar una referencia rápida para importar y previsualizar un objeto de datos relacional, ha importado una tabla de base de datos relacional y previsualizado los datos en la herramienta Developer.

Para acceder a las referencias rápidas, haga clic en **Ayuda > Referencias rápidas**.

Data Quality y creación de perfiles

Utilice las funciones de calidad de datos de la Herramienta del desarrollador para analizar el contenido y la estructura de los datos. Puede mejorar los datos para satisfacer sus necesidades empresariales.

Utilice la Herramienta del desarrollador para diseñar y ejecutar procesos para lograr los siguientes objetivos:

- Crear perfiles de los datos. La creación de perfiles revela el contenido y la estructura de los datos. La creación de perfiles es un paso clave en cualquier proyecto de datos, ya que permite identificar las ventajas y desventajas de los datos y, además, le ayuda a definir el plan de proyecto.
- Crear cuadros de mando para revisar la calidad de los datos. Un cuadro de mando es una representación gráfica de las mediciones de calidad de un perfil.
- Estandarizar valores de datos. Estandarice datos para eliminar errores e inconsistencias detectadas cuando se ejecuta un perfil. Se pueden estandarizar variaciones en la puntuación, formato y ortografía. Por ejemplo, puede asegurarse de que los valores de ciudad, estado y código postal sean uniformes.
- Analizar registros. Analice registros de datos para mejorar la estructura de los registros y obtener información adicional de los datos. Un campo de datos de forma libre se puede dividir en campos que contengan distintos tipos de información. También se puede añadir información a los registros. Por ejemplo, puede marcar registros de clientes como de clientes empresariales o personales.
- Validar direcciones postales. La validación de direcciones evalúa y mejora la exactitud y factibilidad de entrega de los datos de direcciones postales. La validación de direcciones corrige los errores en las direcciones y completa las direcciones parciales comparando registros de direcciones respecto a datos de referencia de agencias postales nacionales. La validación de direcciones también puede añadir información postal para agilizar la entrega del correo y reducir los costes del mismo.
- Buscar registros duplicados. El análisis de registros duplicados compara un conjunto de registros con otro para encontrar valores similares o coincidentes en las columnas de datos seleccionadas. El usuario establece el nivel de similitud que indica una buena coincidencia entre los valores de los campos. También se puede establecer el peso relativo que se le da a cada columna en los cálculos de coincidencia. Por ejemplo, se puede priorizar la información de apellido sobre la de nombre de pila.
- Crear y ejecutar reglas de calidad de datos. Informatica proporciona reglas creadas previamente que se pueden ejecutar o editar para adaptarse a los objetivos de cada proyecto. Se pueden crear reglas en la herramienta Developer.
- Colaborar con usuarios de Informatica. Las reglas y tablas de datos de referencia que se añaden al repositorio de modelos están disponibles para los usuarios de las herramientas Developer y Analyst. Los usuarios pueden colaborar en proyectos; asimismo, varios usuarios pueden ser propietarios de objetos en distintas etapas de un proyecto.
- Exportar asignaciones a PowerCenter. Puede exportar asignaciones a PowerCenter para volver a utilizar los metadatos para la integración de datos físicos o para crear servicios web.

Historial de la guía de autoaprendizaje

HypoStores Corporation es una organización minorista nacional con sede central en Boston y almacenes en varios estados. Integra datos operativos de todos los almacenes del país con los datos guardados en la sede de forma periódica. Recientemente, ha abierto un almacén en Los Ángeles.

La sede incluye un equipo ICC central de administradores, desarrolladores y arquitectos responsables de ofrecer una capa de servicios de datos común para todas las aplicaciones integradas y de inteligencia empresarial. Las aplicaciones de inteligencia empresarial incluyen un sistema CRM que contiene los archivos de datos maestros de clientes utilizados para la facturación y el marketing.

HypoStores Corporation debe realizar las siguientes tareas para integrar los datos de la oficina de Los Ángeles con los de la sede central de Boston:

- Examinar los datos de Boston y Los Ángeles en busca de problemas de calidad de datos.
- Analizar la información de los datos de Los Ángeles.
- Estandarizar la información de direcciones de los datos de Boston y Los Ángeles.
- Validar la exactitud de la información sobre direcciones postales en los datos para fines de CRM.

Estructura de la guía de autoaprendizaje

La *Guía Primeros pasos* contiene guías de autoaprendizaje que incluyen lecciones y tareas.

Lecciones

Cada lección presenta conceptos que le ayudarán a comprender las tareas que deben realizarse en la lección. La lección proporciona los requisitos empresariales a partir del historial general. Los objetivos de la lección destacan las tareas que se realizarán para cumplir los requisitos empresariales. Para cada lección, se indica un tiempo estimado de realización. Después de terminar las tareas de la lección, puede revisar el resumen de la lección.

En el caso de que no esté configurado el entorno de la herramienta, la primera lección de cada guía de autoaprendizaje le ayudará a configurarlo.

Tareas

Las tareas ofrecen instrucciones paso a paso. Lleve a cabo todas las tareas en el orden indicado para realizar la lección.

Guía de autoaprendizaje de Informatica Analyst

En esta guía de autoaprendizaje, un analista inicia una sesión en la herramienta Analyst, crea proyectos y carpetas, crea perfiles y reglas, puntúa datos y crea tablas de referencia.

Las lecciones que puede realizar dependerán de si dispone de los productos Informatica Data Quality o Informatica Data Services.

En la siguiente tabla se describen las lecciones que puede realizar en función del producto que tenga:

Lección	Descripción	Producto
Lección 1. Configuración de Informatica Analyst	Inicio de sesión en la herramienta Analyst y creación de un proyecto y una carpeta para las lecciones de la guía de autoaprendizaje.	Data Quality Data Services
Lección 2. Creación de objetos de datos	Importar un archivo sin formato como objeto de datos y previsualizar los datos.	Data Quality
Lección 3. Creación de perfiles rápidos	Creación de un perfil rápido para hacerse una idea rápida de la calidad de los datos.	Data Quality
Lección 4. Creación de perfiles personalizados	Crear un perfil personalizado para configurar columnas y opciones de muestreo y obtención de detalles.	Data Quality

Lección	Descripción	Producto
Lección 5. Creación de reglas de expresión	Crear reglas de expresión para modificar y perfilar valores de columna.	Data Quality
Lección 6. Creación y ejecución de cuadros de mando	Crear y ejecutar un cuadro de mando para medir el progreso de la calidad de datos a lo largo del tiempo.	Data Quality
Lección 7. Creación de tablas de referencia a partir de resultados de perfil	Crear una tabla de referencia que puede utilizar para estandarizar datos de origen.	Data Quality Data Services
Lección 8. Creación de tablas de referencia	Crear una tabla de referencia para establecer relaciones entre los datos de origen y los valores válidos y estándar.	Data Quality Data Services

Informatica Developer Tool

En esta guía de autoaprendizaje, utilizará Developer tool para llevar a cabo varias operaciones de calidad de datos.

Los usuarios de Informatica Data Quality utilizan Developer tool para diseñar y ejecutar procesos que mejoran la calidad de los datos. Los usuarios de Informatica Data Quality utilizan Developer tool para crear y ejecutar perfiles que analizan el contenido y la estructura de los datos.

Realice las siguientes lecciones en la guía de autoaprendizaje sobre calidad de datos:

Lección 1. Configuración de Informatica Developer

Cree una conexión con un repositorio de modelos gestionado por un servicio de repositorio de modelos en un dominio. Cree un proyecto y una carpeta para guardar el trabajo de las lecciones de la guía de autoaprendizaje. Seleccione un servicio de integración de datos predeterminado.

Lección 2. Importación de objetos de datos físicos

Se definirán procesos de calidad de datos para los archivos de datos de clientes asociados a estos objetos.

Lección 3. Creación de perfiles de datos

La creación de perfiles revela el contenido y la estructura de los datos.

La creación de perfiles incluye el análisis de uniones, que es una forma de análisis que determina si es posible realizar una unión válida entre dos columnas de datos.

Lección 4. Análisis de datos

El análisis realiza los registros de datos y mejora la estructura de los registros. Permite encontrar información útil en los datos y también derivar información nueva a partir de datos actuales.

Lección 5. Estandarización de datos

La estandarización elimina errores de datos y las inconsistencias detectadas durante la creación de perfiles.

Lección 6. Validación de datos de dirección

La validación de direcciones evalúa la precisión y factibilidad de entrega de las direcciones postales y corrige errores y omisiones en estas direcciones.

Requisitos previos de la guía de autoaprendizaje

Antes de poder iniciar las lecciones de la guía de autoaprendizaje, el dominio de Informatica debe ejecutarse al menos con un nodo configurado.

El instalador incluye archivos de la guía de autoaprendizaje que usted utilizará para completar las lecciones. Puede encontrar todos los archivos en las instalaciones del cliente y el servidor:

- Puede encontrar los archivos de la guía de autoaprendizaje en la siguiente ubicación de la ruta de instalación de la herramienta Developer:

`<Directorio de instalación de Informatica>\clients\DeveloperClient\Tutorials`

- Puede encontrar los archivos de la guía de autoaprendizaje en la siguiente ubicación de la ruta de instalación de los servicios:

`<Directorio de instalación de Informatica>\server\Tutorials`

Necesita los siguiente archivos para las lecciones de la guía de autoaprendizaje:

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

Parte I: Primeros pasos con Informatica Analyst

Esta parte incluye los siguientes capítulos:

- [Lección 1. Configuración de la herramienta Analyst de Informatica , 20](#)
- [Lección 2. Cómo crear objetos de datos , 23](#)
- [Lección 3. Creación de perfiles predeterminados, 26](#)
- [Lección 4. Creación de perfiles personalizados, 30](#)
- [Lección 5. Creación de reglas de expresión, 34](#)
- [Lección 6. Creación y ejecución de cuadros de mando, 37](#)
- [Lección 7. Creación de tablas de referencia a partir de columnas de perfil, 43](#)
- [Lección 8. Creación de tablas de referencia, 46](#)

CAPÍTULO 2

Lección 1. Configuración de la herramienta Analyst de Informatica

Este capítulo incluye los siguientes temas:

- [Introducción a la configuración de Informatica Analyst, 20](#)
- [Tarea 1. Inicio de sesión en Informatica Analyst, 21](#)
- [Tarea 2. Creación de un proyecto, 21](#)
- [Tarea 3. Creación de una carpeta, 22](#)
- [Resumen de la configuración de Informatica Analyst, 22](#)

Introducción a la configuración de Informatica Analyst

Antes de empezar con las lecciones de esta guía de autoaprendizaje, debe configurar la Herramienta del analista. Para configurar la Herramienta del analista, inicie sesión en la Herramienta del analista y cree un proyecto y una carpeta donde guardar el trabajo.

El dominio de Informatica es un conjunto de nodos y servicios que definen el entorno de Informatica. Los servicios del dominio incluyen el servicio del analista y el servicio de repositorio de modelos. El servicio del analista ejecuta la Herramienta del analista y el servicio de repositorio de modelos gestiona el repositorio de modelos. Cuando se trabaja con la Herramienta del analista, esta herramienta almacena los activos que se han creado en el repositorio de modelos.

Debe crear un proyecto antes de que pueda crear activos en la Herramienta del analista. Un proyecto contiene activos en la Herramienta del analista. Asimismo, un proyecto puede contener carpetas que almacenen activos relacionados, como objetos de datos que formen parte del mismo requisito de negocio.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

- Inicio de sesión en la Herramienta del analista.
- Creación de un proyecto para almacenar los activos que cree en la Herramienta del analista.
- Creación de una carpeta en el proyecto en la que pueda guardar los activos relacionados.

Requisitos previos

Antes de empezar con esta lección, compruebe los siguientes requisitos previos:

- Un administrador ha configurado un servicio de repositorio de modelos y un servicio del analista en la herramienta Administrator.
- Tiene el nombre de host y el número de puerto de la Herramienta del analista.
- Tiene un nombre de usuario y una contraseña para acceder al servicio del analista. Un administrador puede facilitarle esta información.

Duración

Destine de 5 a 10 minutos para realizar esta lección.

Tarea 1. Inicio de sesión en Informatica Analyst

Inicie sesión en la Herramienta del analista para iniciar la guía de autoaprendizaje.

1. Inicie un navegador Microsoft Internet Explorer o Google Chrome.
2. En el campo Dirección, escriba la URL de Informatica Analyst:
`http[s]://<fully qualified host name>:<port number>/analyst`
3. Si el dominio utiliza la autenticación nativa o de LDAP, introduzca su nombre de usuario y contraseña en la página de inicio de sesión.
4. Seleccione **Nativo** o el nombre de un dominio de seguridad específico.
El campo Dominio de seguridad aparece cuando el dominio de Informatica utiliza la autenticación de LDAP o Kerberos. Si no conoce el dominio de seguridad al que pertenece su cuenta de usuario, póngase en contacto con el administrador de dominios de Informatica.
5. Haga clic en **Iniciar sesión**.
La Herramienta del analista se abre en el espacio de trabajo **Inicio**.

Tarea 2. Creación de un proyecto

En esta tarea, se crea un proyecto que contenga los activos creados en Analyst Tool. Cree un proyecto de la guía de autoaprendizaje que contenga la carpeta del proyecto.

1. En el encabezado **Administrar**, haga clic en **Proyectos**.
Aparece el espacio de trabajo **Proyectos**.
2. En el menú **Acciones**, haga clic en **Nuevo > Proyecto**.
Aparece la ventana **Nuevo proyecto**.
3. Escriba su nombre precedido de "Tutorial_" como nombre del proyecto.
4. Haga clic en **Aceptar**.

Tarea 3. Creación de una carpeta

En esta tarea, tiene que crear una carpeta para almacenar los activos relacionados. Puede crear una carpeta en un proyecto o en otra carpeta. Cree una carpeta con el nombre Clientes para almacenar los activos relacionados en el proyecto de calidad de datos.

1. En el panel **Proyectos**, seleccione el proyecto de la guía de autoaprendizaje.
2. En el menú **Acciones**, haga clic en **Nuevo > Carpeta**.
Se abrirá la ventana **Nueva carpeta**.
3. Escriba Clientes como nombre de carpeta.
4. Haga clic en **Aceptar**.
Aparece la carpeta incluida en el proyecto de esta guía de autoaprendizaje.

Resumen de la configuración de Informatica Analyst

En esta lección ha aprendido que Analyst Tool almacena activos en proyectos y carpetas. Un repositorio de modelos contiene proyectos y carpetas. El servicio Analyst ejecuta la herramienta Analyst. El servicio de repositorio de modelos gestiona el repositorio de modelos. El servicio Analyst y el servicio de repositorio de modelos son servicios de aplicación del dominio Informatica.

Inició sesión en la herramienta Analyst y creó un proyecto y una carpeta.

Ya puede utilizar la herramienta Analyst para llevar a cabo otras lecciones de esta guía de autoaprendizaje.

CAPÍTULO 3

Lección 2. Cómo crear objetos de datos

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de objetos de datos, 23](#)
- [Tarea 1. Crear los objetos de datos de archivo sin formato, 24](#)
- [Tarea 2. Visualización de las propiedades del objeto de datos, 24](#)
- [Resumen de la creación de objetos de datos , 25](#)

Introducción a la creación de objetos de datos

En la Herramienta del analista, un objeto de datos es una representación de datos basados en un archivo plano o tabla de base de datos relacional. Se crea un objeto de archivo sin formato o de tabla y luego se ejecuta un perfil sobre los datos del archivo sin formato o de la base de datos relacional. Cuando se crea un objeto de datos a partir de un archivo sin formato en la Herramienta del analista, se puede cargar el archivo en la memoria caché del archivo sin formato en el equipo en el que se ejecuta la Herramienta del analista o se puede especificar la ubicación de red donde está almacenado el archivo sin formato.

Historial

HypoStores guarda los datos de los clientes de Los Ángeles en archivos sin formato. HypoStores necesita crear un perfil y analizar los datos y realizar tareas de calidad de datos.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

1. Cargar el archivo sin formato en la ubicación de memoria caché del archivo sin formato y crear un objeto de datos.
2. Previsualizar los datos para el objeto de datos de archivo sin formato.

Requisitos previos

Antes de empezar con esta lección, compruebe los siguientes requisitos previos:

- Ha terminado la lección 1 de esta guía de autoaprendizaje.
- Tiene el archivo sin formato LA_Customers.csv. Puede encontrar este archivo en la carpeta <Directorio raíz de la instalación>\<Versión de lanzamiento>\clients\DeveloperClient\Tutorials.

Duración

Destine de 5 a 10 minutos para realizar esta tarea.

Tarea 1. Crear los objetos de datos de archivo sin formato

En esta tarea va a crear un objeto de datos de archivo sin formato a partir del archivo LA_Customers.

1. En la Herramienta del analista, haga clic en **Nuevo > Objeto de datos de archivo sin formato**.
Aparece el asistente **Agregar archivo sin formato**.
2. Seleccione **Explorar y cargar** y haga clic en **Explorar**.
3. Vaya a la ubicación de LA_Customers.csv y haga clic en **Abrir**.
4. Haga clic en **Siguiente**.
En el panel **Elija tipo de importación** se muestran las opciones **Delimitada** y **De anchura fijo**. Seleccione la opción **Delimitado**. La opción predeterminada es **Delimitado**.
5. Haga clic en **Siguiente**.
6. En **Especificar los delimitadores y calificadores de texto utilizados en los datos**, seleccione **Comillas dobles** como calificador de texto.
7. En **Especificar líneas para importar**, seleccione **Importar desde la primera línea** para importar nombres de columnas desde la primera línea que no esté en blanco.
El panel **Vista previa** se actualiza para mostrar los encabezados de columna desde la primera fila.
8. Haga clic en **Siguiente**.
El panel **Atributos de columna** muestra el tipo de datos, la precisión, la escala y el formato de cada columna.
9. Haga clic en **Siguiente**.
El campo **Nombre** muestra LA_Customers.
10. Seleccione el proyecto **Tutorial_** y la carpeta **Customers**.
11. Haga clic en **Finalizar**.
El objeto de datos aparece en el contenido de la carpeta Clientes.

Tarea 2. Visualización de las propiedades del objeto de datos

En esta tarea, puede ver las propiedades del objeto de datos LA_Customers.

1. Haga clic en **Abrir** para abrir el espacio de trabajo **Biblioteca**.
2. En el espacio de trabajo **Biblioteca**, haga clic en **Objetos de datos** en el panel **Activos**.
Se muestra una lista de objetos de datos en el panel **Objetos de datos**.
3. Haga clic en el archivo sin formato **LA_Customers**.

- Se muestra el panel **Vista previa de datos** con los datos obtenidos del objeto de datos LA_Customers.
4. En el panel **Vista previa de datos**, revise la estructura y el contenido del objeto de datos LA_Customers. La Herramienta del analista muestra las primeras 100 filas del objeto de datos de archivo sin formato.
 5. Haga clic en **Propiedades**.
El panel **Propiedades** muestra el nombre, el tipo, la descripción y la ubicación del objeto de datos. También puede ver los nombres y las propiedades de la columna para el objeto de datos.

Resumen de la creación de objetos de datos

En esta lección, ha aprendido que los objetos de datos son representaciones de datos basadas en un origen de archivo sin formato o de base de datos relacional. Ha aprendido que puede crear un objeto de datos de archivo sin formato y previsualizar los datos que contiene.

Ha cargado un archivo sin formato y creado un objeto de datos de un archivo sin formato, previsualizado los datos del objeto de datos y visualizado las propiedades del objeto de datos.

Después de crear un objeto de datos, cree un perfil predeterminado para el objeto de datos en la lección 3 y cree un perfil personalizado para el objeto de datos en la lección 4.

CAPÍTULO 4

Lección 3. Creación de perfiles predeterminados

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de perfiles predeterminados, 26](#)
- [Tarea 1. Crear y ejecutar un perfil predeterminado, 27](#)
- [Tarea 2. Ver los resultados del perfil en una vista de resumen, 28](#)
- [Resumen de la creación de perfiles predeterminados, 28](#)

Introducción a la creación de perfiles predeterminados

Un perfil es el análisis de la calidad de datos basado en el contenido y la estructura de los datos. Un perfil predeterminado es un perfil que se crea con opciones predeterminadas. Utilice un perfil predeterminado para obtener resultados de perfil sin tener que configurar todas las columnas y opciones de un perfil.

Cree y ejecute un perfil predeterminado para analizar la calidad de los datos cuando inicie un proyecto de calidad de datos. Cuando cree un objeto de perfil predeterminado, seleccione el objeto de datos y las columnas del objeto de datos que desee analizar. Un perfil predeterminado omite la columna de perfil y la configuración de opciones. La Herramienta del analista crea los perfiles en el archivo sin formato activo del objeto de datos del archivo sin formato.

Historial

HypoStores desea incorporar a su almacén de datos los datos de la recién adquirida oficina de Los Ángeles. Antes de incorporar los datos al almacén de datos, éste se debe limpiar. Usted es el analista responsable de evaluar la calidad de los datos y de retransmitir la información al desarrollador responsable de la limpieza de los datos. Usted desea ver rápidamente los resultados del perfil y hacerse una idea básica sobre la calidad de los datos.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

1. Crear y ejecutar un perfil predeterminado para el objeto de datos de archivo sin formato LA_Customers.
2. Ver los resultados del perfil.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 2 de esta guía de autoaprendizaje.

Duración

Destine de 5 a 10 minutos para realizar esta lección.

Tarea 1. Crear y ejecutar un perfil predeterminado

En esta tarea, creará un perfil predeterminado para todas las columnas del objeto de datos y utilizará opciones predeterminadas de muestreo y de obtención de detalles.

1. En el espacio de trabajo **Biblioteca**, seleccione un objeto de datos en el panel **Activos**.
2. Haga clic con el botón derecho en el objeto de datos y seleccione **Crear perfil**.
Aparecerá el asistente **Nuevo perfil**.
3. La opción **Origen único** está seleccionada de forma predeterminada. Haga clic en **Siguiente**.
4. En la pantalla **Especificar propiedades generales**, el nombre, la descripción y la ubicación se rellenan de forma predeterminada.
5. Haga clic en **Siguiente**.
6. En la pantalla **Seleccionar origen**, el objeto de datos aparece de forma predeterminada. Puede ver las columnas en el panel **Seleccionar columnas**.
7. Haga clic en **Siguiente**.
8. En la pantalla **Especificar la configuración**, las siguientes opciones están seleccionadas de forma predeterminada:
 - **Ejecutar perfil de columna**
 - **Todas las filas** en el panel **Ejecutar perfil en**
 - **Activo** en el panel **Obtención de detalles**
 - **Excluir de la inferencia los tipos y dominios de datos aprobados en las siguientes ejecuciones del perfil**.
 - Conexión **Nativa** en el panel **Entorno en tiempo de ejecución**.
9. Haga clic en **Siguiente**.
10. En la pantalla **Especificar reglas y filtros**, haga clic en **Guardar y ejecutar** para crear y ejecutar el perfil.
La Herramienta del analista crea y ejecuta el perfil. Los resultados de perfil aparecen en la vista de resumen.

Tarea 2. Ver los resultados del perfil en una vista de resumen

En esta tarea, utilizará la vista de resumen del perfil LA_Customers para obtener un resumen rápido de los resultados del perfil.

1. En el panel **Biblioteca > Activos > Perfiles**, haga clic en el perfil **LA_Customers**.
Los resultados de perfil aparecen en la vista de resumen.
2. En la vista de resumen, haga clic en **Columnas** en el panel **Filtrar por** para ver los resultados del perfil por columnas.
3. Mueva el cursor sobre los gráficos de barras horizontales para ver los valores en porcentajes.
4. En la sección **Porcentaje de valores nulos, únicos y no únicos**, podrá ver los valores nulos, únicos y no únicos en porcentajes para una columna.
5. En la sección **Patrón**, podrá ver varios patrones de la columna como gráficos de barras horizontales. Puede ver los caracteres de los patrones y el número de patrones similares de una columna como un porcentaje si mueve el puntero sobre el gráfico de barras.
6. En la sección **Longitud**, podrá ver la longitud mínima y máxima de los valores de la columna.
7. En la sección **Valor**, podrá ver los valores mínimo y máximo de una columna.
8. En la sección **Tipo de datos**, podrá ver todos los tipos de datos deducidos y todos los tipos de datos documentados de una columna si mueve el cursor sobre los valores.
9. En la sección **Dominio de datos**, podrá ver todos los tipos de dominio deducidos de una columna si mueve el cursor sobre los valores.
10. Para ver los datos atípicos, haga clic en **Acciones > Detectar valor atípico** para detectar valores atípicos en los resultados de perfil.
11. Haga clic en los filtros **Valor atípico del patrón** o **Valor atípico de frecuencia del valor** para ver los valores atípicos en los resultados del perfil.
12. Haga clic en un nombre de columna para ver los resultados de perfil de dicha columna en la vista detallada.

Resumen de la creación de perfiles predeterminados

En esta lección, ha aprendido que un perfil predeterminado muestra resultados de un perfil sin tener que configurar todas las opciones de muestreo de columnas y filas para un perfil. Ha aprendido que puede crear y ejecutar un perfil predeterminado para analizar la calidad de los datos cuando inicie un proyecto de calidad de datos. También ha aprendido que la Herramienta del analista crea perfiles en el archivo sin formato activo del objeto de datos de archivo sin formato.

Ha creado un perfil predeterminado y ha analizado los resultados del perfil. Ha obtenido más información sobre las columnas del perfil, incluidos los valores nulos y los tipos de dato. También ha utilizado los valores y patrones de la columna para identificar problemas de calidad de datos.

Después de analizar los resultados de un perfil rápido, puede realizar las siguientes tareas:

- Crear un perfil personalizado para excluir columnas del perfil e incluir solamente las columnas en las que esté interesado.
- Crear una regla de expresión para crear columnas virtuales y crear un perfil para las mismas.

- Crear una tabla de referencias para incluir los valores de una columna.

CAPÍTULO 5

Lección 4. Creación de perfiles personalizados

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de perfiles personalizados, 30](#)
- [Tarea 1. Crear un perfil personalizado, 31](#)
- [Tarea 2. Ejecutar el perfil, 32](#)
- [Tarea 3. Obtener detalles de los resultados del perfil, 32](#)
- [Resumen de la creación de perfiles personalizados, 33](#)

Introducción a la creación de perfiles personalizados

Un perfil es el análisis de la calidad de datos basado en el contenido y la estructura de los datos. Un perfil personalizado es un perfil que crea cuando desea configurar las columnas, las opciones de muestreo y de obtención de detalles para una creación de perfiles más rápida. Configure las opciones de muestreo para seleccionar las filas de muestra en el origen. Configure las opciones de obtención de detalles para obtener detalles de las filas de datos en los datos de origen o provisionales. Puede optar por ejecutar el perfil en un entorno nativo o de Hive.

Un perfil se crea y ejecuta para analizar la calidad de los datos después de iniciar un proyecto de calidad de datos. Al crear un objeto de perfil, puede comenzar seleccionando el objeto de datos y las columnas del objeto de datos en los que desea ejecutar un perfil.

Historial

HypoStores necesita incorporar a su almacén de datos los datos de la recién adquirida oficina de Los Ángeles. HypoStores desea acceder a la calidad de los datos de niveles de clientes del archivo de datos de clientes de Los Ángeles. Usted es el analista responsable de evaluar la calidad de los datos y de transmitir la información al desarrollador responsable de la limpieza de los datos.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

1. Crear un perfil personalizado para el objeto de datos de archivo sin formato y excluir las columnas con valores nulos.
2. Ejecutar el perfil para analizar el contenido y estructura de la columna CustomerTier.
3. Obtener detalles de las filas para los resultados del perfil.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado las Lecciones 1, 2 y 3 de esta guía de autoaprendizaje.

Duración

Destine de 5 a 10 minutos para realizar esta lección.

Tarea 1. Crear un perfil personalizado

En esta tarea, creará un perfil personalizado. Para crear un perfil personalizado, seleccione el objeto de datos y las columnas en los que desee ejecutar un perfil. También configura las opciones de muestreo y de obtención de detalles.

1. Haga clic en **Nuevo > Perfil**.
Aparecerá el asistente **Nuevo perfil**.
2. La opción **Origen único** está seleccionada de forma predeterminada. Haga clic en **Siguiente**.
3. En la pantalla **Especificar propiedades generales**, defina las siguientes opciones:
 - En el campo Nombre, introduzca **Profile_LA_Customers**.
 - En el campo Ubicación, seleccione la carpeta **Customers**.
4. Haga clic en **Siguiente**.
5. En la pantalla **Seleccionar origen**, haga clic en **Seleccionar**.
Se muestra el cuadro de diálogo **Elegir objeto de datos**.
6. En el cuadro de diálogo **Elegir objeto de datos**, seleccione **LA_Customers**. Haga clic en **Aceptar**.
7. En la pantalla **Seleccionar origen**, desactive las columnas Address2, Address3 y City2.
8. Haga clic en **Siguiente**.
9. En la pantalla **Especificar la configuración**, defina las opciones siguientes:
 - **Ejecutar perfil de columna**.
 - Seleccione la opción **Muestra aleatoria** en el panel **Ejecutar perfil en**.
 - Seleccione la opción **Excluir de la inferencia los tipos y dominios de datos aprobados en las siguientes ejecuciones del perfil**.
 - Seleccione la opción **Provisional** en el panel **Obtención de detalles**.
10. Haga clic en **Siguiente**.
11. En la pantalla **Especificar reglas y filtros**, haga clic en **Guardar y finalizar** para crear y ejecutar el perfil.
La Herramienta del analista crea el perfil y lo muestra en el espacio de trabajo **Detección**. Tiene que ejecutar el perfil para ver los resultados.

Tarea 2. Ejecutar el perfil

En esta tarea, se ejecuta un perfil para crear perfiles en el objeto de datos y mostrar los resultados de los perfiles. La Herramienta del analista crea los perfiles en el archivo sin formato temporal para el objeto de datos del archivo sin formato.

1. Asegúrese de que está en el Espacio de trabajo **Detección**.
Puede ver el perfil **Profile_LA_Customers_Custom** en el espacio de trabajo.
2. Haga clic en **Profile_LA_Customers_Custom** en el espacio de trabajo.
3. Aparece la pantalla del perfil, donde puede optar por modificar el perfil o ejecutarlo. Haga clic en **Ejecutar**.
4. Los resultados del perfil aparecen en la vista de resumen.

Tarea 3. Obtener detalles de los resultados del perfil

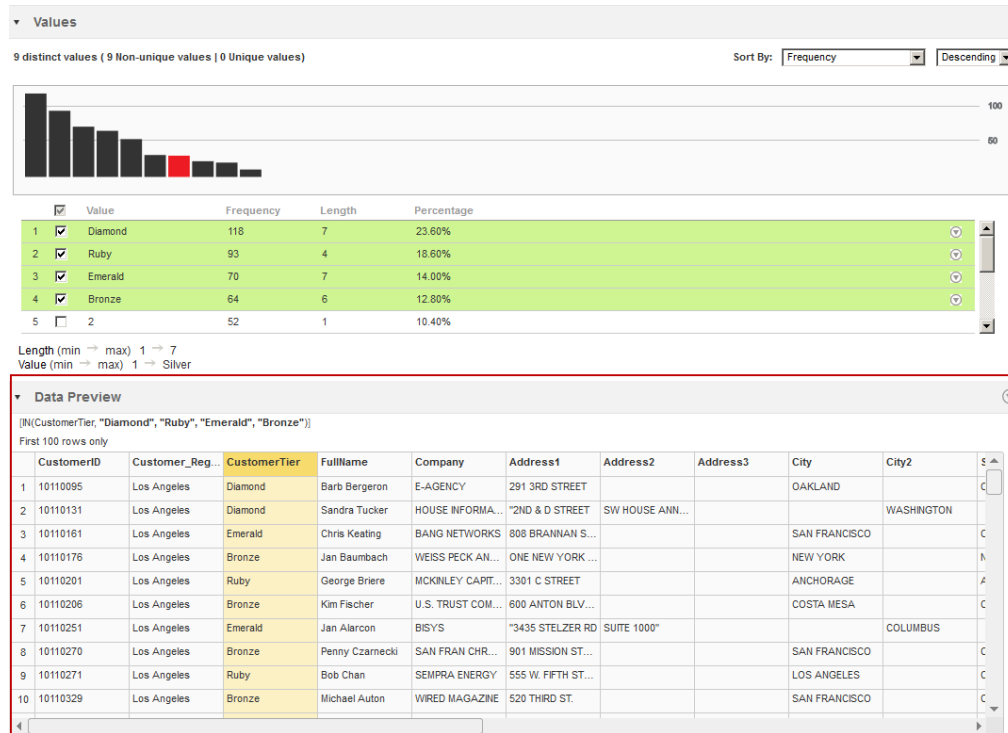
En esta tarea, se obtienen detalles de los valores de la columna CustomerTier para ver las filas de origen en el objeto de datos para el perfil.

1. Compruebe que está en la vista de resumen de los resultados del perfil **Profile_LA_Customers**.
2. Haga clic en la columna **CustomerTier**.
Los resultados de perfil de la columna se mostrarán en la vista detallada.

3. En la vista detallada, seleccione los valores Diamante, Rubí, Esmeralda y Bronce. Haga clic con el botón derecho en los valores del panel **Valores** y seleccione **Obtener detalles**.

Las filas de la columna que tienen el valor Diamante, Rubí, Esmeralda o Bronce aparecen en el panel **Vista previa de datos**.

La siguiente imagen muestra los resultados de obtención de detalles en el panel **Vista previa de datos** de los valores Diamante, Rubí, Esmeralda o Bronce:



El panel **Vista previa de datos** muestra las 100 primeras filas de la columna seleccionada. El título del panel **Vista previa de datos** muestra la lógica utilizada para la columna de origen.

Resumen de la creación de perfiles personalizados

En esta lección, ha aprendido que puede configurar las columnas para las que se crea un perfil y que se pueden configurar las opciones de muestreo y de obtención de detalles. Ha aprendido que puede obtener detalles para ver las filas subyacentes de valores de columna y que puede configurar las columnas que se incluyen en el momento de visualizar los valores de columna.

Ha creado un perfil personalizado que incluía la columna CustomerTier columna, ha ejecutado el perfil y obtenido detalles de las filas subyacentes para la columna CustomerTier en los resultados.

Utilice el objeto de perfil personalizado para crear una regla de expresión en la lección 5.

CAPÍTULO 6

Lección 5. Creación de reglas de expresión

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de reglas de expresión, 34](#)
- [Tarea 1. Crear reglas de expresión y ejecutar el perfil, 35](#)
- [Tarea 2. Ver la salida de la regla de expresión, 35](#)
- [Tarea 3. Editar las reglas de expresión, 36](#)
- [Resumen de la creación de reglas de expresión, 36](#)

Introducción a la creación de reglas de expresión

Las reglas de expresión emplean funciones de expresión y columnas de origen para definir la lógica de las reglas. Puede crear reglas de expresión y añadirlas a un perfil en la Herramienta del analista. Puede asociar una regla de expresión a uno o más perfiles.

La salida de una regla de expresión es una columna virtual del perfil. La Herramienta del analista perfila la columna virtual cuando se ejecuta el perfil.

Puede usar reglas de expresión para validar columnas origen o crear columnas origen adicionales basadas en el valor de las columnas origen.

Ejemplo

HypoStores desea incorporar a su almacén de datos los datos de la recién adquirida oficina de Los Ángeles. HypoStores desea analizar los nombres de los clientes y dividirlos en nombre y apellido. HypoStores desea utilizar reglas de expresión para analizar una columna que contenga nombres y apellidos en columnas virtuales diferenciadas y luego perfilar las columnas. HypoStores también quiere poner las reglas a disposición de otros analistas que necesiten analizar la salida de estas reglas.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

1. Crear reglas de expresión para separar la columna FullName en una columna para el nombre y otra para el apellido. Crear una regla para separar el nombre del nombre completo. Crear otra regla para separar el apellido del nombre completo. Estas reglas se crean para el perfil Profile_LA_Customers.
2. Ejecutar el perfil y visualizar la salida de las reglas en el perfil.

3. Editar las reglas para que las puedan usar otros usuarios de la Herramienta del analista.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha finalizado las lecciones 1 a 4.

Duración

Destine de 10 a 15 minutos para realizar esta lección.

Tarea 1. Crear reglas de expresión y ejecutar el perfil

En esta tarea, se crean dos reglas de expresión para analizar la columna FullName en dos columnas virtuales llamadas FirstName y LastName. Los nombres de la regla son FirstName y LastName.

1. En el espacio de trabajo **Biblioteca**, haga clic en el perfil **Profile_LA_Customers**.
Los resultados de perfil aparecen en la vista de resumen.
2. Haga clic en **Editar**.
Aparecerá el asistente **Perfil**.
3. Haga clic en la pantalla **Especificar reglas y filtros**.
4. En el panel **Reglas**, haga clic en **Acciones > Crear regla**.
5. En el campo **Nombre**, introduzca **FirstName**.
6. En la sección **Expresión**, introduzca la siguiente expresión para separar el nombre de la columna **Nombre**:

```
SUBSTR(FullName,1,INSTR(FullName,' ',-1,1) - 1)
```
7. Haga clic en **Validar**.
8. Haga clic en **Aceptar**.
9. Repita los pasos 4 a 8 para crear una regla denominada **LastName**. Introduzca la siguiente expresión para separar el apellido de la columna **Nombre**:

```
SUBSTR(FullName,INSTR(FullName,' ',-1,1),LENGTH(FullName))
```
10. Haga clic en **Guardar y ejecutar** para guardar y ejecutar el perfil.

Tarea 2. Ver la salida de la regla de expresión

En esta tarea, se visualiza la salida de reglas de expresión que separan nombres y apellidos después de ejecutar un perfil.

1. En la vista de resumen, haga clic en **Editar**.
Aparecerá el asistente Perfil.
2. En el asistente Perfil, haga clic en **Seleccionar origen**.
3. En la pantalla **Seleccionar origen**, marque la casilla junto a **Nombre** en la barra de herramientas para borrar todas las columnas.

Una de las columnas está seleccionada de forma predeterminada porque es necesario seleccionar al menos una columna en la sección **Columnas**.

4. Seleccione la columna **FullName**, y las reglas **FirstName** y **LastName**.

5. Borre cualquier otra columna que se haya seleccionado.

6. Haga clic en **Guardar y ejecutar**.

Los resultados de perfil aparecen en la vista de resumen.

7. Haga clic en la regla **FirstName**. Los resultados de perfil de la regla aparecerán en la vista detallada.

8. Seleccione un valor en el panel **Valores**. Haga clic con el botón derecho en el valor y seleccione **Obtención de detalles**.

Los valores para la columna **FullName**, así como las reglas **FirstName** y **LastName** aparecerán en el panel **Vista previa de datos** junto con otros valores de columna. Tenga en cuenta que la Herramienta del analista divide la columna **FullName** en nombre y apellido.

Tarea 3. Editar las reglas de expresión

En esta tarea, define que las reglas de expresión se puedan reutilizar y que estén disponibles para todos los usuarios de la Herramienta del analista.

1. En la vista de resumen del perfil **Profile_LA_Customers**, haga clic en **Editar**.

Aparecerá el asistente Perfil.

2. Haga clic en la pantalla **Especificar reglas y filtros**.

3. En la pantalla **Especificar reglas y filtros**, seleccione la regla **FirstName** y haga clic en **Acciones > Editar regla**.

A continuación, aparecerá **Editar regla**.

4. Seleccione la opción **¿Desea guardar esta regla como una regla reutilizable?** y elija una ubicación para guardar la regla.

5. Haga clic en **Aceptar**.

6. Seleccione la regla **LastName** y repita los pasos 3 a 5.

7. Haga clic **Guardar y finalizar** para guardar el perfil.

Cualquier usuario de la Herramienta del analista puede utilizar las reglas **FirstName** y **LastName** para dividir una columna con nombres y apellidos en columnas diferenciadas.

Resumen de la creación de reglas de expresión

En esta lección, ha aprendido que las reglas de expresión utilizan funciones de expresión y columnas de origen para definir una lógica de regla. Ha aprendido que la salida de una regla de expresión es una columna virtual del perfil. La herramienta Analyst incluye la columna virtual cuando se ejecuta el perfil.

Ha creado dos reglas de expresión, las ha añadido a un perfil y ha ejecutado el perfil. Ha visualizado la salida de las reglas y las ha puesto a disposición de todos los usuarios de la herramienta Analyst.

CAPÍTULO 7

Lección 6. Creación y ejecución de cuadros de mando

Este capítulo incluye los siguientes temas:

- [Introducción a la creación y ejecución de cuadros de mando, 37](#)
- [Tarea 1. Crear un cuadro de mando a partir de los resultados de un perfil, 38](#)
- [Tarea 2. Ejecutar el cuadro de mandos, 39](#)
- [Tarea 3. Ver el cuadro de mandos, 40](#)
- [Tarea 4. Editar el cuadro de mandos, 40](#)
- [Tarea 5. Configurar umbrales, 41](#)
- [Tarea 6. Ver gráficos de tendencias de puntuación, 41](#)
- [Creación y ejecución de cuadros de mando, 42](#)

Introducción a la creación y ejecución de cuadros de mando

Un cuadro de mandos es una representación gráfica de valores válidos para una columna o la salida de una regla en los resultados del perfil. Utilice cuadros de mandos para medir y controlar el progreso de la calidad de los datos a lo largo del tiempo.

Para crear un cuadro de mandos, añada columnas desde el perfil a un cuadro de mandos como mediciones, asigne ponderaciones a las mediciones y configure los umbrales de puntuación. Puede agregar filtros a los cuadros de mandos que se basen en los datos de origen. Para ejecutar un cuadro de mandos, seleccione los valores válidos de la medición y ejecute el cuadro de mandos para ver las puntuaciones de las mediciones.

Los cuadros de mandos muestran la frecuencia de valores de las columnas de un perfil en forma de puntuaciones. Las puntuaciones reflejan el porcentaje de valores válidos de una medición.

Historial

HypoStores desea incorporar a su almacén de datos los datos de la recién adquirida oficina de Los Ángeles. Antes de que la organización fusione los datos, desea comprobar que se analiza la calidad de los datos en los distintos niveles de clientes y estados. Usted es el analista responsable de supervisar el progreso de los análisis de calidad de los datos. Desea crear un cuadro de mandos a partir de las columnas de perfil de nivel de cliente y estado y visualizar gráficos de tendencia de puntuación para determinar hasta qué punto mejoran las puntuaciones a lo largo del tiempo.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

1. Crear un cuadro de mandos a partir de los resultados del perfil Profile_LA_Customers_Custom para visualizar las puntuaciones correspondientes a las columnas CustomerTier y State.
2. Ejecutar el cuadro de mandos para generar las puntuaciones de las columnas CustomerTier y State.
3. Visualizar el cuadro de mandos para ver las puntuaciones de cada columna.
4. Editar el cuadro de mandos para especificar distintos valores válidos para las puntuaciones.
5. Configurar umbrales de puntuación y ejecutar el cuadro de mandos.
6. Visualizar gráficos de tendencia de puntuación para determinar la mejora de las puntuaciones en el transcurso del tiempo.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 5 de esta guía de autoaprendizaje.

Duración

Dedique 15 minutos a completar las tareas de esta lección.

Tarea 1. Crear un cuadro de mando a partir de los resultados de un perfil

En esta tarea, creará un cuadro de mando a partir del perfil Profile_LA_Customers_Custom para puntuar los valores de las columnas CustomerTier y State.

1. En el espacio de trabajo **Biblioteca**, haga clic en el perfil **Profile_LA_Customers**.
Aparecerá la vista de resumen de los resultados de perfil.
2. En la vista de resumen, seleccione la columna **CustomerTier** y haga clic con el botón derecho en la columna; a continuación, seleccione **Añadir a > Cuadro de mando**.
Aparece el asistente **Añadir a cuadro de mando**.
3. En el asistente **Añadir a cuadro de mando**, la opción **Nuevo cuadro de mando** está seleccionada de forma predeterminada. Haga clic en **Siguiente**.
4. En la pantalla **Paso 2 de 8**, introduzca **sc_LA_Customer** como nombre del cuadro de mando y navegue hasta la carpeta **Clientes** para definirla como ubicación del cuadro de mando.
5. Haga clic en **Siguiente**.
6. En la pantalla **Paso 3 de 8**, seleccione las columnas **CustomerTier** y **Estado** para añadirlas al cuadro de mando.
7. Haga clic en **Siguiente**.
8. En la pantalla **Paso 4 de 8** puede crear, editar o eliminar filtros para las mediciones. En esta guía de autoaprendizaje, no vamos a crear un filtro de cuadros de mando. Haga clic en **Siguiente**.
9. En la pantalla **Paso 5 de 8**, seleccione la medición **CustomerTier** en el panel **Mediciones**.
10. En el panel **Puntuación con: valores**, seleccione todos los valores y haga clic en el botón **Añadir todo** para mover los valores a la sección **Valores válidos**.

Utilice la tecla **Mayús** para seleccionar varios valores.

11. En el panel **Mediciones**, seleccione la medición **Estado** y aquellos valores que tengan un código de estado de dos letras en la sección **Puntuación con: valores**.
12. Haga clic en el botón **Añadir** para mover los valores a la sección **Valores válidos**.
Puede ver el número total de los valores válidos y el porcentaje de valores válidos en la parte superior de la sección.
13. En la sección **Mediciones** de cada medición, acepte la configuración predeterminada de los umbrales de puntuación en la sección **Umbrales de medición**.
14. Haga clic en **Siguiente**.
15. En la pantalla **Paso 6 de 8**, puede seleccionar un grupo de mediciones para añadirlo a las mediciones de forma opcional. De forma predeterminada, la Herramienta del analista añade las mediciones al grupo de mediciones **Predeterminado**.
16. Haga clic en **Siguiente**.
17. En la pantalla **Paso 7 de 8**, haga doble clic en la columna **Ponderación** de la medición **CustomerTier** del panel **Predeterminado - Mediciones**.
Al ejecutar un cuadro de mando, la Herramienta del analista calcula la media ponderada de cada grupo de mediciones según la puntuación y ponderación asignadas a cada medición.
18. Especifique la ponderación de las mediciones **CustomerTier** y **State**.
19. Haga clic en **Siguiente**.
20. En la pantalla **Paso 8 de 8**, la opción **Nativo** está seleccionada de forma predeterminada. Haga clic en **Guardar** para crear el cuadro de mando.
El cuadro de mando aparece en el espacio de trabajo **Cuadros de mando**.

Tarea 2. Ejecutar el cuadro de mandos

En esta tarea, ejecutará el cuadro de mandos `sc_LA_Customer` para generar las puntuaciones de las columnas **CustomerTier** y **State**.

1. En el espacio de trabajo **Biblioteca**, haga clic en **Activos > Cuadros de mandos**.
2. Seleccione un cuadro de mandos del panel **Cuadro de mandos**.
3. Haga clic en **Acciones > Abrir**.
Aparecerá el cuadro de mandos.
4. Haga clic en **Acciones > Ejecutar cuadro de mandos**.
Aparece el cuadro de diálogo **Ejecutar cuadro de mandos**.
5. Compruebe los valores del cuadro de diálogo y haga clic en **Ejecutar**.
El espacio de trabajo **Cuadros de mando** muestra las puntuaciones de las columnas **CustomerTier** y **State**.

Tarea 3. Ver el cuadro de mandos

En esta tarea, visualizará el cuadro de mandos `sc_LA_Customer` para ver las puntuaciones de las columnas `CustomerTier` y `State`.

1. Seleccione la fila **Estado** que contenga la puntuación Estado que desee visualizar.

En la sección **sc_LA_Customer - mediciones**, podrá ver las siguientes propiedades del cuadro de mandos:

- Nombre del cuadro de mandos.
- Número total de filas en el cuadro de mandos.
- Número de filas no válidas.
- La puntuación junto con un gráfico de barras horizontales.
- Tendencia de la puntuación. Puede hacer clic en la tendencia de la puntuación para ver una representación gráfica en la pantalla **Detalle del gráfico de tendencias**.
- Ponderación de la medición.
- Coste de los datos no válidos.
- Tendencia de coste.
- Objeto de datos. Haga clic en el objeto de datos para acceder a la vista previa de datos del objeto de datos en el espacio de trabajo **Detección**.
- Nombre de la regla o la columna.
- Tipo de origen.
- Icono de obtención de detalles.

2. Haga clic en el icono Obtención de detalles de la fila `State`.

Las puntuaciones no válidas de la columna `State` aparecerán en la sección **Filas no válidas** del panel **Obtención de detalles**.

3. Seleccione **Filas válidas** para ver las puntuaciones que sean válidas para la columna Estado.

4. Haga clic en el icono Obtención de detalles de la fila **CustomerTier**.

Todas las puntuaciones de la columna **CustomerTier** son válidas.

Tarea 4. Editar el cuadro de mandos

En esta tarea, editará el cuadro de mandos `sc_LA_Customer` para especificar el valor `Rubí` como no válido para la puntuación `CustomerTier`.

1. Compruebe que se encuentra en el espacio de trabajo **Cuadro de mandos** y que está abierto el cuadro de mandos `sc_LA_Customer`.
2. Seleccione **Acciones > Editar > Mediciones**.
Aparece el cuadro de diálogo **Editar cuadro de mandos**.
3. En la sección **Mediciones**, seleccione **CustomerTier**.
4. En la sección **Puntuación con: valores**, mueva **Rubí** desde la sección **Valores válidos** hasta la sección **Valores disponibles**.

Acepte la configuración predeterminada que aparece en la sección **Umbrales de medición**.

5. Haga clic en **Guardar y ejecutar** para guardar los cambios realizados en el cuadro de mandos y ejecutarlo.
6. Vuelva a visualizar la puntuación CustomerTier.
La puntuación CustomerTier cambia al porcentaje 81,4.

Tarea 5. Configurar umbrales

En esta tarea, configurará los umbrales de la puntuación State del cuadro de mandos sc_LA_Customer para determinar los rangos válidos de los datos de la columna State. Los valores con un código de dos letras, como CA, son aceptables, y los códigos con más de dos letras, como Calif, no son aceptables.

1. Compruebe que se encuentra en el espacio de trabajo **Cuadro de mandos** y que está abierto el cuadro de mandos sc_LA_Customer.
2. Seleccione **Acciones > Editar > Mediciones**.
Aparece el cuadro de diálogo **Editar cuadro de mandos**.
3. En la sección **Mediciones**, seleccione **Estado**.
4. En la sección **Umbrales de medición**, introduzca los siguientes rangos para las puntuaciones Buena e Inaceptable: del 90 al 100%, buena; del 0 al 50%, inaceptable y entre el 51 y el 89%, aceptable.
Los umbrales representan los límites inferiores de los rangos aceptable y bueno.
5. Haga clic en **Guardar y ejecutar** para guardar los cambios realizados en el cuadro de mandos y ejecutarlo.
En el panel **Cuadro de mandos**, visualice los cambios en el porcentaje de puntuación y la puntuación representada en forma de barra para la puntuación State.

Tarea 6. Ver gráficos de tendencias de puntuación

En esta tarea, se visualiza el gráfico de tendencias para la puntuación Estado. Puede visualizar gráficos de tendencias para evaluar puntuaciones en el transcurso del tiempo.

1. Compruebe que se encuentra en el espacio de trabajo **Cuadro de mandos** y que está abierto el cuadro de mandos sc_LA_Customer.
2. Seleccione la fila **State**.
3. Haga clic en **Acciones > Mostrar gráfico de tendencias** o haga clic en la flecha situada bajo la columna **Tendencia de puntuación**.
Aparecerá el cuadro de diálogo **Detalle del gráfico de tendencias**. Puede visualizar los umbrales **Bueno**, **Aceptable** e **Inaceptable** para la puntuación. Los umbrales cambian cada vez que se ejecuta el cuadro de mandos después de editar los valores de las puntuaciones en el cuadro de mandos.
4. Señale a cualquier círculo del gráfico para ver los valores válidos en la sección **Valores válidos** en la parte inferior del gráfico.
5. Haga clic en **Cerrar** para volver al cuadro de mandos.

Creación y ejecución de cuadros de mando

En esta lección, ha aprendido que puede crear un cuadro de mando a partir de los resultados de un perfil. Un cuadro de mando contiene las columnas obtenidas a partir de un perfil. Ha aprendido que puede ejecutar un cuadro de mando para generar puntuaciones para columnas. Ha editado un cuadro de mando para configurar valores válidos y establecer umbrales para puntuaciones. También ha aprendido a visualizar el gráfico de tendencia de puntuación.

Ha creado un cuadro de mando a partir de las columnas CustomerTier y State de un perfil para analizar la calidad de los datos de estas columnas. Ha ejecutado el cuadro de mando para generar puntuaciones para cada columna. Ha editado el cuadro de mando para especificar distintos valores válidos para las puntuaciones. Ha configurado umbrales para una puntuación y ha visualizado el gráfico de tendencia de puntuación.

CAPÍTULO 8

Lección 7. Creación de tablas de referencia a partir de columnas de perfil

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de tablas de referencia a partir de columnas de perfil, 43](#)
- [Tarea 1. Crear una tabla de referencia a partir de columnas de perfil, 44](#)
- [Tarea 2. Editar la tabla de referencia, 45](#)
- [Resumen de la creación de tablas de referencia a partir de columnas de perfil, 45](#)

Introducción a la creación de tablas de referencia a partir de columnas de perfil

Una tabla de referencia contiene datos de referencia que puede utilizar para estandarizar datos de origen. Los datos de referencia pueden incluir valores válidos y estándar. Cree tablas de referencia para establecer relaciones entre los valores de datos de origen y los valores válidos y estándar.

Puede crear una tabla de referencia a partir de los resultados de un perfil. Tras crear una tabla de referencia, puede editar la tabla de referencia para añadir columnas o filas y para añadir o editar valores estándar o válidos. Puede ver los cambios realizados en una tabla de referencia en una traza de auditoría.

Historial

HypoStores desea crear un perfil de los datos para detectar anomalías y estandarizarlos con valores válidos. Usted es el analista responsable de estandarizar los valores válidos en los datos. Desea crear una tabla de referencia basada en valores válidos a partir de las columnas de un perfil.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

1. Crear una tabla de referencias a partir de la columna CustomerTier del perfil Profile_LA_Customers_Custom seleccionando valores válidos para las columnas.
2. Editar la tabla de referencia para configurar distintos valores válidos para las columnas.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 6 de esta guía de autoaprendizaje.

Duración

Dedique 15 minutos a completar las tareas de esta lección.

Tarea 1. Crear una tabla de referencia a partir de columnas de perfil

En esta tarea, se crea una tabla de referencia y se añade la columna CustomerTier desde el perfil Profile_LA_Customers a la tabla de referencia.

1. En el espacio de trabajo **Biblioteca**, haga clic en **Activos > Perfiles**.
2. Haga clic en el perfil **Profile_LA_Customers** para abrir los resultados de perfil en la vista de resumen.
3. En la vista de resumen, seleccione la columna CustomerTier que desee añadir a la tabla de referencia. Haga clic con el botón derecho y seleccione **Añadir a tabla de referencia**.

Se muestra el cuadro de diálogo **Agregar a la tabla de referencia**.

4. Seleccione **Crear una tabla de referencia**.
5. Haga clic en **Siguiente**.
6. En el campo **Nombre**, introduzca **Reftab_CustTier_HypoStores**.
7. Introduzca una descripción y establezca 0 como valor predeterminado.

La Herramienta del analista utiliza el valor predeterminado para cualquier registro de tabla que no contenga ningún valor.

8. Haga clic en **Siguiente**.
9. En la sección **Atributos de columna**, configure las siguientes propiedades de la columna CustomerTier:

Propiedad	Descripción
Nombre	CustomerTier
Tipo de datos	Cadena
Precisión	10
Descripción	Valores de niveles de cliente de referencia

10. Si lo desea, puede crear una columna de descripción para las filas de la tabla de referencia. Introduzca el nombre y la precisión para la columna.
11. Compruebe los valores de la columna CustomerTier de la sección **Vista previa**.
12. Haga clic en **Siguiente**.

Aparece el nombre de tabla de referencia Reftab_CustomerTier_HypoStores. Puede introducir una descripción opcional.

13. En la sección **Guardar en**, seleccione el proyecto de guía de autoaprendizaje donde desee crear la tabla de referencia.
En el panel **Tablas de referencia**: aparece una lista de las tablas de referencia existentes en la ubicación seleccionada.
14. Introduzca una nota de auditoría opcional.
15. Haga clic en **Finalizar**.

Tarea 2. Editar la tabla de referencia

En esta tarea, se edita la tabla Reftab_CustomerTier_HypoStores para añadir valores alternativos para los niveles de cliente.

1. En el espacio de trabajo **Biblioteca**, haga clic en **Activos > Tablas de referencia**.
2. Haga clic en la tabla de referencia Reftab_CustomerTier_HypoStores.
La tabla de referencia se abre en el espacio de trabajo **Diseño**.
3. Para editar una fila, selecciónela y haga clic en **Acciones > Editar** o en el icono **Editar**.
A continuación, aparecerá **Editar fila**. Opcionalmente, seleccione múltiples filas para agregar el mismo valor alternativo a cada fila.
4. Introduzca los siguientes valores alternativos para las filas Diamante, Esmeralda, Oro, Plata y Bronce: 1, 2, 3, 4, 5.
Introduzca una nota de auditoría opcional.
5. Haga clic en **Aplicar** para aplicar los cambios.
6. Haga clic en **Cerrar**
Los valores modificados de la tabla de referencia aparecen en el espacio de trabajo **Diseño**.

Resumen de la creación de tablas de referencia a partir de columnas de perfil

En esta lección, ha aprendido a crear tablas de referencia a partir de los resultados de un perfil para configurar valores válidos para datos de origen.

Ha creado una tabla de referencia a partir de una columna de perfil seleccionando valores válidos para columnas. Ha editado la tabla de referencia para configurar distintos valores válidos para las columnas.

CAPÍTULO 9

Lección 8. Creación de tablas de referencia

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de tablas de referencia, 46](#)
- [Tarea 1. Crear una tabla de referencia, 47](#)
- [Resumen de la creación de tablas de referencia, 47](#)

Introducción a la creación de tablas de referencia

Una tabla de referencia contiene datos de referencia que puede utilizar para estandarizar datos de origen. Los datos de referencia pueden incluir valores válidos y estándar. Cree tablas de referencia para establecer relaciones entre los valores de datos de origen y los valores válidos y estándar.

Se puede crear una tabla de referencia manualmente mediante el editor de tablas de referencia. Utilice la tabla de referencia para definir y estandarizar los datos de origen. Puede compartir la tabla de referencia con un desarrollador para utilizarla en transformaciones de búsqueda o de estandarizador en Developer tool.

Historial

HypoStores desea estandarizar datos para que tengan valores válidos. Usted es el analista responsable de estandarizar los valores válidos en los datos. Desea crear una tabla de referencia para definir códigos de nivel de cliente estándar que hagan referencia a los datos de clientes de Los Ángeles. Más adelante, podrá compartir la tabla de referencia con un desarrollador.

Objetivos

En esta lección, llevará a cabo la tarea siguiente:

- Crear una tabla de referencia con el editor de tablas de referencia para definir códigos de nivel de cliente estándar que hagan referencia a los datos de clientes de Los Ángeles.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 2 de esta guía de autoaprendizaje.

Duración

Dedique 10 minutos a completar las tareas de esta lección.

Tarea 1. Crear una tabla de referencia

En esta tarea, creará la tabla de referencia Reftab_CustomerTier_Codes para estandarizar los valores válidos del los datos del nivel de clientes.

1. Haga clic en **Nueva > Tabla de referencia**.
Aparece el asistente **Nueva tabla de referencia**.
2. Seleccione **Usar el editor de tablas de referencia**.
3. Haga clic en **Siguiente**.
4. Para cada columna que desee incluir en la tabla de referencia, haga clic en el icono **Añadir columna nueva** y configure la propiedades para cada columna.
Añada los siguientes nombres de columna: CustomerID, CustomerTier y Status. Puede reordenar las columnas o borrarlas.
5. Introduzca una descripción opcional y establezca el valor predeterminado como **0**.
La herramienta del analista utiliza el valor predeterminado para cualquier registro de tabla que no contenga ningún valor.
6. Haga clic en **Siguiente**.
7. En el campo **Nombre**, introduzca **Reftab_CustomerTier_Codes**.
8. En la sección **Carpetas**, seleccione la carpeta **Clientes** en el proyecto de la guía de autoaprendizaje.
9. Haga clic en **Finalizar**.
La tabla de referencia aparece en el espacio de trabajo **Diseño**.
10. En el menú **Acciones**, seleccione **Añadir fila** para llenar cada columna de tabla de referencia con los siguientes cuatro valores:
CustomerID = LA1, LA2, LA3, LA4
CustomerTier = 1, 2, 3, 4.
Estado= Activo, Inactivo

Resumen de la creación de tablas de referencia

En esta lección, ha aprendido a crear tablas de referencia mediante el editor de tablas de referencia para crear valores estándar válidos y usarlos con datos de origen.

Ha creado una tabla de referencia mediante el editor de tablas de referencia para estandarizar los valores de niveles de cliente de los datos de clientes de Los Ángeles.

Parte II: Primeros pasos con Informatica Developer

Esta parte incluye los siguientes capítulos:

- [Lección 1. Configuración de Informatica Developer, 49](#)
- [Lección 2: Importar objetos de datos físicos, 53](#)
- [Lección 3. Ejecutar un perfil con datos de origen, 62](#)
- [Lección 4. Análisis de datos, 67](#)
- [Lección 5. Estandarización de datos , 74](#)
- [Lección 6. Validación de datos de dirección, 81](#)

CAPÍTULO 10

Lección 1. Configuración de Informatica Developer

Este capítulo incluye los siguientes temas:

- [Introducción a la configuración de Informatica Developer, 49](#)
- [Tarea 1. Iniciar Informatica Developer, 50](#)
- [Tarea 2. Añadir un dominio, 50](#)
- [Tarea 3. Añadir un repositorio de modelos, 51](#)
- [Tarea 4. Crear un proyecto, 51](#)
- [Tarea 5. Crear una carpeta, 51](#)
- [Tarea 6. Seleccionar un servicio de integración de datos predeterminado., 52](#)
- [Resumen de la configuración de Informatica Developer, 52](#)

Introducción a la configuración de Informatica Developer

Antes de empezar con las lecciones de esta guía de autoaprendizaje, debe iniciar y configurar Developer tool. Para ello, tiene que añadir un dominio. Añada un repositorio de modelos que esté en el dominio y cree un proyecto y una carpeta para guardar el trabajo desarrollado. También debe seleccionar un servicio de integración de datos predeterminado.

El dominio de Informatica es un conjunto de nodos y servicios que definen el entorno de Informatica. Los servicios del dominio incluyen el servicio de repositorio de modelos y el servicio de integración de datos.

El servicio de repositorio de modelos gestiona el repositorio de modelos. El repositorio de modelos es una base de datos relacional que guarda los metadatos para los proyectos que crea en Developer tool. Un proyecto guarda los objetos creados en Developer tool. Asimismo, un proyecto puede contener carpetas que almacenen objetos relacionados, como objetos que formen parte del mismo requisito de negocio.

El servicio de integración de datos realiza las tareas de integración de datos en Developer tool.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

- Iniciar Developer tool e ir al entorno de trabajo de dicha herramienta.

- Añadir un dominio en Developer tool.
- Añadir un repositorio de modelos para poder crear un proyecto.
- Crear un proyecto para almacenar los objetos creados en Developer tool.
- Crear una carpeta dentro del proyecto donde almacenar objetos relacionados.
- Seleccionar un servicio de integración de datos predeterminado para ejecutar las tareas de integración de datos.

Requisitos previos

Antes de empezar con esta lección, compruebe los siguientes requisitos previos:

- Ha instalado Developer tool.
- Dispone de un nombre de dominio, un nombre de host y un número de puerto para conectarse a un dominio. Un administrador del dominio puede facilitarle esta información.
- Un administrador del dominio ha configurado un servicio de repositorio de modelos en la herramienta Administrator.
- Tiene un nombre de usuario y una contraseña para acceder al servicio de repositorio de modelos. Un administrador del dominio puede facilitarle esta información.
- Un administrador del dominio ha configurado un servicio de integración de datos.
- El servicio de integración de datos se está ejecutando.

Duración

Dedique de 5 a 10 minutos a completar las tareas de esta lección.

Tarea 1. Iniciar Informatica Developer

Inicie la herramienta Developer para comenzar la guía de autoaprendizaje.

1. Inicie la herramienta Developer.
Aparece la página **Bienvenida** de la herramienta Developer.
2. Haga clic en el botón **Entorno de trabajo**.
Aparece el entorno de trabajo de la herramienta Developer.

Tarea 2. Añadir un dominio

En esta tarea, se añade un dominio en la herramienta Developer para acceder a un repositorio de modelos.

1. Haga clic en **Ventana > Preferencias**.
Aparece el cuadro de diálogo **Preferencias**.
2. Seleccione **Informatica > Dominios**.
3. Haga clic en **Añadir**.
Aparece el cuadro de diálogo **Nuevo dominio**.
4. Escriba el nombre de dominio, el nombre del host y el número de puerto.

5. Haga clic en **Finalizar**.
6. Haga clic en **Aceptar**.

Tarea 3. Añadir un repositorio de modelos

En esta tarea, se añade el repositorio de modelos que se desea utilizar para almacenar proyectos y carpetas.

1. Haga clic en **Archivo > Conectar a repositorio**.
Aparece el cuadro de diálogo **Conectar a repositorio**.
2. Haga clic en **Explorar** para seleccionar un servicio de repositorio de modelos.
3. Haga clic en **Aceptar**.
4. Haga clic en **Siguiente**.
5. Escriba su nombre de usuario y la contraseña.
6. Seleccione un espacio de nombres.
7. Haga clic en **Finalizar**.
El repositorio de modelos aparece en la vista **Explorador de objetos**.

Tarea 4. Crear un proyecto

En esta tarea, se crea un proyecto para almacenar los objetos creados en la herramienta Developer. Puede crear un proyecto para todas las guías de autoaprendizaje de esta guía.

1. En la vista **Explorador de objetos**, seleccione un servicio de repositorio de modelos.
2. Haga clic en **Archivo > Nuevo > Proyecto**.
Aparece el cuadro de diálogo **Nuevo proyecto**.
3. Escriba su nombre precedido de "Tutorial_" como nombre del proyecto.
4. Haga clic en **Finalizar**.
El proyecto aparece en el servicio de repositorio de modelos, en la vista **Explorador de objetos**.

Tarea 5. Crear una carpeta

En esta tarea, se crea una carpeta para almacenar los objetos relacionados. Puede crear una carpeta para todas las guías de autoaprendizaje de esta guía.

1. En la vista **Explorador de objetos**, seleccione el proyecto al que desee añadir la carpeta.
2. Haga clic en **Archivo > Nueva > Carpeta**.
3. Escriba un nombre para la carpeta.

4. Haga clic en **Finalizar**.

La herramienta Developer añade la carpeta en el proyecto, en la vista **Explorador de objetos**. Expanda el proyecto para ver la carpeta.

Tarea 6. Seleccionar un servicio de integración de datos predeterminado.

En esta tarea, se selecciona un servicio de integración de datos predeterminado para poder ejecutar asignaciones y previsualizar datos.

1. Haga clic en **Ventana > Preferencias**.
Aparece el cuadro de diálogo **Preferencias**.
2. Seleccione **Informatica > Servicios de integración de datos**.
3. Expanda el dominio.
4. Seleccione un servicio de integración de datos.
5. Haga clic en **Establecer como predeterminado**.
6. Haga clic en **Aceptar**.

Resumen de la configuración de Informatica Developer

En esta lección, ha aprendido que el dominio Informatica incluye el servicio de repositorio de modelos y el servicio de integración de datos. El servicio de repositorio de modelos gestiona el repositorio de modelos. Un repositorio de modelos contiene proyectos y carpetas. El servicio de integración de datos realiza tareas de integración de datos.

Ha iniciado y configurado la herramienta Developer. Ha añadido un dominio a la herramienta Developer, ha añadido un repositorio de modelos y ha creado un proyecto y una carpeta. También ha seleccionado un servicio de integración de datos predeterminado.

Ya puede utilizar la herramienta Analyst para llevar a cabo otras lecciones de esta guía de autoaprendizaje.

CAPÍTULO 11

Lección 2: Importar objetos de datos físicos

Este capítulo incluye los siguientes temas:

- [Introducción a la importación de objetos de datos físicos, 53](#)
- [Tarea 1. Importar el objeto de datos del archivo sin formato Boston_Customers, 54](#)
- [Tarea 2. Importar el objeto de datos de archivo sin formato LA_Customers, 60](#)
- [Tarea 3. Importar el objeto de datos del archivo sin formato All_Customers, 61](#)
- [Resumen de la importación de objetos de datos físicos, 61](#)

Introducción a la importación de objetos de datos físicos

Un objeto de datos es una representación de datos basados en un archivo sin formato o tabla de base de datos relacional. Se puede importar un archivo sin formato o una tabla de base de datos relacional como objeto de datos físicos para usarlos como origen o destino en una asignación.

Historial

HypoStores Corporation almacena datos de sus oficinas de Los Ángeles y Boston en archivos sin formato. Usted quiere trabajar con estos datos de clientes en Developer tool. Para ello, tiene que importar cada archivo sin formato como un objeto de datos físicos.

Objetivos

En esta lección, se importan archivos sin formato como objetos de datos físicos. También se especifica el directorio del archivo de origen para que el servicio de integración de datos pueda leer los datos de origen desde el directorio correcto.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado la lección 1 de esta guía de autoaprendizaje.

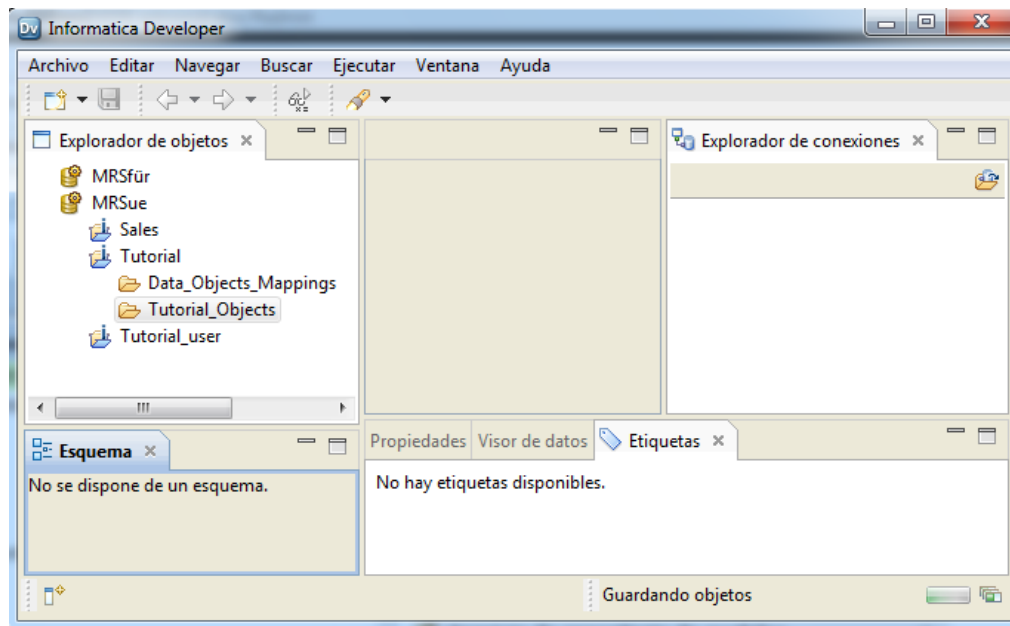
Duración

Destine de 10 a 15 minutos para realizar las tareas de esta lección.

Tarea 1. Importar el objeto de datos del archivo sin formato Boston_Customers

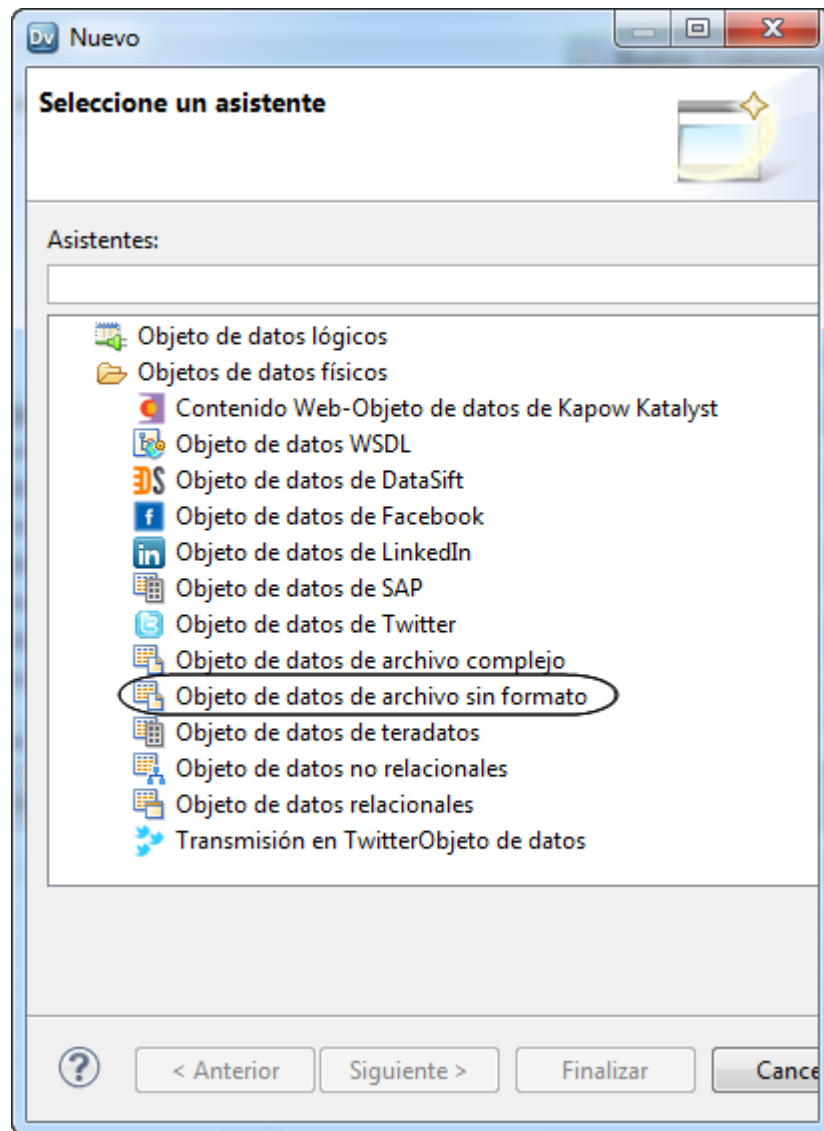
En esta tarea, tiene que importar un objeto de datos físicos desde un archivo que contiene datos de clientes de la oficina de Boston.

1. En la vista **Explorador de objetos**, seleccione la carpeta Tutorial_Objects.



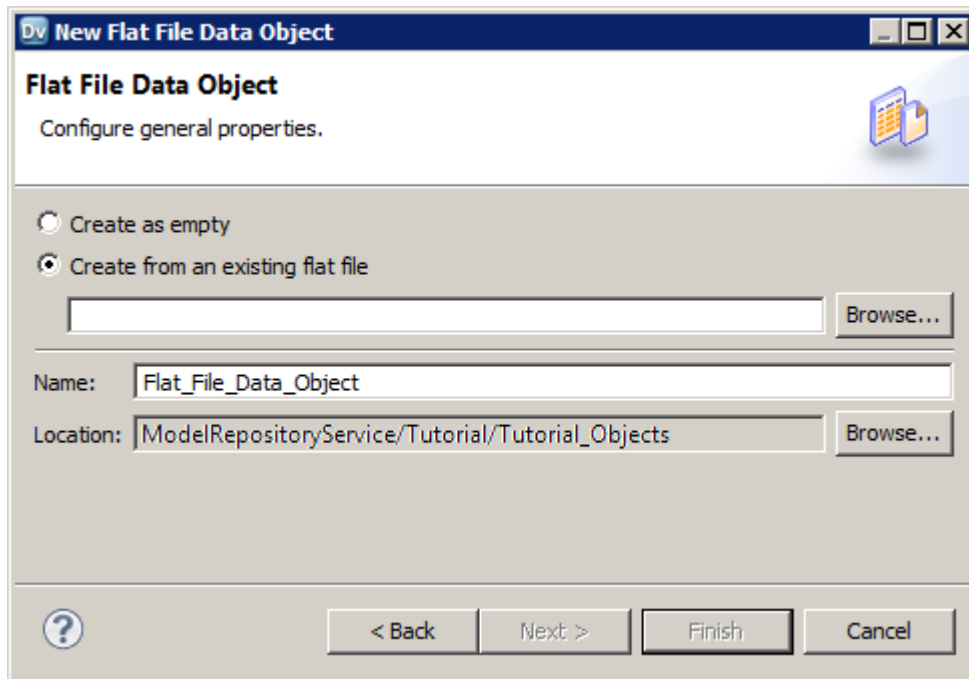
2. Haga clic con el botón derecho en la carpeta Tutorial_Objects y seleccione **Nuevo > Objeto de datos**.

A continuación, aparecerá el cuadro de diálogo **Nuevo**.



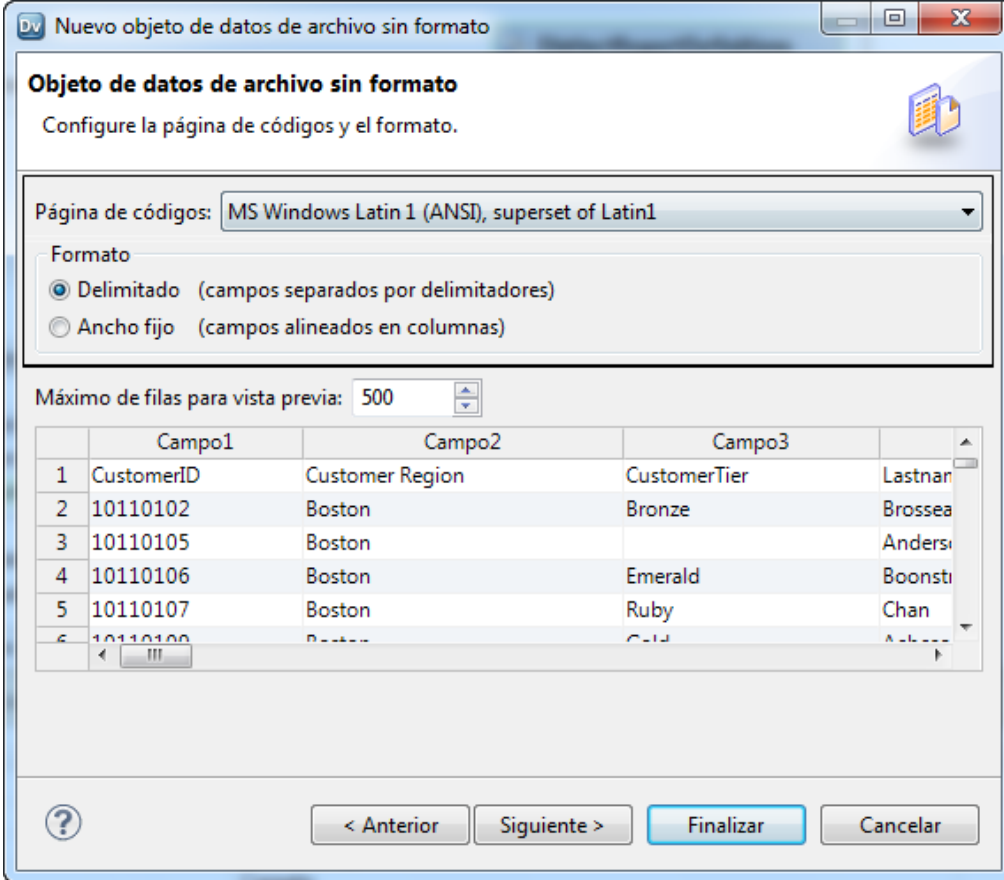
3. Seleccione **Objetos de datos físicos** > **Objeto de datos de archivo sin formato** y haga clic en **Siguiete**.

A continuación, aparecerá el cuadro de diálogo **Nuevo objeto de datos de archivo sin formato**.



4. Seleccione **Crear desde un archivo sin formato existente**.
5. Haga clic en **Examinar** y desplácese hasta Boston_Customers.csv en el directorio siguiente en el equipo de Developer Tool: <directorio de instalación de Informatica>\clientes\DeveloperClient\Guías de aprendizaje
6. Haga clic en **Abrir**.
El asistente nombra al objeto de datos como "Boston_Customers".
7. Haga clic en **Siguiente**.
8. Compruebe que la página de códigos esté establecida en **MS Windows Latín 1 (ANSI)**, **superconjunto de Latín 1** y que el formato esté establecido en **Delimitado**.

El cuadro de diálogo **Nuevo objeto de datos de archivo sin formato** muestra la página de códigos predeterminada, el formato y una vista previa de los datos del archivo sin formato.



Nuevo objeto de datos de archivo sin formato

Configure la página de códigos y el formato.

Página de códigos: MS Windows Latin 1 (ANSI), superset of Latin1

Formato

☒ Delimitado (campos separados por delimitadores)

☐ Ancho fijo (campos alineados en columnas)

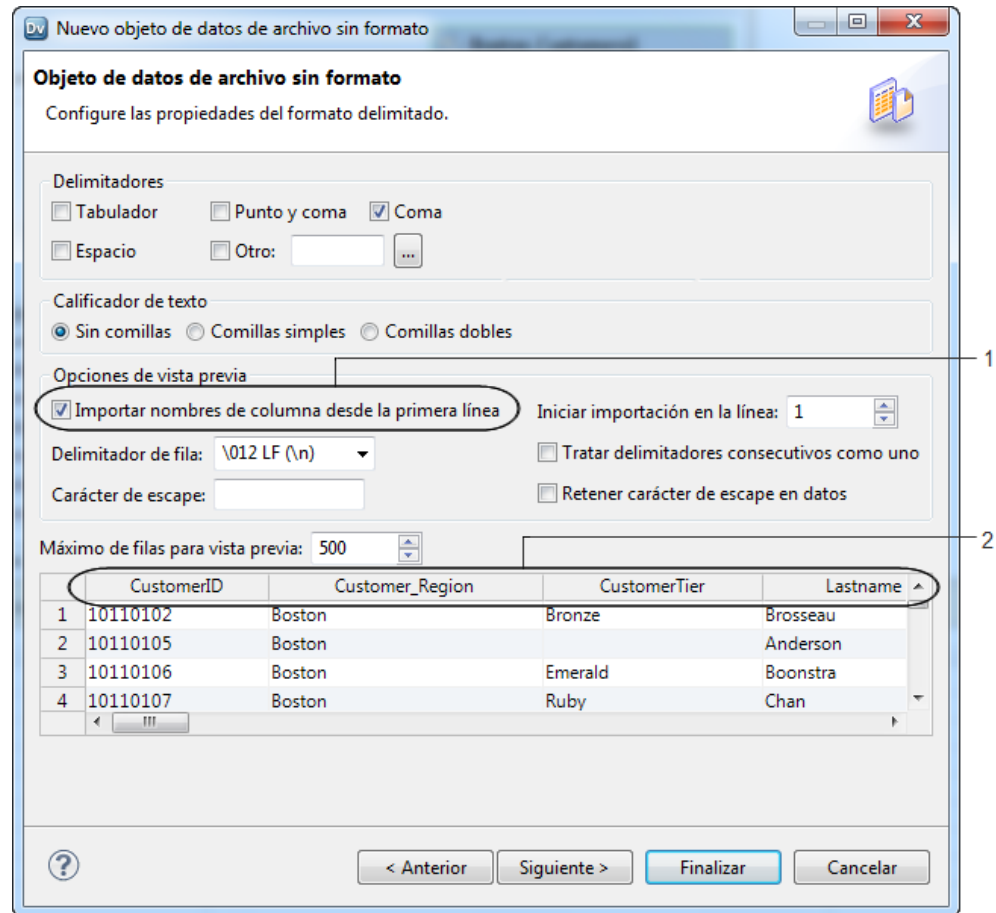
Máximo de filas para vista previa: 500

	Campo1	Campo2	Campo3	
1	CustomerID	Customer Region	CustomerTier	Lastname
2	10110102	Boston	Bronze	Brosseau
3	10110105	Boston		Anderson
4	10110106	Boston	Emerald	Boons
5	10110107	Boston	Ruby	Chan
6	10110108	Boston	Silver	Chen

? < Anterior Siguiente > Finalizar Cancelar

9. Haga clic en **Siguiente**.
10. Seleccione **Importar nombres de columna desde la primera línea**.

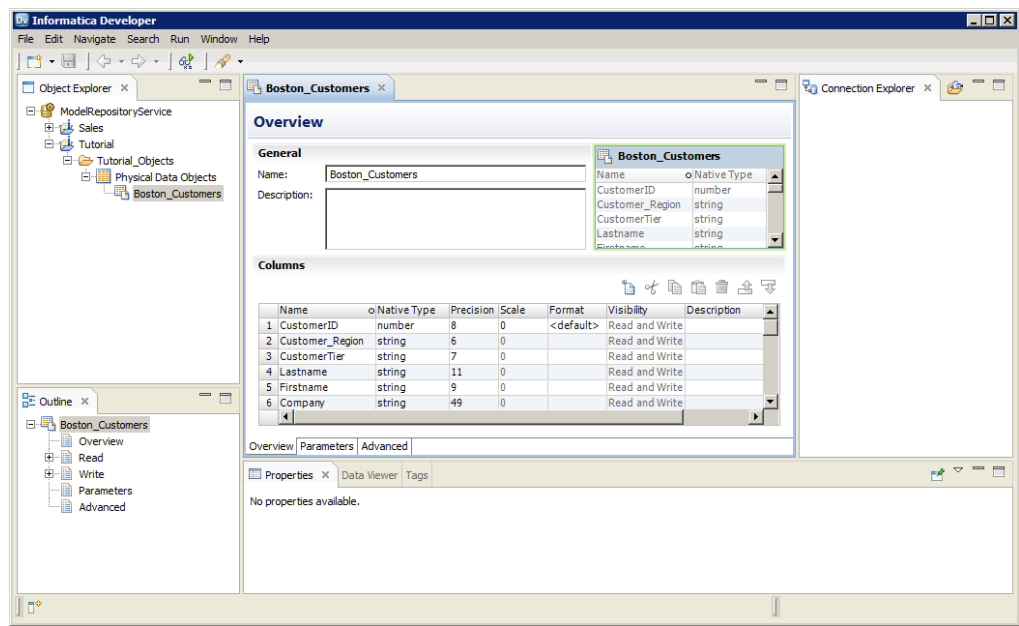
El cuadro de diálogo **Nuevo objeto de datos de archivo sin formato** muestra los nombres de columna en la vista previa de los datos de archivo sin formato.



1. La opción **Importar nombres de columna desde la primera línea**
2. Nombres de columna

11. Haga clic en **Finalizar**.

El objeto de datos físicos Boston_Customers aparece en la carpeta Objetos de datos físicos dentro de la carpeta Tutorial_Objects. La vista **Resumen** muestra el contenido del archivo y se abre en el editor.



12. Haga clic en la vista **Opciones avanzadas**.
La vista **Avanzadas** muestra las propiedades del objeto de datos físicos.
13. En la vista **Avanzadas**, desplácese hasta la sección **Tiempo de ejecución: Lectura**.
14. En la sección **Tiempo de ejecución: Lectura**, establezca **Directorio de archivo de origen** en el siguiente directorio del equipo del servicio de integración de datos: <directorio de instalación de Informatica>\server\Tutorials.

El servicio de integración de datos buscará el archivo de origen en el directorio del servidor del equipo donde se ejecuta dicho servicio. La instalación del servidor contiene una copia de los archivos de la guía de autoaprendizaje. El servicio de integración de datos no puede leer los archivos desde el directorio de instalación del cliente a menos que se cambien los permisos de acceso en el archivo de origen y el directorio.

La siguiente figura muestra un ejemplo del directorio de archivo de origen:

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

Nota: El equipo de Developer Tool debe tener acceso al directorio del archivo de origen en el equipo que ejecuta el servicio de integración de datos. Si Developer Tool no puede acceder al directorio del archivo de origen, dicha herramienta no podrá obtener una vista previa de los datos del archivo de origen ni ejecutar asignaciones que accedan a los datos contenidos en este. Si ejecuta varios servicios de integración de datos, hay un directorio de archivo de origen independiente para cada uno de ellos.

15. Haga clic en la vista **Visor de datos**.

16. En la vista **Visor de datos**, haga clic en **Ejecutar**.
El servicio de integración de datos lee los datos del archivo Boston_Customers y muestra los resultados en la ventana Salida.
17. Haga clic en **Archivo > Guardar** para guardar el objeto de datos físicos Boston_Customers.

Tarea 2. Importar el objeto de datos de archivo sin formato LA_Customers

En esta tarea, tiene que importar un objeto de datos físicos desde un archivo sin formato que contiene datos de clientes de la oficina de Los Ángeles.

1. En la vista **Explorador de objetos**, seleccione el proyecto de la guía de autoaprendizaje.
2. Haga clic en **Archivo > Nuevo > Objeto de datos**.
A continuación, aparece el cuadro de diálogo **Nuevo**.
3. Seleccione **Objetos de datos físicos > Objeto de datos de archivo sin formato** y haga clic en **Siguiente**.
A continuación, aparece el cuadro de diálogo **Nuevo objeto de datos de archivo sin formato**.
4. Seleccione **Crear desde un archivo sin formato existente**.
5. Haga clic en **Examinar** y desplácese hasta LA_Customers.csv en el siguiente directorio: <Directorio de instalación de Informatica>\clients\DeveloperClient\Tutorials
6. Haga clic en **Abrir**.
El asistente nombra el objeto de datos como LA_Customers.
7. Haga clic en **Siguiente**.
8. Compruebe que la página de códigos sea MS Windows Latín 1 (ANSI), supraconjunto de Latín 1.
9. Compruebe que el formato esté delimitado.
10. Haga clic en **Siguiente**.
11. Compruebe que el delimitador establecido se corresponda con la coma.
12. Seleccione **Importar nombres de columna desde la primera línea**.
13. Haga clic en **Finalizar**.
El objeto de datos físicos LA_Customers aparece en **Objetos de datos físicos**, en el proyecto de la guía de autoaprendizaje.
14. Haga clic en **Lectura** y seleccione la transformación de salida.
15. Haga clic en la ficha **Tiempo de espera**, en la vista **Propiedades**.
16. Defina el directorio del archivo de origen como el directorio siguiente en el equipo del servicio de integración de datos: <Directorio de instalación de Informatica>\server\Tutorials
17. Haga clic en **Archivo > Guardar**.

Tarea 3. Importar el objeto de datos del archivo sin formato All_Customers

En esta tarea, tiene que importar un objeto de datos físicos desde un archivo sin formato que contiene datos de clientes de las oficinas de Los Ángeles y Boston.

1. En la vista **Explorador de objetos**, seleccione el proyecto de la guía de autoaprendizaje.
2. Haga clic en **Archivo > Nuevo > Objeto de datos**.
A continuación, aparece el cuadro de diálogo **Nuevo**.
3. Seleccione **Objetos de datos físicos > Objeto de datos de archivo sin formato** y haga clic en **Siguiente**.
Aparece el cuadro de diálogo **Nuevo origen de datos de archivo sin formato**.
4. Seleccione **Crear desde un archivo sin formato existente**.
5. Haga clic en **Examinar** y desplácese hasta All_Customers.csv en el siguiente directorio: <Directorio de instalación de Informatica>\clients\DeveloperClient\Tutorials.
6. Haga clic en **Abrir**.
El asistente nombra el objeto de datos como All_Customers.
7. Haga clic en **Siguiente**.
8. Compruebe que la página de códigos sea MS Windows Latín 1 (ANSI), supraconjunto de Latín 1.
9. Compruebe que el formato esté delimitado.
10. Haga clic en **Siguiente**.
11. Compruebe que el delimitador establecido se corresponda con la coma.
12. Seleccione **Importar nombres de columna desde la primera línea**.
13. Haga clic en **Finalizar**.
El objeto de datos físicos All_Customers aparece en **Objetos de datos físicos**, en el proyecto de la guía de autoaprendizaje.
14. Haga clic en **Lectura** y seleccione la transformación de salida.
15. Haga clic en la ficha **Tiempo de espera**, en la vista **Propiedades**.
16. Defina el directorio del archivo de origen como el directorio siguiente en el equipo del servicio de integración de datos: <Directorio de instalación de Informatica>\server\Tutorials
17. Haga clic en **Archivo > Guardar**.

Resumen de la importación de objetos de datos físicos

En esta lección, ha aprendido que los objetos de datos físicos son representaciones de datos basadas en un archivo sin formato o en una base de datos relacional.

Ha creado objetos de datos físicos a partir de archivos sin formato. También se especifica el directorio del archivo de origen para que el servicio de integración de datos pueda leer los datos de origen desde el directorio correcto.

Los objetos de datos se utilizan como orígenes de asignación en las lecciones sobre calidad de datos.

CAPÍTULO 12

Lección 3. Ejecutar un perfil con datos de origen

Este capítulo incluye los siguientes temas:

- [Introducción a la creación de perfiles de datos, 62](#)
- [Tarea 1. Realizar un análisis de combinación en dos orígenes de datos, 63](#)
- [Tarea 2. Ver los resultados de un análisis de combinación, 64](#)
- [Tarea 3. Ejecutar un perfil en un origen de datos, 65](#)
- [Tarea 4. Ver los resultados de perfil de columna, 65](#)
- [Resumen de datos de creación de perfiles, 66](#)

Introducción a la creación de perfiles de datos

Un perfil es un conjunto de metadatos que describe el contenido y la estructura de un conjunto de datos.

La creación de perfiles y la obtención de datos suelen ser los primeros pasos de un proyecto. Puede ejecutar un perfil para evaluar la estructura de datos y para verificar que las columnas de datos contengan los tipos de información que usted espera. En el caso de que un perfil revele problemas en los datos, puede definir pasos en el proyecto para corregir dichos problemas. Por ejemplo, si un perfil revela que una columna contiene valores de una longitud mayor que la esperada, puede diseñar procesos de calidad de datos para eliminar o corregir los valores problemáticos.

Un perfil que analiza la calidad de datos de las columnas seleccionadas se llama perfil de columna.

Nota: También puede usar Developer tool para descubrir la clave primaria, la clave externa y relaciones de dependencia funcional, así como analizar condiciones de combinación en columnas de datos.

Un perfil de columna proporciona los siguientes hechos sobre los datos:

- La cantidad de valores únicos y nulos de cada columna, expresada en forma de número y de porcentaje.
- Los patrones de datos en cada columna y las frecuencias con que ocurren estos valores.
- Estadísticas sobre los valores de columna, tales como las longitudes máxima y mínima de los valores, y el primer y último valor de cada columna.
- Para perfiles de análisis de combinación, el grado de solapamiento entre dos columnas de datos, mostrado como diagrama de Venn y como valor de porcentaje. Utilice perfiles de análisis de combinación para identificar posibles problemas con condiciones de unión de columnas.

Puede ejecutar un perfil de columna en cualquier etapa de un proyecto para medir la calidad de los datos y comprobar que los cambios efectuados en los datos cumplan los objetivos del proyecto. Puede ejecutar un perfil de columna en una transformación de una asignación para indicar el efecto que ejercerá la transformación en los datos.

Ejemplo

HypoStores desea comprobar que los datos de los clientes no contengan errores, inconsistencias ni información duplicada. Antes de que HypoStores diseñe los procesos para lograr los objetivos de calidad de datos, necesita cuantificar la calidad de sus archivos de datos de origen y confirmar que los datos están listos para procesarse.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

- Realizar un análisis de combinación en el origen de datos `Boston_Customers` y en el origen de datos `LA_Customers`.
- Visualizar los resultados del análisis de combinación para determinar si se puede o no fusionar correctamente los datos de las dos oficinas.
- Ejecutar un perfil en el origen de datos `All_Customers`.
- Visualizar los resultados de la creación de perfil de columna para observar los valores y patrones de los datos.

Requisitos previos

Antes de empezar con esta lección, verifique los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 2 de esta guía de autoaprendizaje.

Tiempo necesario

- Dedique 20 minutos a completar esta lección.

Tarea 1. Realizar un análisis de combinación en dos orígenes de datos

En esta tarea, se realiza un análisis de combinación de los orígenes de datos `Boston_Customers` y `LA_Customers` para ver las condiciones de combinación.

1. Seleccione la carpeta del autoaprendizaje y haga clic en **Archivo > Nuevo > Perfil**.
2. Seleccione **Perfil de detección empresarial**.
3. Haga clic en **Siguiente**.
4. En el campo **Nombre**, introduzca **Tutorial_Profile**.
5. Haga clic en **Finalizar**.
El perfil `Tutorial_Profile` aparece en el explorador de objetos.
6. Arrastre los orígenes de datos **`Boston_Customers`** y **`LA_Customers`** al editor en el lado derecho.
Sugerencia: Mantenga pulsada la tecla Mayús para seleccionar varios objetos de datos.
7. Haga clic con el botón derecho sobre el nombre de un objeto de datos y seleccione **Perfil de combinación**.
Aparece el asistente **Nuevo perfil de combinación**.

8. En el campo **Nombre**, introduzca **Análisis de combinación**.
9. Compruebe que Boston_Customers y LA_Customers aparecen como objetos de datos y haga clic en **Siguiente**.
10. Compruebe que la columna **CustomerID** esté seleccionada en los dos orígenes de datos. Desplácese por el panel del asistente para ver las columnas de ambos conjuntos de datos. Haga clic en **Siguiente**.
11. Haga clic en **Añadir** para añadir condiciones de combinación. Aparece el cuadro de diálogo **Condición de combinación**.
12. En la sección **Columnas**, haga clic en **Añadir fila**.
13. Haga doble clic en la primera fila de la columna de la izquierda y seleccione **CustomerID**.
14. Haga doble clic en la primera fila de la columna de la derecha y seleccione **CustomerID**.
15. Haga clic en **Aceptar** y, a continuación, haga clic en **Finalizar**.
16. Si la Herramienta del desarrollador le solicita guardar los cambios, haga clic en **Sí**. La Herramienta del desarrollador ejecuta el perfil.

Nota: No cierre el perfil. Verá los resultados del perfil en la siguiente tarea.

Tarea 2. Ver los resultados de un análisis de combinación

En esta tarea, se visualizan los resultados de un análisis de combinación en la vista Resultado de combinación del perfil JoinAnalysis.

1. Haga clic en la ficha **JoinAnalysis** en el editor.
2. En la sección **Resultado de combinación**, haga clic en la primera fila.
La sección **Detalles** muestra un diagrama de Venn y la clave de colores que detalla los resultados del análisis de combinación.
3. Compruebe que la columna **Filas de combinación** dé como resultado cero para el número de filas que contienen una combinación.
Este valor indica que los campos CustomerID no tienen valores duplicados. Se pueden fusionar correctamente los dos orígenes de datos.
4. Para ver los valores de CustomerID del objeto de datos LA_Customers, haga doble clic en el círculo con el nombre **LA_Customers** en el diagrama de Venn.

Sugerencia: Haga doble clic en los círculos del diagrama de Venn para ver las filas de datos. Si los círculos forman intersecciones en el diagrama de Venn, haga doble clic en la intersección para ver los valores de datos comunes a ambos conjuntos de datos.

El visor de datos muestra los valores de CustomerID del objeto de datos LA_Customers.

Tarea 3. Ejecutar un perfil en un origen de datos

En esta tarea, se ejecuta un perfil en el origen de datos All_Customers para visualizar el contenido y la estructura de los datos.

1. En la vista **Explorador de objetos**, localice los objetos de datos del proyecto de autoaprendizaje.
2. Seleccione el origen de datos **All_Customers**.
3. Haga clic en **Archivo > Nuevo > Perfil**.
A continuación, aparece el cuadro de diálogo **Nuevo**.
4. Seleccione **Perfil**.
5. Haga clic en **Siguiente**.
6. En el campo **Nombre**, introduzca **All_Customers**.
7. Haga clic en **Finalizar**.

El perfil All_Customers se abre en el editor y se ejecuta el perfil.

Tarea 4. Ver los resultados de perfil de columna

En esta tarea, se visualizan los resultados de creación de perfiles de columna del objeto de datos All_Customers y se examinan los valores y patrones contenidos en los datos.

1. Haga clic en **Ventana > Mostrar vista > Progreso** para ver el progreso del perfil All_Customers.
Se abre la vista Progreso.
2. Cuando la vista Progreso comunique que el perfil All_Customers ha terminado de ejecutarse, haga clic en la vista **Resultados** en el editor.
3. En la sección **Creación de perfiles de columna**, haga clic en la columna **CustomerTier**.
La sección **Detalles** muestra todos los valores contenidos en la columna CustomerTier y muestra información sobre la frecuencia de aparición de los valores en el conjunto de datos.
4. En la sección **Detalles**, haga doble clic **Rubí**.
El visor de datos se ejecuta y muestra los registros en los que la columna CustomerTier contiene el valor **Rubí**.
5. En la sección **Creación de perfiles de columna**, haga clic en la columna **OrderAmount**.
6. En la sección **Detalles**, haga clic en la lista **Mostrar** y seleccione **Patrones**.
La sección **Detalles** muestra los patrones encontrados en la columna OrderAmount. La cadena **9(5)** de la columna Patrón se refiere a los registros que contienen importes de pedidos de cinco cifras. La cadena **9(4)** se refiere a registros que contienen importes de pedidos de cuatro cifras.
7. En la columna **Patrón**, haga doble clic en la cadena **9(4)**.
El visor de datos se ejecuta y muestra los registros en los que la columna OrderAmount contiene un importe de pedido de cuatro cifras.
8. En la sección **Detalles**, haga clic en la lista **Mostrar** y seleccione **Estadísticas**.
La sección **Detalles** muestra estadísticas de la columna OrderAmount, incluidos el valor promedio, la desviación estándar, las longitudes máxima y mínima, los cinco valores más comunes y los cinco valores menos comunes.

Resumen de datos de creación de perfiles

En esta lección, ha aprendido que un perfil proporciona información sobre el contenido y la estructura de los datos.

Ha aprendido que puede realizar un análisis de unión de dos objetos de datos y visualizar el grado de solapamiento entre los mismos. También ha aprendido que puede ejecutar un perfil de columna en un objeto de datos y visualizar los valores, patrones y estadísticas relacionados con cada columna del objeto de datos.

Ha creado el perfil `JoinAnalysis` para determinar si los datos del objeto de datos `Boston_Customers` pueden fusionarse con los datos del objeto de datos `LA_Customers`. Ha visualizado los resultados de este perfil y ha determinado que todos los valores de la columna `CustomerID` son únicos y que puede fusionar los objetos de datos correctamente.

Ha creado el perfil `All_Customers` y ha ejecutado un perfil de columna en el objeto de datos `All_Customers`. Ha visualizado los resultados de este perfil para descubrir valores, patrones y estadísticas para las columnas en el objeto de datos `All_Customers`. Por último, ha ejecutado el visor de datos para visualizar filas que contienen los valores y patrones seleccionados, permitiéndole verificar la calidad de los datos.

CAPÍTULO 13

Lección 4. Análisis de datos

Este capítulo incluye los siguientes temas:

- [Resumen del análisis de datos, 67](#)
- [Tarea 1. Crear un objeto de datos de destino, 68](#)
- [Tarea 2. Crear una asignación para analizar datos, 70](#)
- [Tarea 3. Ejecutar un perfil en la transformación de analizador, 72](#)
- [Tarea 4. Ejecutar la asignación, 72](#)
- [Tarea 5. Ver la salida de la asignación, 72](#)
- [Resumen del análisis de datos, 73](#)

Resumen del análisis de datos

Puede analizar los datos para identificar uno o más elementos de datos de un campo de entrada y escribir cada uno de ellos en un campo de salida distinto.

El análisis le permite tener un mayor control de la información de cada columna. Por ejemplo, considere un campo de datos que contenga el nombre completo de una persona, `Bob Smith`. Puede usar la transformación de analizador para partir el nombre completo en columnas de datos independientes que contengan el nombre y el apellido. Después de analizar los datos de las nuevas columnas, puede crear operaciones de calidad de datos personalizadas para cada una de ellas.

Puede configurar la transformación de analizador para que utilice *conjuntos de tokens* para analizar columnas de datos como cadenas de componentes. Un conjunto de tokens identifica elementos de datos como palabras, códigos postales, números de teléfono y números de seguridad social.

También puede usar la transformación de analizador para analizar datos que coincidan con entradas de una tabla de referencia o con expresiones regulares personalizadas que introduzca.

Historial

HypoStores desea que el formato de los archivos de datos de clientes de la oficina de Los Ángeles coincida con el formato de los archivos de datos de la oficina de Boston. Los datos de clientes de la oficina de Los Ángeles almacenan el nombre del cliente en una columna, `FullName`, mientras que los datos de clientes de la oficina de Boston almacenan el nombre del cliente en dos columnas diferenciadas, `FirstName` y `LastName`. HypoStores necesita analizar los datos de la columna `FullName` de Los Ángeles como nombres y apellidos, de modo que el formato de los datos de Los Ángeles coincida con el formato de los datos de Boston.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

- Crear y configurar un objeto de datos `LA_Customers_tgt` para que contenga datos analizados.
- Crear una asignación para analizar la columna `FullName` como dos columnas diferenciadas, `FirstName` y `LastName`.
- Añadir el objeto de datos `LA_Customers` a la asignación para conectarlo a los datos de origen.
- Añadir el objeto de datos `LA_Customers_tgt` a la asignación para crear un objeto de datos de destino.
- Añadir una transformación de analizador a la asignación y configurarla para que use un conjunto de tokens para analizar nombres completos como nombres y apellidos.
- Ejecutar un perfil en la transformación de analizador para revisar los datos antes de generar el origen de datos de destino.
- Ejecutar la asignación para generar nombres analizados.
- Ejecutar el visor de datos para visualizar la salida de la asignación.

Requisitos previos

Antes de empezar con esta lección, compruebe que cumple los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 2 de esta guía de autoaprendizaje.

Duración

Dedique 20 minutos a completar las tareas de esta lección.

Tarea 1. Crear un objeto de datos de destino

En esta tarea, se crea un objeto de datos `LA_Customers_tgt` en el que se pueden escribir nombres analizados.

Para crear un objeto de datos de destino, lleve a cabo los siguientes pasos:

1. Cree un objeto de datos `LA_Customers_tgt` basado en el archivo `LA_Customers.csv`.
2. Configure las opciones de lectura y escritura para el objeto de datos, incluyendo los nombres y ubicaciones de los archivos.
3. Añada las columnas `Firstname` y `Lastname` al objeto de datos `LA_Customers_tgt`.

Paso 1. Crear un objeto de datos `LA_Customers`

En este paso, se crea un objeto de datos `LA_Customers_tgt` basado en el archivo `LA_Customers.csv`.

1. Haga clic en **Archivo > Nuevo > Objeto de datos**.
Se abre la ventana **Nuevo**.
2. Seleccione **Objeto de datos de archivo sin formato** y haga clic en **Siguiente**.
3. Compruebe que **Crear a partir de archivo sin formato existente** esté seleccionada.
4. Haga clic en **Explorar** y localice `LA_Customers.csv` en el siguiente directorio: `<Directorio de instalación de Informatica>\clients\DeveloperClient\Tutorials`
5. Haga clic en **Abrir**.

6. En el campo **Nombre**, introduzca `LA_Customers_tgt`.
7. Haga clic en **Siguiente**.
8. Haga clic en **Siguiente**.
9. En la sección **Previsualizar opciones**, seleccione **Importar nombres de columna desde la primera línea** y haga clic en **Siguiente**.
10. Haga clic en **Finalizar**.
El objeto de datos `LA_Customers_tgt` aparece en el editor.

Paso 2. Configurar opciones de lectura y escritura

En este paso, se configuran las opciones de lectura y escritura para el objeto de datos `LA_Customers_tgt`, incluyendo los nombres y ubicaciones de los archivos.

1. Compruebe que el objeto de datos `LA_Customers_tgt` esté abierto en el editor.
2. En el editor, seleccione la vista **Lectura**.
3. Haga clic en **Ventana > Mostrar vista > Propiedades**.
4. En la vista **Propiedades**, seleccione la vista **Tiempo de ejecución**.
5. En la columna **Valor**, haga doble clic en el nombre del archivo de origen y escriba `LA_Customers_tgt.csv`.
6. En la columna **Valor**, haga doble clic para resaltar el directorio del archivo de origen.
7. Haga clic con el botón derecho sobre el nombre resaltado y seleccione **Copiar**.
8. En el editor, seleccione la vista **Escritura**.
9. En la vista **Propiedades**, seleccione la vista **Tiempo de ejecución**.
10. En la columna **Valor**, haga doble clic en la entrada **Directorio de archivo de salida**.
11. Haga clic con el botón derecho y seleccione **Pegar** para pegar la ubicación que ha copiado desde la vista **Lectura**.
12. En la columna **Valor**, haga doble clic en la entrada **Opciones de encabezado** y elija **Nombres de campo de salida**.
13. En la columna **Valor**, haga doble clic en el **Nombre de archivo de salida** y escriba `LA_Customers_tgt.csv`.
14. Haga clic en **Archivo > Guardar** para guardar el objeto de datos.

Paso 3. Añadir columnas al objeto de datos

En este paso, se añaden las columnas `Firstname` y `Lastname` al objeto de datos `LA_Customers_tgt`.

1. En la vista Explorador de objetos, localice los objetos de datos del proyecto de la guía de autoaprendizaje.
2. Haga doble clic en el objeto de datos `LA_Customers_tgt`.
El objeto de datos `LA_Customers_tgt` se abre en el editor.
3. Compruebe que la vista **Resumen** esté seleccionada.
4. Seleccione la columna **FullName** y haga clic en el botón **Nueva** para añadir una columna.
Aparece una columna llamada `FullName1`.
5. Cambie el nombre de la columna a `Firstname`. Haga clic en el campo **Precisión** e introduzca "30".

6. Seleccione la columna `Firstname` y haga clic en el botón **Nueva** para añadir una columna.
Aparece una columna llamada `FirstName1`.
7. Cambie el nombre de la columna a `Lastname`. Haga clic en el campo **Precisión** e introduzca "30".
8. Haga clic en **Archivo > Guardar** para guardar el objeto de datos.

Tarea 2. Crear una asignación para analizar datos

En esta tarea, se crea una asignación y se configura para usar objetos de datos y una transformación del analizador.

Para crear una asignación para analizar datos, lleve a cabo los siguientes pasos:

1. Cree una asignación.
2. Añada objetos de datos de origen y de destino a la asignación.
3. Añada una transformación de analizador a la asignación.
4. Configure la transformación del analizador para analizar la columna origen que contiene el nombre completo del cliente en dos columnas individuales que contengan el nombre y el apellido.

Paso 1. Crear una asignación

En este paso, se crea y da nombre a la asignación.

1. En la vista Explorador de objetos, seleccione su proyecto de guía de autoaprendizaje.
2. Haga clic en **Archivo > Nueva > Asignación**.
Se abre la ventana **Nueva asignación**.
3. En el campo **Nombre**, introduzca `ParserMapping`.
4. Haga clic en **Finalizar**.
La asignación se abre en el editor.

Paso 2. Añadir objetos de datos a la asignación

En este paso, se añade el objeto de datos `LA_Customers` y el objeto de datos `LA_Customers_tgt` a la asignación.

1. En la vista Explorador de objetos, localice los objetos de datos de la guía de autoaprendizaje.
2. Seleccione el objeto de datos `LA_Customers` y arrástrelo hasta el editor.
Se abre la ventana **Añadir objeto de datos físicos a asignación**.
3. Compruebe que **Lectura** esté seleccionado y haga clic en **Aceptar**.
El objeto de datos aparece en el editor.
4. En la vista Explorador de objetos, localice los objetos de datos de la guía de autoaprendizaje.
5. Seleccione el objeto de datos `LA_Customers_tgt` y arrástrelo hasta el editor.
Se abre la ventana **Añadir objeto de datos físicos a asignación**.
6. Seleccione **Escritura** y haga clic en **Aceptar**.
El objeto de datos aparece en el editor.

7. Seleccione los puertos `CustomerID`, `CustomerTier` y `FullName` en el objeto de datos `LA_Customers`. Arrastre los puertos hasta el puerto `CustomerID` del objeto de datos `LA_Customers_tgt`.

Sugerencia: Mantenga pulsada la tecla **CTRL** para seleccionar varios puertos.

Los puertos del objeto de datos `LA_Customers` se conectan con los puertos correspondientes en el objeto de datos `LA_Customers_tgt`.

Paso 3. Añadir una transformación de analizador a la asignación

En este paso, se añade una transformación de analizador a la asignación `ParserMapping`.

1. Seleccione el editor que contenga la asignación `ParserMapping`.
2. En la paleta Transformación, seleccione la transformación de analizador.
3. Haga clic en el editor.
Se abre la ventana **Nueva transformación de analizador**.
4. Compruebe que **Analizador de tokens** esté seleccionado y haga clic en **Finalizar**.
La transformación del analizador aparece en el editor.
5. Seleccione el puerto `FullName` en el objeto de datos `LA_Customers` y arrastre el puerto al grupo Entrada de la transformación del analizador.
El puerto `FullName` aparece en la transformación de analizador y se conecta con el puerto `FullName` del objeto de datos.

Paso 4. Configurar la transformación de analizador

En este paso, se configura la transformación de analizador para analizar la columna que contiene el nombre completo de un cliente en columnas diferenciadas que contienen el nombre y el apellido.

1. Seleccione el editor que contenga la asignación `ParserMapping`.
2. Haga clic en la transformación de analizador.
3. Haga clic en **Ventana > Mostrar vista > Propiedades**.
4. En la vista Propiedades, seleccione la vista **Estrategias**.
5. Haga clic en **Nueva**. Se muestra el asistente de nueva estrategia.
6. Haga clic en la flecha de selección en la columna Entradas y elija el puerto `FullName`.
7. Seleccione el delimitador de espacio de carácter `[\s]`.
8. Haga clic en **Siguiente**.
9. Seleccione la operación **Analizar con conjunto de tokens** y haga clic en **Siguiente**.
10. Seleccione **Conjuntos de tokens fijos (sólo salida individual)** y elija el conjunto de tokens **No definido**.
11. Haga clic en el campo **Salidas** y seleccione **Nueva**.
12. En el cuadro de diálogo **Salidas de operación**, cambie el nombre de la salida a `Undefined_Output`.
13. Haga clic en **Finalizar**.
14. En la transformación de analizador, haga clic en el puerto `Undefined_Output` y arrástrelo hasta el puerto `FirstName` del objeto de datos `LA_customers_tgt`.
Aparece una conexión entre los puertos.
15. En la transformación de analizador, haga clic en el puerto `OverflowField` y arrástrelo al puerto `LastName` en el objeto de datos `LA_customers_tgt`.

Aparece una conexión entre los puertos.

16. Haga clic en **Archivo > Guardar** para guardar la asignación.

Tarea 3. Ejecutar un perfil en la transformación de analizador

En esta tarea, se ejecuta un perfil en la transformación de analizador para verificar que se ha configurado la transformación de analizador para analizar correctamente el nombre completo.

1. Seleccione el editor que contenga la asignación `ParserMapping`.
2. Haga clic con el botón derecho en la transformación de analizador y seleccione **Perfil ahora**.
El perfil se ejecuta y se abre en el editor.
3. En el editor, haga clic en la vista **Resultados** para mostrar el resultado de la operación de creación de perfil.
4. Seleccione la columna `Undefined_output` para mostrar información sobre la columna en la sección **Detalles**.
Los valores contenidos en la columna `Undefined_output` aparecen en la sección **Detalles**, junto con estadísticas de frecuencia y porcentaje para cada valor.
5. Visualice los datos y compruebe que solamente aparecen nombres en la columna `Undefined_output`.

Tarea 4. Ejecutar la asignación

En esta tarea, se ejecuta la asignación para crear la salida de asignación.

1. Seleccione el editor que contenga la asignación `ParserMapping`.
2. Haga clic en **Ejecutar > Ejecutar asignación**.
La asignación se ejecuta y escribe la salida en el archivo `LA_Customers_tgt.csv`.

Tarea 5. Ver la salida de la asignación

En esta tarea, se ejecuta el visor de datos para visualizar la salida de asignación.

1. En la vista Explorador de objetos, localice el objeto de datos `LA_Customers_tgt` del objeto de datos del proyecto de la guía de autoaprendizaje y haga doble clic en el objeto de datos.
El objeto de datos se abre en el editor.
2. Haga clic en **Ventana > Mostrar vista > Visor de datos**.
Se abre la vista Visor de datos.
3. En el visor de datos, haga clic en **Ejecutar**.
El visor de datos se ejecuta y muestra los datos.

4. Compruebe que las columnas `FirstName` y `LastName` muestren datos analizados correctamente.

Resumen del análisis de datos

En esta lección se le ha explicado que el análisis de datos identifica los elementos de datos de un campo de entrada y escribe cada uno de ellos en una columna nueva.

Ha aprendido a usar la transformación de analizador para analizar los datos. También ha aprendido que puede crear un perfil para una transformación en una asignación a fin de analizar la salida de esa transformación. Por último, ha aprendido que puede visualizar la salida de una asignación mediante el visor de datos.

Ha creado y configurado el objeto de datos `LA_Customers_tgt` para que contenga la salida analizada. Ha creado una asignación para analizar los datos. En esta asignación, ha configurado una transformación de analizador con un conjunto de tokens para que se analicen los nombres y apellidos de la columna `FullName` del archivo de clientes de Los Ángeles. Ha configurado la asignación para grabar los datos analizados en las columnas `Firstname` y `Lastname` del objeto de datos `LA_Customers_tgt`. También ha ejecutado un perfil para visualizar la salida de la transformación antes de ejecutar la asignación. Por último, ha ejecutado la asignación y ha utilizado el visor de datos para visualizar las nuevas columnas de datos del objeto de datos `LA_Customers_tgt`.

CAPÍTULO 14

Lección 5. Estandarización de datos

Este capítulo incluye los siguientes temas:

- [Resumen de la estandarización de datos, 74](#)
- [Tarea 1. Crear un objeto de datos de destino, 75](#)
- [Tarea 2. Crear una asignación para estandarizar datos, 77](#)
- [Tarea 3. Ejecutar la asignación, 79](#)
- [Tarea 4. Ver la salida de la asignación, 79](#)
- [Resumen de la estandarización de datos, 80](#)

Resumen de la estandarización de datos

La estandarización de datos mejora la calidad de estos al eliminar los errores e inconsistencias que contienen.

Para mejorar la calidad de los datos, estandarice datos que contengan los siguientes tipos de valores:

- Valores incorrectos
- Valores con información correcta, aunque con formato incorrecto.
- Valores de los cuales se desea derivar información nueva.

Utilice la transformación de estandarizador para buscar estos valores en los datos. Puede elegir uno de los siguientes tipos de operación de búsqueda:

- **Texto.** Busque cadenas personalizadas introducidas. Quite estas cadenas o reemplácelas con un texto personalizado.
- **Tabla de referencia.** Busque cadenas que se encuentren en una tabla de referencia seleccionada. Quite estas cadenas o reemplácelas con entradas de una tabla de referencia o un texto personalizado.

Por ejemplo, puede configurar la transformación de estandarizador para estandarizar datos de dirección que contengan las cadenas personalizadas `Street` y `St.` utilizando la cadena de reemplazo `ST.` La transformación de estandarizador sustituye los términos de la búsqueda con el término `ST.` y escribe el resultado en una nueva columna de datos.

Historial

HypoStores necesita estandarizar sus datos de direcciones de clientes para que todas las direcciones utilicen términos uniformes. Los datos de dirección del objeto de datos `All_Customers` contienen entradas con un formato inconsistente para términos comunes como `Street`, `Boulevard`, `Avenue`, `Drive` y `Park`.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

- Crear y configurar un objeto de datos `All_Customers_Stdz_tgt` para que contenga datos estandarizados.
- Crear una asignación para estandarizar los términos de dirección `Street`, `Boulevard`, `Avenue`, `Drive` y `Park` en un formato uniforme.
- Añadir el objeto de datos `All_Customers` a la asignación para conectarlo a los datos de origen.
- Añadir el objeto de datos `All_Customers_Stdz_tgt` a la asignación para crear un objeto de datos de destino.
- Añadir una transformación de estandarizador a la asignación y configurarlo para estandarizar los términos de las direcciones.
- Ejecutar la asignación para generar datos de dirección estandarizados.
- Ejecutar el visor de datos para visualizar la salida de la asignación.

Requisitos previos

Antes de empezar con esta lección, compruebe que cumple los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 2 de esta guía de autoaprendizaje.

Duración

Dedique 15 minutos a completar esta lección.

Tarea 1. Crear un objeto de datos de destino

En esta tarea, se crea un objeto de datos `All_Customers_Stdz_tgt` en el que se pueden escribir datos estandarizados.

Para crear un objeto de datos de destino, lleve a cabo los siguientes pasos:

1. Cree un objeto de datos `All_Customers_Stdz_tgt` basado en el archivo `All_Customers.csv`.
2. Configure las opciones de lectura y escritura para el objeto de datos, incluyendo los nombres y ubicaciones de los archivos.

Paso 1. Crear un objeto de datos `All_Customers_Stdz_tgt`

En este paso, se crea un objeto de datos `All_Customers_Stdz_tgt` basado en el archivo `All_Customers.csv`.

1. Haga clic en **Archivo > Nuevo > Objeto de datos**.
Se abre la ventana **Nuevo**.
2. Seleccione **Objeto de datos de archivo sin formato** y haga clic en **Siguiente**.
3. Compruebe que **Crear a partir de archivo sin formato existente** esté seleccionada.

4. Haga clic en **Explorar** y localice `All_Customers.csv` en el siguiente directorio: <Directorio de instalación de Informatica>\clients\DeveloperClient\Tutorials
 5. Haga clic en **Abrir**.
 6. En el campo `Nombre`, introduzca `All_Customers_Stdz_tgt`.
 7. Haga clic en **Siguiente**.
 8. Haga clic en **Siguiente**.
 9. En la sección **Previsualizar opciones**, seleccione **Importar nombres de columna desde la primera línea** y haga clic en **Siguiente**.
 10. Haga clic en **Finalizar**.
- El objeto de datos `All_Customers_Stdz_tgt` aparece en el editor.

Paso 2. Configurar opciones de lectura y escritura

En este paso, se configuran las opciones de lectura y escritura para el objeto de datos `All_Customers_Stdz_tgt`, incluyendo los nombres y ubicaciones de los archivos.

1. Compruebe que el objeto de datos `All_Customers_Stdz_tgt` esté abierto en el editor.
2. En el editor, seleccione la vista **Lectura**.
3. Haga clic en **Ventana > Mostrar vista > Propiedades**.
4. En la vista **Propiedades**, seleccione la vista **Tiempo de ejecución**.
5. En la columna **Valor**, haga doble clic en el nombre del archivo de origen y escriba `All_Customers_Stdz_tgt.csv`.
6. En la columna **Valor**, haga doble clic en la entrada **Directorio de archivo de origen**.
7. Haga clic con el botón derecho sobre el nombre resaltado y seleccione **Copiar**.
8. En el editor, seleccione la vista **Escritura**.
9. En la vista **Propiedades**, seleccione la vista **Tiempo de ejecución**.
10. En la columna **Valor**, haga doble clic en la entrada **Directorio de archivo de salida**.
11. Haga clic con el botón derecho y seleccione **Pegar** para pegar la ubicación que ha copiado desde la vista **Lectura**.
12. En la columna **Valor**, haga doble clic en la entrada **Opciones de encabezado** y elija `Nombres de campo de salida`.
13. En la columna **Valor**, haga doble clic en el **Nombre de archivo de salida** y escriba `All_Customers_Stdz_tgt.csv`.
14. Haga clic en **Archivo > Guardar** para guardar el objeto de datos.

Tarea 2. Crear una asignación para estandarizar datos

En esta tarea, se crea una asignación y se configura para usar objetos de datos y una transformación del estandarizador.

Para crear una asignación para estandarizar datos, lleve a cabo los siguientes pasos:

1. Crear una asignación.
2. Añada objetos de datos de origen y destino a la asignación.
3. Añada una transformación de estandarizador a la asignación.
4. Configure la transformación del estandarizador para estandarizar términos de direcciones comunes en formatos uniformes.

Paso 1. Crear una asignación

En este paso, se crea y da nombre a la asignación.

1. En la vista **Explorador de objetos**, seleccione su proyecto de guía de autoaprendizaje.
2. Haga clic en **Archivo > Nueva > Asignación**.
Se abre la ventana **Nueva asignación**.
3. En el campo **Nombre**, introduzca `StandardizerMapping`.
4. Haga clic en **Finalizar**.

La asignación se abre en el editor.

Paso 2. Añadir objetos de datos a la asignación

Este paso, se añaden el objeto de datos `All_Customers` y el objeto de datos `All_Customers_Stdz_tgt` a la asignación.

1. En la vista **Explorador de objetos**, localice los objetos de datos del proyecto de autoaprendizaje.
2. Seleccione el objeto de datos `All_Customers` y arrástrelo hasta el editor.
Se abre la ventana **Añadir objeto de datos físicos a asignación**.
3. Compruebe que **Lectura** esté seleccionado y haga clic en **Aceptar**.
El objeto de datos aparece en el editor.
4. En la vista **Explorador de objetos**, localice los objetos de datos del proyecto de autoaprendizaje.
5. Seleccione el objeto de datos `All_Customers_Stdz_tgt` y arrástrelo hasta el editor.
Se abre la ventana **Añadir objeto de datos físicos a asignación**.
6. Seleccione **Escritura** y haga clic en **Aceptar**.
El objeto de datos aparece en el editor.
7. Seleccione todos los puertos del objeto de datos `All_Customers`. Arrastre los puertos hasta el puerto `CustomerID` del objeto de datos `All_Customers_Stdz_tgt`.

Sugerencia: Mantenga pulsada la tecla Mayús para seleccionar varios puertos. Quizás tenga que desplazarse hacia abajo en la lista de puertos para seleccionar todos los puertos.

Los puertos del objeto de datos `All_Customers` se conectan con los puertos correspondientes en el objeto de datos `All_Customers_Stdz_tgt`.

Paso 3. Añadir una transformación de estandarizador a la asignación

En este paso, se añade una transformación de estandarizador para estandarizar cadenas en los datos de una dirección.

1. Seleccione el editor que contenga la asignación `StandardizerMapping`.
2. En la paleta Transformación, seleccione la transformación de estandarizador.
3. Haga clic en el editor.
Una transformación de estandarizador llamada `NewStandardizer` aparece en la asignación.
4. Para cambiar el nombre de la transformación de estandarizador, haga doble clic en la barra de título de la transformación y escriba `AddressStandardizer`.
5. Seleccione el puerto `Address1` en el objeto de datos `All_Customers` y arrastre el puerto al grupo **Input** de la transformación de estandarizador.

Un puerto llamado `Address1` aparece en el grupo de entrada. El puerto se conecta con el puerto `Address1` del objeto de datos `All_Customers`.

Nota: Se añade un puerto de salida a la transformación cuando se configura una estrategia de estandarización.

Paso 4. Configurar la transformación de estandarizador

En este paso, se configura la transformación de estandarizador para estandarizar los términos de dirección en los datos de origen.

Nota: En esta tarea, definirá cinco operaciones de estandarización. Cada operación reemplaza una cadena de la columna de entrada con una cadena nueva.

1. Seleccione el editor que contenga la asignación `StandardizerMapping`.
2. Haga clic en la transformación de estandarizador.
3. Haga clic en **Ventana > Mostrar vista > Propiedades**.
4. En la vista Propiedades, seleccione **Estrategias**.
5. Haga clic en **Nuevo**. Se muestra el asistente de nueva estrategia.
6. Haga clic en la flecha de selección en la columna Inputs y elija el puerto de entrada `Address1`.
El campo Salidas muestra `Address1` como puerto de salida.
7. Seleccione los delimitadores de espacio de carácter y de coma `[\s]` y `[,]`. Si lo desea, seleccione las opciones para eliminar los espacios finales.
8. Haga clic en **Siguiente**.
9. Seleccione la operación **Reemplazar cadenas personalizadas** y haga clic en **Siguiente**.
10. En **Propiedades**, haga clic en **Nueva**.
11. Edite los campos **Cadenas personalizadas** y **Reemplazar con** de modo que contengan el primer par de cadenas de la siguiente tabla:

Cadenas personalizadas	Reemplazar por
STREET	ST.

Cadenas personalizadas	Reemplazar por
BOULEVARD	BLVD.
AVENUE	AVE.
DRIVE	DR.
PARK	PK.

- Repita los pasos 9 a 12 para definir las operaciones de estandarización para todas las cadenas de la tabla.
- Arrastre el puerto de salida `Address1` al puerto `Address1` en el objeto de datos `All_Customers_Stdz_tgt`.
- Haga clic en **Archivo > Guardar** para guardar la asignación.

Tarea 3. Ejecutar la asignación

En esta tarea, se ejecuta la asignación para escribir direcciones estandarizadas en el objeto de datos de salida.

- Seleccione el editor que contenga la asignación `StandardizerMapping`.
- Haga clic en **Ejecutar > Ejecutar asignación**.

La asignación se ejecuta y escribe la salida en el archivo `All_Customers_Stdz_tgt.csv`.

Tarea 4. Ver la salida de la asignación

En esta tarea, ejecutará el visor de datos para visualizar la salida de asignación y comprobar que los datos de dirección se hayan estandarizado correctamente.

- En la vista **Explorador de objetos**, localice el objeto de datos `All_Customers_Stdz_tgt` del proyecto de guía de autoaprendizaje y haga doble clic en el objeto de datos.
El objeto de datos se abre en el editor.
- Haga clic en **Ventana > Mostrar vista > Visor de datos**.
Se abre la vista Visor de datos.
- En el visor de datos, haga clic en **Ejecutar**.
El visor de datos muestra la salida de asignación.
- Compruebe que la columna `Address1` muestre datos estandarizados correctamente. Por ejemplo, todas las instancias de la cadena `STREET` deben haberse reemplazado por la cadena `ST`.

Resumen de la estandarización de datos

En esta lección, ha aprendido que puede estandarizar datos para eliminar errores e inconsistencias en los datos.

Ha aprendido que puede utilizar una transformación de estandarizador para estandarizar cadenas en una columna de entrada. También ha aprendido que puede visualizar la salida de una asignación mediante el visor de datos.

Ha creado y configurado el objeto de datos `All_Customers_Stdz_tgt` para que contenga una salida estandarizada. Ha creado una asignación para estandarizar los datos. En esta asignación, ha configurado una transformación de estandarizador para estandarizar la columna `Address1` del objeto de datos `All_Customers`. Ha configurado la asignación para escribir una salida estandarizada en el objeto de datos `All_Customers_Stdz_tgt`. Por último, ha ejecutado la asignación y ha utilizado el visor de datos para visualizar los datos estandarizados del objeto de datos `All_Customers_Stdz_tgt`.

CAPÍTULO 15

Lección 6. Validación de datos de dirección

Este capítulo incluye los siguientes temas:

- [Resumen de la validación de datos de direcciones, 81](#)
- [Tarea 1. Crear un objeto de datos de destino , 82](#)
- [Tarea 2. Crear una asignación para validar direcciones, 84](#)
- [Tarea 3. Configurar la transformación de validación de dirección, 85](#)
- [Tarea 4. Ejecutar la asignación, 89](#)
- [Tarea 5. Ver la salida de la asignación, 89](#)
- [Resumen de la validación de datos de direcciones, 92](#)

Resumen de la validación de datos de direcciones

La validación de direcciones es el proceso que consiste en evaluar y mejorar la calidad de las direcciones postales. Evalúa la calidad de las direcciones comparando direcciones de entrada con un conjunto de datos de referencia o direcciones válidas. Mejora la calidad de las direcciones mediante la identificación de valores de dirección incorrectos y utilizando el conjunto de datos de referencia para crear campos que contengan valores correctos.

Una dirección es válida cuando se puede entregar. Una dirección tiene un formato correcto y contiene información real sobre la calle, la ciudad y el código postal; si los datos no dan como resultado una dirección con factibilidad de entrega, la dirección no es válida. La herramienta Developer utiliza conjuntos de datos de referencia de direcciones para comprobar la factibilidad de entrega de las direcciones de entrada. Informatica proporciona conjuntos de datos de referencia de direcciones.

Un conjunto de datos de referencia de direcciones contiene datos que describen todas las direcciones con factibilidad de entrega de un país. El proceso de validación de direcciones busca en el conjunto de datos de referencia la dirección que se parece más a los datos de la dirección de entrada. Si el proceso encuentra una coincidencia cercana en el conjunto de datos de referencia, escribe valores nuevos para cualquier valor de datos incorrecto o incompleto. El proceso crea un conjunto de códigos alfanuméricos que describen el tipo de coincidencia detectada entre la dirección de entrada y las direcciones de referencia. También puede reestructurar la dirección y puede añadir información que no se encuentre en la dirección de entrada, como un sufijo de código postal de cuatro dígitos para una dirección de Estados Unidos.

Utilice la transformación del validador de direcciones para construir procesos de validación de direcciones en la herramienta Developer. Esta transformación multigrupo contiene un conjunto de puertos de entrada y

salida predefinidos que corresponden a todos los campos posibles de una dirección de entrada. Cuando se configura una transformación del validador de direcciones, se define el conjunto de datos de referencia y se crea una plantilla de dirección de entrada y de salida mediante los puertos de transformación. En esta lección, se configura la transformación para validar datos de direcciones de Estados Unidos.

Historial

HypoStores necesita tener datos de direcciones correctos y completos para asegurarse de que sus campañas de envío directo de correspondencia y otros elementos de correspondencia lleguen a sus clientes. Disponer de datos de dirección correctos y completos también reduce el coste de las operaciones de envío de correspondencia para la organización. Asimismo, HypoStores necesita que los datos de sus clientes incluyan direcciones con un formato de impresión lo suficientemente flexible para incluir direcciones de longitudes distintas.

Para cumplir estos requisitos empresariales, el equipo ICC de HypoStores crea una asignación de validación de direcciones en la herramienta Developer.

Objetivos

En esta lección, llevará a cabo las tareas siguientes:

- Crear un objeto de datos de destino que contenga los campos de dirección y códigos coincidentes validados.
- Crear una asignación con un objeto de datos de origen, un objeto de datos destino y una transformación del validador de direcciones.
- Configurar la transformación del validador de direcciones para validar los datos de las direcciones de los clientes.
- Ejecutar la asignación para validar los datos de las direcciones y revisar las salidas de los códigos coincidentes para verificar la validez de los datos de las direcciones.

Requisitos previos

Antes de empezar con esta lección, compruebe los siguientes requisitos previos:

- Ha terminado las Lecciones 1 y 2 de esta guía de autoaprendizaje.
- Los datos de referencia de direcciones de Estados Unidos están instalados en el dominio y registrados en la herramienta Administrator. Póngase en contacto con el administrador de Informática para comprobar que los datos de direcciones de Estados Unidos estén instalados en el sistema. Los datos de referencia se instalan mediante el programa de instalación de contenido Data Quality.

Duración

Dedique 25 minutos a completar esta lección.

Tarea 1. Crear un objeto de datos de destino

En esta tarea, se crea un objeto de datos de destino, se configuran las opciones de escritura y se añaden puertos.

Para crear y configurar el objeto de datos de destino, lleve a cabo los siguientes pasos:

1. Cree un objeto de datos All_Customers_av_tgt basado en el archivo All_Customers.csv.
2. Configure las opciones de lectura y escritura para el objeto de datos, incluyendo los nombres y ubicaciones de los archivos.

3. Añada puertos al objeto de datos para recibir los valores de código correlacionados generados por la transformación del validador de direcciones.

Paso 1. Crear el objeto de datos All_Customers_av_tgt

En este paso, se crea un objeto de datos `All_Customers_av_tgt` basado en el archivo `All_Customers.csv`.

1. Haga clic en **Archivo > Nuevo > Objeto de datos**.
Se abre la ventana **Nuevo**.
2. Seleccione **Objeto de datos de archivo sin formato** y haga clic en **Siguiente**.
3. Compruebe que **Crear a partir de archivo sin formato existente** está seleccionada. Haga clic en **Explorar** junto a esta selección, busque el archivo `All_Customers.csv` y haga clic en **Abrir**.
4. En el campo **Nombre**, introduzca `All_Customers_av_tgt`.
5. Haga clic en **Siguiente**.
6. Haga clic en **Siguiente**.
7. En la sección **Previsualizar opciones**, seleccione **Importar nombres de columna de la primera línea** y haga clic en **Siguiente**.
8. Haga clic en **Finalizar**.

El objeto de datos `All_Customers_av_tgt` aparece en el editor.

Paso 2. Configurar opciones de lectura y escritura

En este paso, se configuran las opciones de lectura y escritura para el objeto de datos `All_Customers_av_tgt`, incluyendo el nombre y la ubicación del archivo de destino.

1. Compruebe que el objeto de datos `All_Customers_av_tgt` esté abierto en el editor.
2. En el editor, seleccione la vista **Lectura**.
3. Seleccione **Ventana > Mostrar vista > Propiedades**.
4. En la vista **Propiedades**, seleccione la vista **Tiempo de ejecución**.
5. En la columna **Valor**, haga doble clic en el nombre del archivo de origen y escriba `All_Customers_av_tgt.csv`.
6. En la columna **Valor**, haga doble clic para resaltar la ruta del directorio del archivo de origen.
7. Haga clic con el botón derecho sobre la ruta resaltada y seleccione **Copiar**.
8. En el editor, seleccione la vista **Escritura**.
9. En la vista **Propiedades**, seleccione la vista **Tiempo de ejecución**.
10. En la columna **Valor**, haga doble clic en la entrada **Directorio de archivo de salida**.
11. Haga clic con el botón derecho en esta entrada y seleccione **Pegar** para añadir la ruta que ha copiado desde la vista **Lectura**.
12. En la columna **Valor**, haga doble clic en la entrada **Opciones de encabezado** y elija **Nombres de campo de salida**.
13. En la columna **Valor**, haga doble clic en el **Nombre de archivo de salida** y escriba `All_Customers_av_tgt.csv`.
14. Seleccione **Archivo > Guardar** para guardar el objeto de datos.

Paso 3. Añadir puertos al objeto de datos

En este paso, se añaden dos puertos al objeto de datos `All_Customers_av_tgt` para que la transformación del validador de direcciones pueda escribir valores de códigos correlacionados en el archivo de destino. Nombre los puertos como `MailabilityScore` y `MatchCode`.

El valor `MailabilityScore` describe la capacidad de entrega de la dirección de entrada. El valor `MatchCode` describe el tipo de correlación que realiza la transformación entre la dirección de entrada y las direcciones de los datos de referencia.

1. En la vista **Explorador de objetos**, localice los objetos de datos del proyecto de autoaprendizaje.
2. Haga doble clic en el objeto de datos `All_Customers_av_tgt`.
El objeto de datos `All_Customers_av_tgt` se abre en el editor.
3. Compruebe que **Resumen** esté seleccionada.
4. Seleccione el puerto final en la lista de puertos. Este puerto se llama `MiscDate`.
5. Haga clic en **Nuevo**.
Aparece un puerto llamado `MiscDate1`.
6. Cambie el nombre del puerto `MiscDate1` a `MailabilityScore`.
7. Seleccione el puerto `MailabilityScore`.
8. Haga clic en **Nuevo**.
Aparece un puerto llamado `MailabilityScore1`.
9. Cambie el nombre del puerto `MailabilityScore1` a `MatchCode`.
10. Haga clic en **Archivo > Guardar** para guardar el objeto de datos.

Tarea 2. Crear una asignación para validar direcciones

En esta tarea, se crea una asignación y se añaden objetos de datos y una transformación del validador de direcciones.

Para crear la asignación y añadir los objetos que necesita, lleve a cabo los siguientes pasos:

1. Cree un objeto de asignación.
2. Añada objetos de datos de origen y de destino a la asignación.
3. Añada una transformación de validador de direcciones a la asignación.

Paso 1. Crear una asignación

En este paso, se crea y da nombre a la asignación.

1. En la vista **Explorador de objetos**, seleccione su proyecto de guía de autoaprendizaje.
2. Seleccione **Archivo > Nueva > Asignación**.
Se abre la ventana **Nueva asignación**.
3. En el campo **Nombre**, introduzca `ValidateAddresses`.

4. Haga clic en **Finalizar**.
La asignación se abre en el editor.

Paso 2. Añadir objetos de datos a la asignación

Este paso, se añaden los objetos de datos de origen y destino a la asignación.

`All_Customers` es el objeto de datos de origen para la asignación. La transformación del validador de direcciones lee los datos de este objeto. `All_Customers_av_tgt` es el objeto de datos de destino para la asignación. Este objeto lee datos de la transformación del validador de direcciones.

1. En la vista **Explorador de objetos**, localice los objetos de datos del proyecto de autoaprendizaje.
2. Seleccione el objeto de datos `All_Customers` y arrástrelo hasta el editor.
Se abre la ventana **Añadir objeto de datos físicos a asignación**.
3. Compruebe que **Lectura** esté seleccionado y haga clic en **Aceptar**.
El objeto de datos aparece en el editor.
4. En la vista **Explorador de objetos**, localice los objetos de datos del proyecto de autoaprendizaje.
5. Seleccione el objeto de datos `All_Customers_av_tgt` y arrástrelo hasta el editor.
Se abre la ventana **Añadir objeto de datos físicos a asignación**.
6. Seleccione **Escritura** y haga clic en **Aceptar**.
El objeto de datos aparece en el editor.
7. Haga clic en **Guardar**.

Paso 3. Añadir una transformación de validador de direcciones a la asignación

En este paso, se añade una transformación del validador de direcciones a la asignación que contiene los objetos de origen y destino.

Una vez realizado este paso, se puede configurar la transformación y conectar sus puertos con los objetos de datos.

1. Seleccione el editor que contenga la asignación `ValidateAddresses`.
2. En la paleta Transformación, seleccione la transformación de validador de direcciones.
3. Haga clic en el editor.
La transformación del validador de direcciones aparece en el editor.

Tarea 3. Configurar la transformación de validación de dirección

En esta tarea, se configura la transformación del validador de direcciones para leer y validar direcciones del origen de datos `All_Customers`.

Nota: La transformación del validador de direcciones contiene una serie de puertos de entrada y salida predefinidos. Seleccione los puertos que necesite y conéctelos a los objetos de la asignación.

Para configurar la transformación, lleve a cabo los siguientes pasos:

1. Seleccione el país predeterminado para la validación de dirección.
2. Configure los puertos de entrada de la transformación.
3. Configure los puertos de salida de la transformación.
4. Conecte puertos de origen de datos no utilizados con el destino de datos.

Paso 1. Establecer el país predeterminado para la validación de dirección

En este paso, seleccione el país predeterminado para la validación de dirección. Los archivos de datos de referencia de dirección que la transformación del validador de dirección utiliza están organizados por país. Al seleccionar el país predeterminado, debe identificar el conjunto de datos de país que la transformación aplicará a cualquier dirección de entrada que no contenga información de país.

1. Seleccione la transformación del validador de direcciones en el editor.
2. En **Propiedades**, haga clic en **Configuración general**.
3. En el menú **País predeterminado**, seleccione `Estados Unidos`.

Paso 2. Configurar los puertos de entrada de la transformación del validador de direcciones

En este paso, se seleccionan puertos de entrada de transformación y se conectan al objeto de datos `All_Customers_av`.

La transformación del validador de direcciones contiene varios grupos de puertos de entrada predefinidos. Seleccione los puertos de entrada correspondientes a los campos de la dirección de entrada y añádalos a la transformación.

Mantenga pulsada la tecla Ctrl en los siguientes pasos para seleccionar múltiples puertos en una única operación.

1. Seleccione la transformación del validador de direcciones en el editor.
2. Bajo **Propiedades**, haga clic en **Plantillas**.
3. Expanda el grupo de puertos **Modelo básico**.
4. Expanda el grupo de puertos **Híbrido** y seleccione los siguientes puertos:

Nombre de puerto	Descripción
Línea de dirección de entrega 1	Datos de la dirección, como el nombre de la calle y el número del edificio.
Localidad completa 1	Nombre de la ciudad o localidad.
Código postal 1	Código postal.
Provincia 1	Nombre de la provincia o estado.
Nombre de país	Nombre o abreviatura del país.

Nota: Mantenga pulsada la tecla Ctrl para seleccionar múltiples puertos en una única operación.

5. En la barra de herramientas situada encima de la lista de nombres de puerto, haga clic en **Añadir puerto a transformación**.

Esta barra de herramientas está visible cuando se selecciona **Plantillas**.

Los puertos seleccionados aparecen en la transformación en el editor de asignaciones.

6. Conecte los puertos origen con los puertos de la transformación del validador de direcciones como se indica a continuación:

Puerto de origen	Puerto de transformación del validador de direcciones
Address1	Línea de dirección de entrega 1
CITY	Localidad completa 1
Código postal	Código postal 1
Estado	Provincia 1
País	Nombre de país

Paso 3. Configurar los puertos de salida de la transformación del validador de direcciones

En este paso, selecciona los puertos de salida de transformación y conecta estos puertos al objeto de datos `All_Customers_av_tgt`.

La transformación del validador de direcciones contiene varios grupos de puertos de salida predefinidos. Seleccione los puertos que definan la estructura de dirección que necesite y añádalos a la transformación.

También puede seleccionar puertos que contengan información sobre el tipo de validación lograda para cada dirección.

1. Seleccione la transformación del validador de direcciones en el editor de asignaciones.
2. En **Propiedades**, haga clic en **Plantillas**.
3. Expanda el grupo de puertos **Modelo básico**.
4. Expanda el grupo de puertos de salida **Elementos de dirección** y seleccione el siguiente puerto:

Nombre de puerto	Descripción
Calle completa 1	Datos de dirección de una calle, como el nombre de la calle y el número del edificio.

5. Expanda el grupo de puertos de salida **Elementos de última línea** y seleccione los siguientes puertos:

Nombre de puerto	Descripción
Localidad completa 1	Nombre de la ciudad o localidad.
Código postal 1	Código postal.
Abreviatura de provincia 1	Identificador de la provincia o estado.

Nota: Mantenga pulsada la tecla Ctrl para seleccionar múltiples puertos en una única operación.

6. Expanda el grupo de puertos de salida **País** y seleccione el siguiente puerto:

Nombre de puerto	Descripción
Nombre del país 1	Nombre del país.

7. Expanda el grupo de puertos de salida **Información de estado** y seleccione los siguientes puertos:

Nombre de puerto	Descripción
Puntuación por correo	Puntuación que representa la posibilidad de una entrega postal satisfactoria.
Código de coincidencia	Código que representa el grado de similitud entre la dirección de entrada y los datos de referencia.

8. En la barra de herramientas situada encima de la lista de nombres de puerto, haga clic en **Añadir puerto a transformación**.

Esta barra de herramientas está visible cuando se selecciona **Plantillas**.

9. Conecte los puertos de la transformación del validador de direcciones con los puertos `All_Customers_av_tgt` del modo que se indica a continuación:

Puerto de transformación del validador de direcciones	Puerto de destino
Calle completa 1	Address1
Localidad completa 1	CITY
Código postal 1	ZIP
Abreviatura de provincia 1	Estado
Nombre del país 1	País
Puntuación por correo	MailabilityScore
Código de coincidencia	MatchCode

Paso 4. Conectar puertos de origen de datos no utilizados con el destino de datos

En este paso, se conectan los puertos no utilizados en el origen de datos `All_Customers` con el destino de datos.

- Conecte los puertos no utilizados en el origen de datos con los puertos que tengan los mismos nombres en el destino de datos.

Tarea 4. Ejecutar la asignación

En esta tarea, se ejecuta la asignación para crear la salida de asignación.

1. Seleccione el editor que contenga la asignación `ValidateAddresses`.
2. Seleccione **Ejecutar > Ejecutar asignación**.

La asignación se ejecuta y escribe la salida en el archivo `All_Customers_av_tgt.csv`.

Tarea 5. Ver la salida de la asignación

En esta tarea, se ejecuta el visor de datos para visualizar la salida de una asignación. Compruebe la calidad de las direcciones validadas examinando los valores escritos en las columnas Puntuación de viabilidad de envío de correo y Código de coincidencia del objeto de datos de destino.

El valor Código de coincidencia es un código alfanumérico que representa el tipo de validación realizada por la asignación en la dirección.

El valor Puntuación de viabilidad de envío de correo es un valor de un dígito que indica la factibilidad de entrega de la dirección.

1. En la vista Explorador de objetos, busque el objeto de datos `All_Customers_av_tgt` del proyecto de guía de autoaprendizaje y haga doble clic en el objeto de datos.

El objeto de datos se abre en el editor.

2. Seleccione **Ventana > Mostrar vista > Visor de datos**.

Se abre el visor de datos.

3. Haga clic en **Ejecutar** en el visor de datos.

El visor de datos muestra la salida de la asignación.

4. Desplácese por los resultados de la asignación hasta que las columnas Puntuación de viabilidad de envío de correo y Código de coincidencia sean visibles.

5. Revise los valores de la columna Puntuación de viabilidad de envío de correo.

Las puntuaciones pueden ir de 0 a 5. Las direcciones con puntuaciones superiores tienen más probabilidad entregarse correctamente.

6. Revise los valores de la columna Código de coincidencia.

El código de coincidencia es un código alfanumérico. El carácter alfabético indica el tipo de validación realizada por la transformación, mientras que el dígito indica la calidad de la dirección definitiva.

En la tabla siguiente se describen los valores del código de coincidencia:

Código	Descripción
A1	La búsqueda de códigos de dirección ha encontrado una dirección parcial o completa para el código de entrada.
A0	La búsqueda de códigos de dirección no ha encontrado ninguna dirección para el código de entrada.

Código	Descripción
C4	Corregido. Todos los elementos relevantes para el envío postal se han comprobado.
C3	Corregido. Algunos elementos no se pueden comprobar.
C2	Corregidos, aunque no se ha podido determinar el estado de entrega debido a la ausencia de datos de referencia.
C1	Corregidos, aunque no se ha podido determinar el estado de entrega debido a que la estandarización del usuario ha introducido errores.
I4	Los datos no se pueden corregir completamente, pero hay una única coincidencia con una dirección en los datos de referencia.
I3	Los datos no se pueden corregir completamente, pero hay varias coincidencias con direcciones en los datos de referencia.
I2	Los datos no se pueden corregir. El modo por lotes devuelve direcciones sugeridas parciales.
I1	Los datos no se pueden corregir. El modo por lotes no se puede sugerir una dirección.
N7	Error de validación. La validación no se produjo porque la validación de una sola línea no está desbloqueada.
N6	Error de validación. La validación no se produjo porque la validación de una sola línea no es compatible con el país de destino.
N5	Error de validación. La validación no se produjo porque la base de datos de referencia ha caducado.
N4	Error de validación. La validación no se produjo porque los datos de referencia están dañados o tienen un formato incorrecto.
N3	Error de validación. La validación no se produjo porque los datos de país no se pueden desbloquear.
N2	Error de validación. La validación no se produjo porque la base de datos de referencia requerida no está disponible.
N1	Error de validación. La validación no se produjo porque el país no se reconoce o no es compatible.
Q3	Modo de lista de sugerencias. La validación de direcciones puede recuperar una o más direcciones completas a partir de los datos de referencia de direcciones que se corresponden con la dirección de entrada.
Q2	Modo de lista de sugerencias. La validación de direcciones puede combinar los elementos de la dirección de entrada y los elementos de los datos de referencia de direcciones para crear una dirección completa.

Código	Descripción
Q1	Modo de lista de sugerencias. La validación de direcciones no puede sugerir una dirección completa. Para generar una sugerencia de dirección completa, añada datos a la dirección de entrada.
Q0	Modo de lista de sugerencias. No hay suficientes datos de entrada para generar una sugerencia.
RB	País reconocido por la abreviatura. Reconoce códigos de país ISO de dos y de tres caracteres. También puede reconocer abreviaturas habituales como "GER" para Alemania.
RA	País reconocido por el ajuste Forzar país de la transformación.
R9	País reconocido por el ajuste País predeterminado de la transformación.
R8	País reconocido por el nombre de país.
R7	País reconocido por el nombre de país, pero la transformación identifica errores en los datos de país.
R6	País reconocido por los datos de territorio.
R5	País reconocido por los datos de provincia.
R4	País reconocido por los datos de municipio principal.
R3	País reconocido por el formato de dirección.
R2	País reconocido por un script.
R1	País no reconocido porque hay varias coincidencias disponibles.
R0	País no reconocido.
S4	Modo de análisis. La dirección se ha analizado correctamente.
S3	Modo de análisis. La dirección se ha analizado y se han obtenido varios resultados.
S1	Modo de análisis. Se ha producido un error de análisis debido a que un formato de entrada no coincide.
V4	Verificado. Los datos de entrada son correctos. La validación de direcciones ha comprobado todos los elementos relevantes para el envío postal y las entradas coinciden de manera perfecta.
V3	Verificado. Los datos de entrada son correctos, pero algunos o todos los elementos se han estandarizado, o bien la entrada contiene nombres que han caducado o que son exónimos.

Código	Descripción
V2	Verificado. Los datos de entrada son correctos, pero algunos elementos no se han podido verificar debido a que los datos de referencia están incompletos.
V1	Verificado. Los datos de entrada son correctos, pero la estandarización del usuario ha afectado la factibilidad de entrega de forma negativa. Por ejemplo, el código postal tiene una longitud demasiado corta.

Resumen de la validación de datos de direcciones

En esta lección, ha aprendido que la validación de direcciones compara los datos de direcciones de entrada con datos de referencia y devuelve la versión más precisa posible de la dirección.

Ha aprendido que el proceso de validación de direcciones también devuelve información de estado sobre la calidad de cada dirección.

Ha aprendido que los usuarios de la herramienta Administrator ejecutan el programa de instalación de contenido de calidad de datos para instalar datos de referencia de direcciones.

También ha aprendido que la transformación del validador de direcciones es una transformación de varios grupos, y que puede seleccionar de los grupos de puertos los puertos de entrada y de salida para la transformación. Los puertos de entrada que seleccione determinan el contenido de la dirección que se valida. Los puertos de salida determinan el contenido del registro de la dirección final.

APÉNDICE A

Preguntas frecuentes

Este apéndice incluye los siguientes temas:

- [Preguntas frecuentes de Informatica Analyst, 93](#)
- [Preguntas frecuentes de Informatica Developer, 93](#)

Preguntas frecuentes de Informatica Analyst

Revise las preguntas frecuentes para encontrar respuestas a las preguntas acerca de Informatica Analyst.

¿Se puede usar una sola cuenta de usuario para acceder a la Herramienta del administrador, del desarrollador y del analista?

Sí. Se puede conceder permiso a un usuario para acceder a las tres herramientas. No es necesario crear cuentas de usuario independientes para cada aplicación cliente.

¿Dónde se almacenan mis datos de referencia?

Puede utilizar la Herramienta del desarrollador y la Herramienta del analista para crear y compartir objetos de datos de referencia. El repositorio de modelos almacena los metadatos del objeto de datos de referencia. La base de datos de datos de referencia almacena los valores de datos de la tabla de referencia. Configure la base de datos de datos de referencia en el servicio de administración de contenido.

Preguntas frecuentes de Informatica Developer

Revise las preguntas frecuentes para encontrar respuestas a las preguntas acerca de Informatica Developer.

¿Qué diferencia hay entre un origen y un destino en PowerCenter y un objeto de datos físicos en la herramienta Developer?

En PowerCenter, se crea una definición de origen para incluirla como origen de asignación. Se crea una definición de destino para incluirla como destino de asignación. En la herramienta Developer, se crea un objeto de datos físicos que se puede usar como origen o destino de asignación.

¿Qué diferencia hay entre una asignación en la herramienta Developer y una asignación en PowerCenter?

Una asignación en PowerCenter especifica cómo mover datos entre orígenes y destinos. Una asignación en la herramienta Developer especifica cómo se mueven los datos entre la entrada y la salida de una asignación.

Una asignación en PowerCenter debe incluir una o más definiciones de origen, calificadores de origen y definiciones de destino. Una asignación en PowerCenter también puede incluir métodos abreviados, transformaciones y mapplets.

Una asignación en la herramienta Developer debe incluir una entrada y una salida de asignación. Una asignación de la herramienta Developer también puede incluir transformaciones y mapplets.

La herramienta Developer tiene los siguientes tipos de asignaciones:

- Asignación que mueve datos entre orígenes y destinos. Este tipo de asignación solamente difiere de una asignación de PowerCenter en que no puede usar métodos abreviados y no utiliza un calificador de origen.
- Asignación de objetos de datos lógicos Una asignación en un modelo de objeto de datos lógicos. Una asignación de objeto de datos lógicos puede contener un objeto de datos lógicos como entrada de asignación y un objeto de datos como salida de asignación. También puede contener uno o más objetos de datos físicos como entrada de asignación y un objeto de datos lógicos como salida de asignación.
- Asignación de tabla virtual Una asignación en un servicio de datos SQL. Contiene un objeto de datos como entrada de asignación y una tabla virtual como salida de asignación.
- Asignación de procedimiento virtual almacenado. Define un conjunto de lógica empresarial en un servicio de datos SQL. Contiene una transformación de parámetro de entrada u objeto de datos físicos como entrada de asignación y una transformación de parámetro de salida u objeto de datos físicos como salida de asignación.

¿Qué diferencia hay entre un mapplet en PowerCenter y un mapplet en la herramienta Developer?

Un mapplet en PowerCenter y en la herramienta Developer es un objeto reutilizable que contiene un conjunto de transformaciones. Se puede usar la lógica de transformación en múltiples asignaciones.

Un mapplet PowerCenter puede contener definiciones de origen o transformaciones de entrada como entrada de asignación. Debe contener transformaciones de salida como salida del mapplet.

Un mapplet de la herramienta Developer puede contener objetos de datos o transformaciones de entrada como salida del mapplet. Puede contener objetos de datos o transformaciones de salida como salida del mapplet. Una asignación en la herramienta Developer también incluye las siguientes funciones:

- Se puede validar un mapplet como regla. Se utiliza una regla en un perfil.
- Un mapplet puede contener otros mapplets.

¿Qué diferencia hay entre un mapplet y una regla?

Se puede validar un mapplet como regla. Una regla es una lógica empresarial que define las condiciones que se aplican a los datos de origen al ejecutarse un perfil. Se puede validar un mapplet como regla cuando el mapplet cumple los siguientes requisitos:

- Contiene una transformación de entrada y salida.
- El mapplet no contiene ninguna transformación activa.
- No especifica cardinalidad entre grupos de entrada.

INDICE

C

configuración de Developer tool
resumen [49](#)
configuración de la Herramienta del analista
resumen [20](#)
creación de cuadros de mandos
resumen [37](#)
creación de objetos de datos
resumen [23](#)
creación de perfiles de datos
resumen [62](#)
creación de perfiles personalizados
resumen [30](#)
creación de perfiles predeterminados
resumen [26](#)

creación de reglas de expresión
resumen [34](#)
creación de tablas de referencia a partir de columnas
resumen [43](#)

I

importar objeto de datos físicos
resumen [53](#)

T

tablas de referencia
resumen [46](#)