



Informatica® Data Quality
9.6.1 HotFix 2

Guide de démarrage

© Copyright Informatica LLC 1998, 2018

Ce logiciel et sa documentation contiennent des informations appartenant à Informatica Corporation, protégées par la loi sur le droit d'auteur et fournies dans le cadre d'un accord de licence prévoyant des restrictions d'utilisation et de divulgation. Toute ingénierie inverse du logiciel est interdite. Il est interdit de reproduire ou transmettre sous quelque forme et par quelque moyen que ce soit (électronique, photocopie, enregistrement ou autre) tout ou partie de ce document sans le consentement préalable d'Informatica Corporation. Ce logiciel peut être protégé par des brevets américains et/ou internationaux, ainsi que par d'autres brevets en attente.

L'utilisation, la duplication ou la divulgation du Logiciel par le gouvernement américain est sujette aux restrictions décrites dans l'accord de licence applicable du logiciel conformément aux documents DFARS 227.7202-1(a) et 227.7702-3(a) (1995), DFARS 252.227-7013[©](1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19 ou FAR 52.227-14 (ALT III) le cas échéant.

Les informations dans ce produit ou cette documentation sont sujettes à modification sans préavis. Si vous rencontrez des problèmes dans ce produit ou la documentation, veuillez nous en informer par écrit.

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 et Informatica Master Data Management sont des marques de commerce ou des marques déposées d'Informatica Corporation aux États-Unis et dans d'autres juridictions du monde. Tous les autres noms de société ou de produit peuvent être des marques de commerce ou des marques déposées de leurs détenteurs respectifs.

Des portions de ce logiciel et/ou de la documentation sont sujettes au copyright détenu par des tierces parties, dont Copyright DataDirect Technologies. Tous droits réservés. Copyright © Sun Microsystems. Tous droits réservés. Copyright © RSA Security Inc. Tous droits réservés. Copyright © Ordinal Technology Corp. Tous droits réservés. Copyright © Aandacht c.v. Tous droits réservés. Copyright Genivia, Inc. Tous droits réservés. Copyright Isomorphic Software. Tous droits réservés. Copyright © Meta Integration Technology, Inc. Tous droits réservés. Copyright © Intalio. Tous droits réservés. Copyright © Oracle. Tous droits réservés. Copyright © Adobe Systems Incorporated. Tous droits réservés. Copyright © DataArt, Inc. Tous droits réservés. Copyright © ComponentSource. Tous droits réservés. Copyright © Microsoft Corporation. Tous droits réservés. Copyright © Rogue Wave Software, Inc. Tous droits réservés. Copyright © Teradata Corporation. Tous droits réservés. Copyright © Yahoo! Inc. Tous droits réservés. Copyright © Glyph & Cog, LLC. Tous droits réservés. Copyright © Thinkmap, Inc. Tous droits réservés. Copyright © Clearpace Software Limited. Tous droits réservés. Copyright © Information Builders, Inc. Tous droits réservés. Copyright © OSS Nokalva, Inc. Tous droits réservés. Copyright Edifecs, Inc. Tous droits réservés. Copyright Cleo Communications, Inc. Tous droits réservés. Copyright © International Organization for Standardization 1986. Tous droits réservés. Copyright © ej-technologies GmbH. Tous droits réservés. Copyright © Jaspersoft Corporation. Tous droits réservés. Copyright © International Business Machines Corporation. Tous droits réservés. Copyright © yWorks GmbH. Tous droits réservés. Copyright © Lucent Technologies. Tous droits réservés. Copyright © Université de Toronto. Tous droits réservés. Copyright © Daniel Veillard. Tous droits réservés. Copyright © Uniconic, Inc. Copyright IBM Corp. Tous droits réservés. Copyright © MicroQuill Software Publishing, Inc. Tous droits réservés. Copyright © PassMark Software Pty Ltd. Tous droits réservés. Copyright © LogiXML, Inc. Tous droits réservés. Copyright © 2003-2010 Lorenzi Davide. Tous droits réservés. Copyright © Red Hat, Inc. Tous droits réservés. Copyright © The Board of Trustees of the Leland Stanford Junior University. Tous droits réservés. Copyright © EMC Corporation. Tous droits réservés. Copyright © Flexera Software. Tous droits réservés. Copyright © Jinfonet Software. Tous droits réservés. Copyright © Apple Inc. Tous droits réservés. Copyright © Telerik Inc. Tous droits réservés. Copyright © BEA Systems. Tous droits réservés. Copyright © PDFlib GmbH. Tous droits réservés. Copyright © Orientation in Objects GmbH. Tous droits réservés. Copyright © Tanuki Software, Ltd. Tous droits réservés. Copyright © Ricebridge. Tous droits réservés. Copyright © Sencha, Inc. Tous droits réservés. Copyright © Scalable Systems, Inc. Tous droits réservés. Copyright © jQWidgets. Tous droits réservés. Copyright © Tableau Software, Inc. Tous droits réservés. Copyright © MaxMind, Inc. Tous droits réservés. Copyright © TMate Software s.r.o. Tous droits réservés. Copyright © MapR Technologies Inc. Tous droits réservés.

Ce produit inclut des logiciels développés par Apache Software Foundation (<http://www.apache.org/>), et/ou d'autres logiciels sous licence et sous diverses versions Apache License (la « Licence »). Vous pouvez obtenir une copie de ces licences à l'adresse suivante : <http://www.apache.org/licenses/>. Sauf dispositions contraires de la loi en vigueur ou accord écrit, le logiciel distribué sous cette licence est livré « EN L'ÉTAT », SANS GARANTIE NI CONDITION D'AUCUNE SORTE, expresse ou implicite. Se reporter aux Licences pour la langue spécifique régissant les droits et limitations dans le cadre des Licences.

Ce produit inclut des logiciels développés par Mozilla (<http://www.mozilla.org/>), copyright de logiciel The JBoss Group, LLC, tous droits réservés ; copyright de logiciel © 1999-2006 de Bruno Lowagie et Paulo Soares et d'autres logiciels sous licence et sous diverses versions du GNU Lesser General Public License Agreement, accessible sur <http://www.gnu.org/licenses/lgpl.html>. Les matériaux sont fournis gratuitement par Informatica, « en l'état », sans garantie d'aucune sorte, expresse ou implicite, notamment les garanties implicites de conformité légale et d'usage normal.

Le produit inclut les logiciels ACE(TM) et TAO(TM), copyright Douglas C. Schmidt et son groupe de recherche à Washington University, University of California, Irvine, et Vanderbilt University, Copyright (©) 1993-2006, tous droits réservés.

Ce produit inclut des logiciels développés par OpenSSL Project pour une utilisation dans OpenSSL Toolkit (copyright The OpenSSL Project. Tous droits réservés) et la redistribution de ce logiciel est sujette aux termes publiés sur <http://www.openssl.org> et <http://www.openssl.org/source/license.html>.

Ce produit inclut le logiciel Curl, copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://curl.haxx.se/docs/copyright.html>. L'autorisation d'utiliser, copier, modifier et distribuer ce logiciel à toute fin, avec ou sans rémunération, est accordée par les présentes, à la condition que la notification de copyright ci-dessus et cette notification d'autorisation apparaissent dans toutes les copies.

Le produit inclut des logiciels sous copyright 2001-2005 (©) MetaStuff, Ltd. Tous droits réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.dom4j.org/license.html>.

Le produit inclut des logiciels sous copyright © 2004-2007, The Dojo Foundation. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://dojotoolkit.org/license>.

Ce produit inclut le logiciel ICU sous copyright de International Business Machines Corporation et autres. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

Ce produit inclut des logiciels sous copyright © 1996-2006 Per Bothner. Tous Droits Réservés. Votre droit à utiliser de tels matériels est défini dans la licence qui peut être consultée sur <http://www.gnu.org/software/kawa/Software-License.html>.

Ce produit inclut le logiciel OSSP UUID sous copyright © 2002 Ralf S. Engelschall, copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.opensource.org/licenses/mit-license.php>.

Ce produit inclut des logiciels développés par Boost (<http://www.boost.org/>) ou sous licence de logiciel Boost. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur http://www.boost.org/LICENSE_1_0.txt.

Ce produit inclut des logiciels sous copyright © 1997-2007 University of Cambridge. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.pcre.org/license.txt>.

Ce produit inclut des logiciels sous copyright © 2007 The Eclipse Foundation. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://www.eclipse.org/org/documents/epl-v10.php> et <http://www.eclipse.org/org/documents/edl-v10.php>.

Ce produit comprend des logiciels sous licence dont les conditions se trouvent aux adresses : <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/licence.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/javaxservice/>, <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/lyokato/libgeohash/blob/master/LICENSE>, <https://github.com/hjiang/jsonxx/blob/master/LICENSE>, <https://code.google.com/p/lz4/>, <https://github.com/jedisct1/libodium/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>; et <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>.

Ce produit inclut un logiciel sous licence Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), licence Common Development Distribution License (<http://www.opensource.org/licenses/cddl1.php>), licence Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), licence Sun Binary Code License Agreement Supplemental License Terms, licence BSD (<http://www.opensource.org/licenses/bsd-license.php>), le nouvelle licence BSD License (<http://opensource.org/licenses/BSD-3-Clause>), la licence MIT (<http://www.opensource.org/licenses/mit-license.php>), la licence Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) et la licence publique du développeur initial Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

Ce produit inclut des logiciels sous copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. Tous Droits Réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions publiées sur <http://xstream.codehaus.org/license.html>. Ce produit inclut des logiciels développés par Indiana University Extreme! Lab. Pour plus d'informations, veuillez vous rendre sur <http://www.extreme.indiana.edu/>.

Ce produit inclut des logiciels sous copyright © 2013 Frank Balluffi et Markus Moeller. Tous droits réservés. Les autorisations et limitations concernant ce logiciel sont sujettes aux conditions de la licence MIT.

Ce logiciel est protégé par les brevets américains 5 794 246 ; 6 014 670 ; 6 016 501 ; 6 029 178 ; 6 032 158 ; 6 035 307 ; 6 044 374 ; 6 092 086 ; 6 208 990 ; 6 339 775 ; 6 640 226 ; 6 789 096 ; 6 823 373 ; 6 850 947 ; 6 895 471 ; 7 117 215 ; 7 162 643 ; 7 243 110 ; 7 254 590 ; 7 281 001 ; 7 421 458 ; 7 496 588 ; 7 523 121 ; 7 584 422 ; 7 676 516 ; 7 720 842 ; 7 721 270 ; 7 774 791 ; 8 065 266 ; 8 150 803 ; 8 166 048 ; 8 166 071 ; 8 200 622 ; 8 224 873 ; 8 271 477 ; 8 327 419 ; 8 386 435 ; 8 392 460 ; 8 453 159 ; 8 458 230 ; 8 707 336 ; 8 886 617 et RE44 478, par des brevets internationaux et d'autres brevets en instance.

EXCLUSION DE RESPONSABILITÉ : Informatica Corporation fournit cette documentation « en l'état », sans garantie d'aucune sorte, explicite ou implicite, notamment les garanties implicites de non-infraction, de conformité légale ou d'usage normal. Informatica Corporation ne garantit pas que ce logiciel et cette documentation sont exempts d'erreurs. Les informations fournies dans ce logiciel ou cette documentation peuvent inclure des inexactitudes techniques ou des erreurs typographiques. Les informations contenues dans ce logiciel et sa documentation sont sujettes à modification à tout moment sans préavis.

AVIS

Ce produit Informatica (le « Logiciel ») inclut certains pilotes (les « Pilotes DataDirect ») de DataDirect Technologies, une société de Progress Software Corporation (« DataDirect ») qui sont sujets aux conditions suivantes :

1. LES PILOTES DATADIRECT SONT FOURNIS « EN L'ÉTAT », SANS GARANTIE D'AUCUNE SORTE, EXPRESSE OU IMPLICITE, NOTAMMENT LES GARANTIES IMPLICITES DE CONFORMITÉ LÉGALE, D'USAGE NORMAL ET DE NON-INFRACTION.
2. DATADIRECT OU SES FOURNISSEURS TIERS NE POURRONT EN AUCUN CAS ÊTRE TENUS RESPONSABLES ENVERS LE CLIENT UTILISATEUR FINAL DE TOUT DOMMAGE DIRECT, ACCESSOIRE, INDIRECT, SPÉCIAL, CONSÉCUTIF OU AUTRE RÉSULTANT DE L'UTILISATION DES PILOTES ODBC, QU'ILS SOIENT INFORMÉS OU NON À L'AVANCE DE LA POSSIBILITÉ DE TELS DOMMAGES. CES LIMITATIONS S'APPLIQUENT À TOUTES LES CAUSES D'ACTION, NOTAMMENT TOUTE INFRACTION AU CONTRAT, INFRACTION À LA GARANTIE, NÉGLIGENCE, RESPONSABILITÉ STRICTE, REPRÉSENTATION INCORRECTE ET AUTRES TORTS.

Date de publication: 2018-06-09

Sommaire

Préface.....	8
Ressources Informatica.	8
Portail Mon support Informatica.	8
Documentation Informatica.	8
Matrices de disponibilité de produit Informatica.	8
Site Web Informatica.	9
Bibliothèque de procédures Informatica.	9
Base de connaissances Informatica.	9
Canal YouTube du support Informatica.	9
Informatica Marketplace.	9
Informatica Velocity.	9
Support client international Informatica.	10
 Chapitre 1: Présentation de la mise en route.....	 11
Présentation du domaine Informatica.	11
Fonctionnalités disponibles.	13
Introduction à Informatica Analyst.	14
Informatica Developer - Présentation.	14
Page d'accueil Informatica Developer.	15
Feuilles d'aide.	15
Qualité des données et profilage.	16
Scénario du tutoriel.	17
Structure du tutoriel.	17
Tutoriel de Informatica Analyst.	17
Outil Informatica Developer.	18
Prérequis du tutoriel.	19
 Partie I: Débuter avec Informatica Analyst.....	 20
 Chapitre 2: Leçon 1. Configuration de Informatica Analyst.....	 21
Configuration de Informatica Analyst - Présentation.	21
Tâche 1. Se connecter à Informatica Analyst.	22
Tâche 2. Créer un projet.	22
Tâche 3. Créer un dossier.	22
Configuration de Informatica Analyst - Résumé.	23
 Chapitre 3: Leçon 2. Création d'objets de données.....	 24
Création d'objets de données - Présentation.	24
Tâche 1. Créer les objets de données de fichier plat.	25
Tâche 2. Afficher les propriétés de l'objet de données.	25

Création d'objets de données - Résumé.	26
Chapitre 4: Leçon 3. Création de profils rapides.	27
Création de profils rapides - Présentation.	27
Tâche 1. Créer et exécuter un profil rapide.	28
Tâche 2. Afficher les résultats de profil.	28
Création de profils rapides - Résumé.	29
Chapitre 5: Leçon 4. Création de profils personnalisés.	30
Création de profils personnalisés -Présentation.	30
Tâche 1. Créer un profil personnalisé.	31
Tâche 2. Exécuter le profil.	32
Tâche 3. Développer les résultats de profil.	32
Création de profils personnalisés - Résumé.	32
Chapitre 6: Leçon 5. Création de règles d'expression.	33
Création de règles d'expression - Présentation.	33
Tâche 1. Créer des règles d'expression et exécuter le profil.	34
Tâche 2. Afficher la sortie de la règle d'expression.	35
Tâche 3. Éditer les règles d'expression.	35
Création de règles d'expression - Résumé.	36
Chapitre 7: Leçon 6. Création et exécution de fiches d'évaluation.	37
Création et de exécution de fiches d'évaluation - Présentation.	37
Tâche 1. Créer une fiche d'évaluation à partir des résultats de profil.	38
Tâche 2. Exécuter la fiche d'évaluation.	39
Tâche 3. Afficher la fiche d'évaluation.	39
Tâche 4. Éditer la fiche d'évaluation.	40
Tâche 5. Configurer les seuils.	40
Tâche 6. Afficher les graphes de tendance de score.	40
Création et de exécution de fiches d'évaluation - Résumé.	41
Chapitre 8: Leçon 7. Création de tables de référence à partir des colonnes de profil.	42
Création de tables de référence à partir des colonnes de profil - Présentation.	42
Tâche 1. Créer une table de référence à partir des colonnes de profil.	43
Tâche 2. Éditer la table de référence.	44
Création de tables de référence à partir des colonnes de profil - Résumé.	44
Chapitre 9: Leçon 8. Création de tables de référence.	45
Création de tables de référence - Présentation.	45
Tâche 1. Créer une table de référence.	46
Création de tables de référence - Résumé.	46

Partie II: Débuter avec Informatica Developer.....	47
Chapitre 10: Leçon 1. Configuration de Informatica Developer.	48
Configuration de Informatica Developer - Présentation.	48
Tâche 1. Démarrer Informatica Developer.	49
Tâche 2. Ajouter un domaine.	49
Tâche 3. Ajouter un référentiel modèle.	50
Tâche 4. Créer un projet.	50
Tâche 5. Créer un dossier.	50
Tâche 6. Sélectionner un service Data Integration Service par défaut..	51
Configuration de Informatica Developer - Résumé.	51
Chapitre 11: Leçon 2 : Importation d'objets de données physiques.	52
Importation d'objets de données physiques - Présentation.	52
Tâche 1. Importer l'objet de données de fichier plat Boston_Customers.	53
Tâche 2. Importer l'objet de données de fichier plat LA_Customers.	53
Tâche 3. Importer l'objet de données de fichier plat All_Customers.	54
Importation d'objets de données physiques - Résumé.	55
Chapitre 12: Leçon 3. Profilage des données.	56
Profilage des données - Présentation.	56
Tâche 1. Effectuer une analyse de jointure sur deux sources de données.	57
Tâche 2. Afficher les résultats de l'analyse de jointure.	58
Tâche 3. Exécuter un profil sur une source de données.	59
Tâche 4. Afficher les résultats de profilage de colonne.	59
Profilage des données Résumé.	60
Chapitre 13: Leçon 4. Analyse des données.	61
Présentation de l'analyse des données.	61
Tâche 1. Création d'un objet de données cible.	62
Étape 1. Création d'un objet de données LA_Customers_tgt.	62
Étape 2. Configuration des options Lire et Écrire.	63
Étape 3. Ajout de colonnes à l'objet de données.	63
Tâche 2. Création d'un mappage pour analyser les données.	64
Étape 1. Création d'un mappage.	64
Étape 2. Ajout d'objets de données au mappage.	64
Étape 3. Ajout d'une transformation Analyseur au mappage.	65
Étape 4. Configuration de la transformation Analyseur.	65
Tâche 3. Exécution d'un profil sur la transformation Analyseur.	66
Tâche 4. Exécution du mappage.	66
Tâche 5. Affichage de la sortie du mappage.	66
Résumé de l'analyse des données.	67

Chapitre 14: Leçon 5. Normalisation des données 68

Présentation de la normalisation des données.	68
Tâche 1. Création d'un objet de données cible.	69
Étape 1. Création d'un objet de données All_Customers_Stdz_tgt.	69
Étape 2. Configuration des options Lire et Écrire.	70
Tâche 2. Création d'un mappage pour normaliser les données.	70
Étape 1. Création d'un mappage.	71
Étape 2. Ajout d'objets de données au mappage.	71
Étape 3. Ajout d'une transformation Normalisation au mappage.	71
Étape 4. Configuration de la transformation Normalisation.	72
Tâche 3. Exécution du mappage.	73
Tâche 4. Affichage de la sortie du mappage.	73
Résumé de la normalisation des données.	73

Chapitre 15: Leçon 6. Validation des données d'adresse..... 74

Présentation de la validation des données d'adresse.	74
Tâche 1. Création d'un objet de données cible	75
Étape 1. Créez l'objet de données All_Customers_av_tgt.	76
Étape 2. Configuration des options Lire et Écrire.	76
Étape 3. Ajout de ports à l'objet de données.	77
Tâche 2. Création d'un mappage pour valider les adresses.	77
Étape 1. Création d'un mappage.	77
Étape 2. Ajout d'objets de données au mappage.	78
Étape 3. Ajout d'une transformation du validateur d'adresses au mappage.	78
Tâche 3. Configuration de la transformation du validateur d'adresses.	78
Étape 1. Définition du jeu de données de référence d'adresses par défaut.	79
Étape 2. Configuration des ports d'entrée de la transformation du validateur d'adresses.	79
Étape 3. Configuration de la transformation du validateur d'adresses.	80
Étape 4. Connexion des ports inutilisés de la source de données à la cible des données.	81
Tâche 4. Exécution du mappage.	82
Tâche 5. Affichage de la sortie du mappage.	82
Résumé de la validation des données d'adresse.	85

Annexe A: Forum Aux Questions (FAQ)..... 86

FAQ Informatica Analyst.	86
Foire Aux Questions (FAQ) Informatica Developer	86

Préface

Le *Guide de démarrage* Data Quality a été rédigé à l'attention des développeurs et analystes de qualité des données. Il fournit des tutoriels pour aider les utilisateurs débutants à se familiariser avec Informatica Developer et Informatica Analyst. Ce guide suppose une compréhension des concepts de qualité des données, des concepts de fichiers plats et de bases de données relationnelles, ainsi que des moteurs de base de données dans votre environnement.

Ressources Informatica

Portail Mon support Informatica

En tant que client Informatica, vous pouvez accéder au portail Mon support Informatica sur <http://mysupport.informatica.com>.

Ce site contient des informations sur les produits et les groupes d'utilisateurs, des bulletins d'information, un lien vers le système de gestion des dossiers d'assistance à la clientèle d'Informatica (ATLAS), une bibliothèque de procédures Informatica, une base de connaissances Informatica, ainsi que la documentation nécessaire sur les produits Informatica et l'accès à sa communauté d'utilisateurs.

Documentation Informatica

L'équipe Documentation d'Informatica s'efforce de fournir une documentation précise et utilisable. N'hésitez pas à contacter l'équipe Documentation d'Informatica par courriel à l'adresse infa_documentation@informatica.com pour lui faire part de vos questions, commentaires ou suggestions concernant cette documentation. Ces commentaires et suggestions nous permettront d'améliorer notre documentation. Veuillez préciser si vous acceptez d'être contacté au sujet de ces commentaires.

L'équipe Documentation met à jour la documentation chaque fois que nécessaire. Pour obtenir la toute dernière version de la documentation concernant votre produit, consultez la Documentation de produit sur <http://mysupport.informatica.com>.

Matrices de disponibilité de produit Informatica

Les matrices de disponibilité de produit (PAM) indiquent les versions des systèmes d'exploitation, les bases de données et les autres types de sources et cibles de données pris en charge par une version d'un produit. Vous pouvez consulter les PAM sur le portail Mon Support Informatica à l'adresse <https://mysupport.informatica.com/community/my-support/product-availability-matrices>.

Site Web Informatica

Vous pouvez accéder au site Web d'entreprise Informatica sur <http://www.informatica.com>. Le site contient des informations sur Informatica, son expertise, les événements à venir et les bureaux de vente. Vous y trouverez aussi des informations sur ses produits et ses partenaires. Les rubriques de service du site fournissent des informations importantes sur le support technique, la formation et l'éducation, ainsi que les services d'implémentation.

Bibliothèque de procédures Informatica

En tant que client Informatica, vous avez accès à la bibliothèque de procédures Informatica sur <http://mysupport.informatica.com>. La bibliothèque de procédures Informatica est une collection de ressources destinée à vous familiariser avec les produits Informatica et leurs fonctionnalités. Elle regroupe des articles et des démonstrations interactives qui permettent de résoudre des problèmes courants et de comparer les fonctionnalités et les comportements, et qui vous guident lors de la réalisation de tâches concrètes spécifiques.

Base de connaissances Informatica

En tant que client Informatica, vous avez accès à la base de connaissances Informatica sur <http://mysupport.informatica.com>. Utilisez la base de connaissances pour rechercher des solutions documentées aux problèmes techniques connus concernant les produits Informatica. Vous y trouverez également la réponse aux questions les plus fréquentes, des livres blancs et des conseils techniques. N'hésitez pas à contacter l'équipe Base de connaissances Informatica par courriel à l'adresse KB_Feedback@informatica.com pour lui faire part de vos questions, commentaires et suggestions concernant la base de connaissances.

Canal YouTube du support Informatica

Vous pouvez accéder au canal YouTube du support Informatica sur <http://www.youtube.com/user/INFASupport>. Le canal YouTube du support Informatica contient des vidéos concernant les solutions qui vous guident dans l'exécution de tâches spécifiques. Si vous avez des questions, commentaires ou suggestions concernant le canal YouTube du support Informatica, contactez l'équipe de support YouTube par courriel à l'adresse supportvideos@informatica.com ou envoyez un tweet à @INFASupport.

Informatica Marketplace

Informatica Marketplace est un forum où développeurs et partenaires peuvent partager des solutions qui permettent d'augmenter, d'étendre ou d'améliorer les implémentations d'intégration de données. En tirant profit des centaines de solutions disponibles sur Marketplace, vous pouvez améliorer votre productivité et accélérer le temps d'implémentation de vos projets. Vous pouvez accéder à Informatica Marketplace à l'adresse <http://www.informaticamarketplace.com>.

Informatica Velocity

Vous pouvez accéder à Informatica Velocity à l'adresse <http://mysupport.informatica.com>. Développé à partir de l'expérience concrète de centaines de projets de gestion de données, Informatica Velocity représente le savoir collectif de nos consultants, qui ont travaillé avec des entreprises du monde entier pour planifier, développer, déployer et tenir à jour des solutions de gestion des données efficaces. Si vous avez des questions, des commentaires et des suggestions sur Informatica Velocity, contactez le support des services professionnels Informatica à l'adresse ips@informatica.com.

Support client international Informatica

Vous pouvez contacter un centre de support client par téléphone ou via l'assistance en ligne.

L'assistance en ligne requiert un nom d'utilisateur et un mot de passe. Vous pouvez demander un nom d'utilisateur et un mot de passe sur <http://mysupport.informatica.com>.

Les numéros de téléphone du support client international Informatica sont disponibles sur le site Web Informatica à l'adresse

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

CHAPITRE 1

Présentation de la mise en route

Ce chapitre comprend les rubriques suivantes :

- [Présentation du domaine Informatica, 11](#)
- [Introduction à Informatica Analyst, 14](#)
- [Informatica Developer - Présentation, 14](#)
- [Scénario du tutoriel, 17](#)
- [Structure du tutoriel, 17](#)

Présentation du domaine Informatica

Informatica utilise une architecture orientée service qui offre la possibilité de dimensionner les services et de partager les ressources entre plusieurs machines. Le domaine Informatica est l'unité principale de gestion et d'administration des services.

Vous pouvez vous connecter à Informatica Administrator après avoir installé Informatica. Vous pouvez utiliser Administrator Tool pour gérer le domaine et configurer les services d'application afin de pouvoir accéder aux clients d'application restants.

Informatica Data Quality est composé des éléments suivants :

- Clients d'application. Groupe de clients que vous utilisez pour accéder aux fonctionnalités sous-jacentes d'Informatica. Les clients d'application font des demandes au gestionnaire de service ou aux services d'application.
- Services d'application. Groupe de services qui représente la fonctionnalité basée sur serveur. Un domaine Informatica peut contenir un sous-ensemble de services d'application. Vous configurez les services d'application dont les clients d'application ont besoin.
- Entrepôt de profils. Base de données relationnelle utilisée par le service d'intégration de données pour stocker les résultats de profil.
- Référentiels. Groupe de bases de données relationnelles qui stockent des métadonnées sur des objets et processus requis pour traiter les demandes des clients d'application.
- Gestionnaire de service. Service intégré au domaine pour gérer toutes les opérations du domaine. Le gestionnaire de service exécute les services d'application, ainsi que des fonctions de domaine, notamment les authentifications, les autorisations et les connexions.

Le tableau suivant répertorie les clients d'application, à l'exception d'Administrator Tool, des services d'application et des référentiels requis par le client :

Client d'application	Services d'application	Référentiels
Data Analyzer	Service de rapports	Référentiel de l'analyseur de données
Service de rapports et de tableaux de bord Informatica	Service de rapports et de tableaux de bord	Référentiel Jaspersoft
Informatica Analyst	<ul style="list-style-type: none"> - Service Analyst - Service de gestion de contenu - Service d'intégration de données - Service de Référentiel Modèle - Service de recherche 	Référentiel modèle
Informatica Developer	<ul style="list-style-type: none"> - Service Analyst - Service de gestion de contenu - Service d'intégration de données - Service de Référentiel Modèle 	Référentiel modèle
Metadata Manager	<ul style="list-style-type: none"> - Service Metadata Manager - service d'intégration PowerCenter - Service de référentiel PowerCenter 	<ul style="list-style-type: none"> - Référentiel Metadata Manager - Référentiel PowerCenter
Client PowerCenter	<ul style="list-style-type: none"> - service d'intégration PowerCenter - Service de référentiel PowerCenter 	Référentiel PowerCenter
Console Hub des services Web	<ul style="list-style-type: none"> - service d'intégration PowerCenter - Service de référentiel PowerCenter - Hub de services Web 	Référentiel PowerCenter

Un client d'application Informatica n'accède pas aux services d'application suivants :

- Service d'écoute PowerExchange. Gère l'écouteur PowerExchange pour les mouvements de données massifs et la capture de données différentielles. Le service d'intégration PowerCenter se connecte à l'écouteur PowerExchange via le service d'écoute.
- service de journalisation PowerExchange. Gère l'enregistreur PowerExchange pour Linux, UNIX et Windows afin de capturer les données modifiées et les enregistrer dans les fichiers journaux de l'enregistreur PowerExchange. Les données modifiées peuvent provenir des journaux de reprise DB2, les journaux de rétablissement Oracle, d'une base de données de distribution Microsoft SQL Server ou des sources de données sur un système i5/OS ou z/OS.
- Service SAP BW. Écoute les demandes RFC issues de SAP BI et demande que le service d'intégration PowerCenter exécute le flux à des fins d'extraction de ou de chargement dans SAP BI.

Fonctionnalités disponibles

Les produits Informatica utilisent un ensemble commun d'applications. Les fonctionnalités du produit que vous pouvez utiliser dépendent de votre licence de produit.

Le tableau suivant décrit les options de licence et les fonctionnalités applicatives disponibles avec chaque option :

Option de licence	Fonctionnalités d'Informatica Developer	Fonctionnalités d'Informatica Analyst
Data Quality	<ul style="list-style-type: none"> - Créer et exécuter des mappages - Création et exécution de règles - Profilage qui comprend l'utilisation du profil de découverte des données d'entreprise et la découverte de clé primaire, de clé étrangère et de dépendance fonctionnelle - Conserver les résultats de profils inférés - Fiches d'évaluation - Exportation d'objets vers PowerCenter 	<ul style="list-style-type: none"> - Profilage qui comprend la découverte des données d'entreprise - Utiliser la recherche de découverte pour trouver l'emplacement des données et des métadonnées dans les référentiels de profilage - Fiches d'évaluation - Conserver les résultats de profils inférés - Créer et exécuter des règles de profilage - Gestion de table de référence - Créer des spécifications de règle et compiler les spécifications de règle dans les mapplets - Gestion des enregistrements erronés et dupliqués
Data Services	<ul style="list-style-type: none"> - Création de modèles d'objet de données logique - Création et exécution de mappages avec les transformations Data Services - Création de services de données SQL - Création de services Web - Exportation d'objets vers PowerCenter 	<ul style="list-style-type: none"> - Gestion de table de référence
Options de Data Services et de profilage	<ul style="list-style-type: none"> - Création de modèles d'objet de données logique - Création et exécution de mappages avec les transformations Data Services - Création de services de données SQL - Création de services Web - Exportation d'objets vers PowerCenter - Création et exécution de règles avec les transformations Data Services - Profilage 	<ul style="list-style-type: none"> - Gestion de table de référence

Introduction à Informatica Analyst

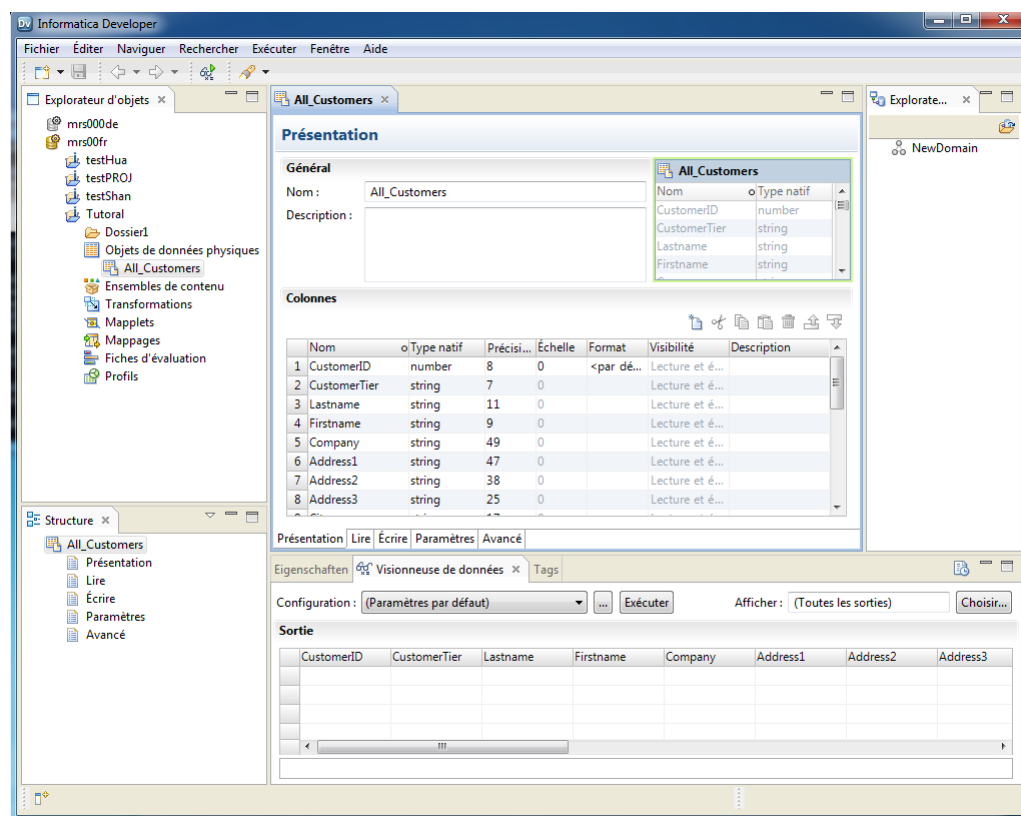
Informatica Analyst est un client d'application basé sur le Web que les analystes peuvent utiliser pour analyser, nettoyer, normaliser, effectuer le profil et évaluer les données dans une entreprise.

Selon votre licence, les analystes et développeurs professionnels utilisent l'outil Analyst pour une collaboration orientée données. Vous pouvez effectuer un profilage de colonnes et de règles, traiter des fiches d'évaluation et gérer les enregistrements incorrects et les enregistrements dupliqués. Vous pouvez également gérer les données de référence et fournir les données aux programmeurs dans le cadre d'une solution de qualité des données.

Informatica Developer - Présentation

Informatica Developer est un client d'application utilisé par les développeurs pour concevoir et implémenter des solutions de données.

La figure suivante illustre l'outil Developer :



L'outil Developer peut afficher les vues suivantes :

Vue Explorateur d'objets

Affiche les projets, les dossiers et les objets à l'intérieur des projets et des dossiers.

Vue Explorateur de connexion

Affiche les connexions aux bases de données relationnelles.

Vue Structure

Affiche les objets qui dépendent d'un objet sélectionné dans la vue **Explorateur d'objets**.

Vue Feuilles d'aide

Affiche la feuille d'aide que vous ouvrez. Pour ouvrir une feuille d'aide, cliquez sur **Aide > Feuilles d'aide** et sélectionnez une feuille d'aide.

vue Propriétés

Affiche les propriétés d'un objet sélectionné dans l'éditeur.

Vue Visionneuse de données

Affiche les données source et les résultats de profil, et prévisualise la sortie d'une transformation.

Vue Dépendances d'objets

Affiche les dépendances d'objets lorsque vous affichez, modifiez ou supprimez un objet.

Affichage des alertes

Affiche des alertes concernant le statut de la connexion.

Vue Journal de validation

Affiche les erreurs de validation d'objet.

L'outil Developer peut également afficher d'autres vues. Vous pouvez masquer et déplacer des vues vers un autre emplacement dans l'espace de travail de l'outil Developer. Cliquez sur **Fenêtre > Afficher la vue** pour sélectionner les vues à afficher.

Page d'accueil Informatica Developer

La première fois que vous ouvrez l'outil Developer, la page d'accueil apparaît. Utilisez la page d'accueil pour en savoir plus sur l'outil Developer, le configurer et commencer à l'utiliser.

La page d'accueil affiche les options suivantes :

- **Présentation.** Cliquez sur le bouton Présentation pour obtenir une présentation des solutions de qualité des données et des services de données.
- **Premiers pas.** Cliquez sur le bouton Premiers pas pour en savoir plus sur la configuration de l'outil Developer et l'accès aux leçons Informatica Data Quality et Informatica Data Services.
- **Tutoriels.** Cliquez sur le bouton Tutoriels pour accéder aux leçons concernant les solutions de qualité et de services de données.
- **Ressources Web.** Cliquez sur le bouton Ressources Web pour accéder à un lien vers mysupport.informatica.com, où vous pouvez accéder à la bibliothèque de procédures Informatica. La Bibliothèque de procédures Informatica contient les articles sur Informatica Data Quality, Informatica Data Services, et d'autres produits Informatica.
- **Espace de travail.** Cliquez sur le bouton Espace de travail pour commencer à travailler dans l'outil Developer.

Feuilles d'aide

L'outil Developer inclut des feuilles d'aide faisant partie de l'aide en ligne. Une feuille d'aide vous guide, étape par étape, à travers une ou plusieurs tâches dans l'outil Developer.

Lorsque vous consultez une feuille d'aide, vous pouvez exécuter les tâches afin d'en vérifier les résultats. Par exemple, après avoir terminé une feuille d'aide pour importer et prévisualiser un objet de données

relationnelles, vous avez importé une table de base de données relationnelles et prévisualisé les données dans l'outil Developer.

Pour accéder aux feuilles d'aide, cliquez sur **Aide > Feuilles d'aide**.

Qualité des données et profilage

Utilisez les fonctionnalités de qualité des données dans l'outil Developer pour analyser le contenu et la structure de vos données. Vous pouvez améliorer les données afin de répondre aux besoins de votre entreprise.

Utilisez l'outil Developer pour concevoir et exécuter des processus qui réalisent les objectifs suivants :

- Effectuer le profil des données. Le profilage révèle le contenu et la structure de vos données. Le profilage est une étape clé dans tout projet de données, car il permet d'identifier les points forts et les points faibles de vos données, tout en vous aidant à définir le plan de votre projet.
- Créer des fiches d'évaluation pour examiner la qualité des données. Une fiche d'évaluation est une représentation graphique des mesures de qualité dans un profil.
- Normaliser les valeurs des données. Normalisez les données pour éliminer les erreurs et les incohérences détectées lors de l'exécution d'un profil. Vous pouvez normaliser les variations de ponctuation, de formatage et d'orthographe. Par exemple, vous pouvez vous assurer que les valeurs de ville, état et code postal sont cohérentes.
- Analyser les enregistrements. Analysez les enregistrements de données pour améliorer la structure des enregistrements et tirer des informations supplémentaires de vos données. Vous pouvez diviser un champ de données de forme libre unique en champs qui contiennent différents types d'informations. Vous pouvez aussi ajouter des informations à vos enregistrements. Par exemple, vous pouvez indiquer qu'un enregistrement concerne un client particulier ou d'entreprise.
- Valider les adresses postales. La validation des adresses évalue et améliore l'exactitude et la possibilité de livraison de vos données d'adresse postale. La validation des adresses corrige les erreurs dans les adresses et complète les adresses partielles en comparant les enregistrements d'adresse avec des données de référence provenant de transporteurs postaux nationaux. La validation des adresses peut aussi ajouter des informations postales qui accélèrent la distribution du courrier tout en réduisant les coûts associés.
- Rechercher les enregistrements dupliqués. Une analyse d'enregistrements dupliqués compare un ensemble d'enregistrements entre eux pour détecter les doublons ou valeurs comparables dans les colonnes de données sélectionnées. Vous devez définir le niveau de similarité qui indique une bonne correspondance entre des valeurs de champ. Vous pouvez aussi définir le poids relatif donné à chaque colonne dans les calculs de correspondance. Par exemple, vous pouvez donner la priorité aux informations de nom de famille sur les informations de prénom.
- Créer et exécuter des règles de qualité des données. Informatica fournit des règles prédéfinies que vous pouvez exécuter ou modifier pour correspondre aux objectifs de votre projet. Vous pouvez créer des règles dans l'outil Developer.
- Collaborer avec les utilisateurs d'Informatica. Les règles et tables de données de référence que vous ajoutez au référentiel modèle sont accessibles aux utilisateurs des outils Developer et Analyst. Les utilisateurs peuvent collaborer sur des projets et différents utilisateurs peuvent s'approprier des objets à différentes phases d'un projet.
- Exporter des mappages vers PowerCenter. Vous pouvez exporter des mappages vers PowerCenter afin de réutiliser les métadonnées pour une intégration physique des données ou pour créer des services Web.

Scénario du tutoriel

HypoStores Corporation est une organisation de grande distribution à l'échelle nationale qui, de son siège social à Boston, gère des magasins dans plusieurs états. Elle intègre régulièrement les données opérationnelles de ses magasins à travers le pays dans l'entrepôt de données du siège social. Elle a récemment ouvert un magasin à Los Angeles.

Le siège social comprend une équipe ICC centrale formée d'administrateurs, de développeurs et d'architectes chargés de fournir une couche de services de données commune à toutes les applications composites et BI (Business Intelligence). Les applications BI comprennent un système de gestion de la clientèle CRM (Customer Relation Management) qui contient les fichiers maîtres des données clients utilisés pour la facturation et le marketing.

HypoStores Corporation doit effectuer les tâches suivantes pour intégrer les données des opérations de Los Angeles à celles du siège social de Boston.

- Vérifier que les données de Boston et de Los Angeles ne comportent pas de problème en termes de qualité.
- Analyser les informations à partir des données de Los Angeles.
- Normaliser les informations d'adresse pour les données de Boston et de Los Angeles.
- Valider l'exactitude des informations d'adresse postale dans les données à des fins de CRM.

Structure du tutoriel

Le *Guide de démarrage* contient des tutoriels composés de leçons et de tâches.

Leçons

Chaque leçon présente des concepts qui vous aideront à comprendre les tâches à réaliser dans la leçon. La leçon fournit des spécifications d'entreprise extraites du scénario global. Les objectifs de la leçon sont de mettre en évidence les tâches que vous exécuterez pour remplir les spécifications de l'entreprise. Chaque leçon donne une durée approximative de son suivi. Lorsque vous avez terminé les tâches de la leçon, vous pouvez en consulter le résumé.

Si l'environnement de l'outil n'est pas configuré, la première leçon de chaque tutoriel vous permet de le faire.

Tâches

Les tâches fournissent des instructions étape par étape. Effectuez toutes les tâches dans l'ordre indiqué pour suivre la leçon.

Tutoriel de Informatica Analyst

Au cours de ce tutoriel, un analyste se connecte à l'outil Analyst pour créer des projets et dossiers, des profils et règles, des données de score, ainsi que des tables de référence.

Les leçons que vous pouvez suivre dépendent des produits dont vous disposez : Informatica Data Quality ou Informatica Data Services.

Le tableau suivant décrit les leçons que vous pouvez suivre selon votre produit :

Leçon	Description	Produit
Leçon 1. Configuration d'Informatica Analyst	Connectez-vous à l'outil Analyst et créez un projet et un dossier pour les leçons du tutoriel.	Qualité des données Data Services
Leçon 2. Création d'objets de données	Importez un fichier plat comme objet de données et prévisualisez les données.	Qualité des données
Leçon 3. Création de profils rapides	Création d'un profil rapide pour obtenir rapidement une idée de la qualité des données.	Qualité des données
Leçon 4. Création de profils personnalisés	Créez un profil personnalisé pour configurer des colonnes et des options d'échantillonnage et de développement.	Qualité des données
Leçon 5. Création de règles d'expression	Créez des règles d'expression pour modifier et effectuer le profil des valeurs de colonne.	Qualité des données
Leçon 6. Création et exécution de fiches d'évaluation	Créez et exécutez une fiche d'évaluation pour mesurer la progression de la qualité des données dans le temps.	Qualité des données
Leçon 7. Création de tables de référence à partir des résultats de profil	Créez une table de référence que vous pouvez utiliser pour normaliser les données source.	Qualité des données Data Services
Leçon 8. Création de tables de référence	Créez une table de référence pour établir des relations entre les données source et des valeurs standard valides.	Qualité des données Data Services

Outil Informatica Developer

Dans ce tutoriel, vous allez utiliser l'outil Developer pour effectuer plusieurs opérations de qualité des données.

Les utilisateurs d'Informatica Data Quality utilisent l'outil Developer pour créer et exécuter des profils qui analysent le contenu et la structure de données.

Les utilisateurs de Informatica Data Quality utilisent l'outil Developer pour concevoir et exécuter des processus qui améliorent la qualité des données.

Suivez les leçons suivantes dans le tutoriel Data Quality :

Leçon 1. Configuration de Informatica Developer

Créez une connexion à un référentiel modèle géré par un service Service de Référentiel Modèle dans un domaine. Créez un projet et un dossier afin d'y stocker votre travail pour les leçons du tutoriel. Sélectionnez un service Data Integration Service par défaut.

Leçon 2. Importation d'objets de données physiques

Vous allez définir des processus de qualité des données pour les fichiers de données clients associés à ces objets.

Leçon 3. Profilage des données

Le profilage révèle le contenu et la structure de vos données.

Le profilage inclut une analyse de jointure, à savoir une forme d'analyse qui détermine si une jointure valide est possible entre deux colonnes de données.

Leçon 4. Analyse de données

L'analyse enrichit vos enregistrements de données et améliore la structure des enregistrements. Elle peut trouver des informations utiles dans vos données et tirer de nouvelles informations des données actuelles.

Leçon 5. Normalisation des données

La normalisation élimine les erreurs de données et les incohérences détectées lors du profilage.

Leçon 6. Validation des données d'adresse

La validation des adresses évalue l'exactitude et la possibilité de livraison de vos adresses postales et répare les erreurs et omissions dans les adresses.

Prérequis du tutoriel

Avant de commencer les leçons du tutoriel, le domaine Informatica doit être actif avec au moins un nœud configuré.

Le programme d'installation comprend des fichiers de tutoriel dont vous aurez besoin pour suivre les leçons. Vous trouverez tous les fichiers à la fois dans les installations client et serveur :

- Vous trouverez les fichiers du tutoriel dans l'emplacement suivant du chemin d'installation de l'outil Developer :
`<Répertoire d'installation Informatica>\clients\DeveloperClient\Tutorials`
- Vous trouverez les fichiers du tutoriel dans l'emplacement suivant du chemin d'installation des services :
`<Répertoire d'installation Informatica>\server\Tutorials`

Vous avez besoin des fichiers suivants pour les leçons du tutoriel :

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

Partie I : Débuter avec Informatica Analyst

Cette partie contient les chapitres suivants :

- [Leçon 1. Configuration de Informatica Analyst, 21](#)
- [Leçon 2. Création d'objets de données, 24](#)
- [Leçon 3. Création de profils rapides, 27](#)
- [Leçon 4. Création de profils personnalisés, 30](#)
- [Leçon 5. Création de règles d'expression, 33](#)
- [Leçon 6. Création et exécution de fiches d'évaluation, 37](#)
- [Leçon 7. Création de tables de référence à partir des colonnes de profil, 42](#)
- [Leçon 8. Création de tables de référence, 45](#)

CHAPITRE 2

Leçon 1. Configuration de Informatica Analyst

Ce chapitre comprend les rubriques suivantes :

- [Configuration de Informatica Analyst - Présentation, 21](#)
- [Tâche 1. Se connecter à Informatica Analyst, 22](#)
- [Tâche 2. Créer un projet, 22](#)
- [Tâche 3. Créer un dossier, 22](#)
- [Configuration de Informatica Analyst - Résumé, 23](#)

Configuration de Informatica Analyst - Présentation

Avant de commencer les leçons de ce tutoriel, vous devez configurer Analyst Tool. Pour configurer Analyst Tool, connectez-vous à Analyst Tool et créez un projet et un dossier pour stocker votre travail.

Le domaine Informatica est une collection de nœuds et de services qui définissent l'environnement Informatica. Les services du domaine incluent Analyst Service et Service de Référentiel Modèle. Analyst Service exécute Analyst Tool et Service de Référentiel Modèle gère le référentiel modèle. Lorsque vous travaillez dans l'outil Analyst, celui-ci stocke les actifs que vous créez dans le référentiel modèle.

Vous devez créer un projet avant de pouvoir créer des objets dans l'outil Analyst. Un projet contient des actifs de l'outil Analyst. Un projet peut aussi contenir des dossiers qui stockent les actifs associés, tels que des objets de données qui font partie des mêmes spécifications d'entreprise.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Vous connecter à Analyst Tool.
- Créer un projet pour stocker les objets définis dans l'outil Analyst.
- Créer un dossier dans le projet susceptible de stocker des objets associés.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Un administrateur a configuré Service de Référentiel Modèle et Analyst Service dans Administrator Tool.
- Vous disposez du nom d'hôte et du numéro de port pour Analyst Tool.

- Vous disposez d'un nom d'utilisateur et d'un mot de passe pour accéder à Analyst Service. Vous pouvez obtenir ces informations auprès de l'administrateur.

Délai

Prévoyez 5 à 10 minutes pour cette leçon.

Tâche 1. Se connecter à Informatica Analyst

Connectez-vous à l'outil Analyst pour commencer le tutoriel.

1. Démarrez le navigateur Microsoft Internet Explorer ou Google Chrome.
2. Dans le champ Adresse, entrez l'URL d'Informatica Analyst :
`http[s]://<host name>:<port number>/analyst`
3. Si le domaine utilise l'authentification LDAP ou native, entrez votre nom d'utilisateur et mot de passe sur la page de connexion.
4. Sélectionnez **Natif** ou le nom d'un domaine de sécurité spécifique.

Le champ Domaine de sécurité apparaît lorsque le domaine Informatica utilise LDAP ou l'authentification Kerberos. Si vous ne connaissez pas le domaine de sécurité auquel appartient votre compte utilisateur, contactez l'administrateur de domaine Informatica.

5. Cliquez sur **Connexion**.

L'outil Analyst s'ouvre dans l'espace de travail **Démarrer**.

Tâche 2. Créer un projet

Dans cette tâche, vous allez créer un projet pour contenir les actifs que vous créez dans l'outil Analyst. Créez un projet de tutoriel qui contiendra le dossier du projet.

1. Dans l'en-tête **Gérer**, cliquez sur **Projets**.
L'espace de travail **Projets** s'affiche.
2. Dans le menu **Actions**, cliquez sur **Nouveau > Projet**.
La fenêtre **Nouveau projet** s'affiche.
3. Entrez votre nom avec le préfixe « Tutoriel_ » comme nom du projet.
4. Cliquez sur **OK**.

Tâche 3. Créer un dossier

Dans cette tâche, vous allez créer un dossier pour stocker les actifs apparentés. Vous pouvez créer un dossier dans un projet ou dans un autre dossier. Créez un dossier nommé Clients pour stocker les objets associés au projet de qualité des données.

1. Dans le panneau **Projets**, sélectionnez le projet tutoriel.

2. Dans le menu **Actions**, cliquez sur **Nouveau > Dossier**.

La fenêtre **Nouveau dossier** s'affiche.

3. Saisissez Clients pour le nom de dossier.

4. Cliquez sur **OK**.

Le dossier apparaît sous le projet tutoriel.

Configuration de Informatica Analyst - Résumé

Dans cette leçon, vous avez appris que l'outil Analyst stocke les objets dans des projets et dossiers. Un référentiel modèle contient des projets et des dossiers. Analyst Service exécute Analyst Tool. Service de Référentiel Modèle gère le référentiel modèle. Analyst Service et Service de Référentiel Modèle sont des services d'application dans le domaine Informatica.

Vous vous êtes connecté à Analyst Tool et avez créé un projet et un dossier.

Maintenant, vous pouvez utiliser Developer Tool pour suivre les autres leçons de ce tutoriel.

CHAPITRE 3

Leçon 2. Création d'objets de données

Ce chapitre comprend les rubriques suivantes :

- [Création d'objets de données - Présentation, 24](#)
- [Tâche 1. Créer les objets de données de fichier plat, 25](#)
- [Tâche 2. Afficher les propriétés de l'objet de données, 25](#)
- [Création d'objets de données - Résumé, 26](#)

Création d'objets de données - Présentation

Dans l'outil Analyst, un objet de données est une représentation de données basée sur un fichier plat ou une table de base de données relationnelles. Vous créez un fichier plat ou objet de table puis exécutez un profil par rapport aux données du fichier plat ou à la table de base de données relationnelles. Lorsque vous créez un objet de données de fichier plat dans l'outil Analyst, vous pouvez charger le fichier dans le cache du fichier plat sur la machine qui exécute l'outil Analyst ou vous pouvez spécifier l'emplacement réseau où le fichier est stocké.

Scénario

HypoStores conserve les données clients de Los Angeles dans des fichiers plats. HypoStores a besoin d'effectuer le profil et analyser les données et d'effectuer des tâches de qualité des données.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

1. Charger le fichier plat à l'emplacement du cache de fichier plat et créer un objet de données.
2. Prévisualiser les données pour l'objet de données de fichier plat.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez effectué la leçon 1 de ce tutoriel.
- Vous disposez du fichier plat LA_Customers.csv. Vous pouvez télécharger le fichier <Installation Root Directory>\<Release Version>\clients\DeveloperClient\Tutorials.

Timing

Prévoyez 5 à 10 minutes pour effectuer cette tâche.

Tâche 1. Créer les objets de données de fichier plat

Dans cette tâche, vous créez un objet de données fichier plat à partir du fichier LA_Customers.

1. Dans l'outil Analyst, cliquez sur **Nouveau > Objet de données de fichier plat**.
L'assistant **Ajouter un fichier plat** s'affiche.
2. Sélectionnez **Parcourir et charger** et cliquez sur **Parcourir**.
3. Allez à l'emplacement de LA_Customers.csv et cliquez sur **Ouvrir**.
4. Cliquez sur **Suivant**.
Le panneau **Choisissez le type d'importation** affiche les options **Délimité** et **Largeur fixe**. Sélectionnez l'option **Délimité**. L'option par défaut est **Délimité**.
5. Cliquez sur **Suivant**.
6. Dans **Spécifier les délimiteurs et qualificateurs de texte utilisés dans vos données**, sélectionnez **Guillemets doubles** comme qualificateur de texte.
7. Dans **Spécifier les lignes à importer**, sélectionnez **Importer depuis la première ligne** pour importer les noms de colonne à partir de la première ligne non vide.
Le panneau **Aperçu** est actualisé pour afficher les titres de colonnes depuis la première ligne.
8. Cliquez sur **Suivant**.
Le panneau **Attributs des colonnes** affiche le type de données, la précision, l'échelle et le format de chaque colonne.
9. Cliquez sur **Suivant**.
Le champ **Nom** affiche LA_Customers.
10. Sélectionnez le projet **Tutorial_** et le dossier **Customers**.
11. Cliquez sur **Terminer**.
L'objet de données s'affiche dans les contenus de dossier pour le dossier Clients.

Tâche 2. Afficher les propriétés de l'objet de données

Dans cette tâche, vous pouvez afficher les propriétés de l'objet de données LA_Customers.

1. Dans le panneau **Prévisualisation des données**, consultez la structure et le contenu de l'objet de données LA_Customers.
L'outil Analyst affiche les 100 premières lignes de l'objet de données de fichier plat.
2. Cliquez sur **Propriétés**.
Le panneau **Propriétés** affiche le nom, la description et l'emplacement de l'objet de données. Vous pouvez également voir les noms de colonne et les propriétés de colonne pour l'objet de données.

Création d'objets de données - Résumé

Dans cette leçon, vous avez appris que les objets de données sont des représentations de données basées sur un fichier plat ou sur une table de base de données relationnelles. Vous avez appris que vous pouvez créer un objet de données de fichier plat et y prévisualiser les données.

Vous avez chargé un fichier plat et créé un objet de données de fichier plat, prévisualisé les données pour l'objet de données, et affiché les propriétés pour l'objet de données.

Après avoir créé un objet de données, vous créez un profil rapide pour l'objet de données dans la Leçon 3, et vous créez un profil personnalisé pour l'objet de données dans la Leçon 4.

CHAPITRE 4

Leçon 3. Création de profils rapides

Ce chapitre comprend les rubriques suivantes :

- [Création de profils rapides - Présentation, 27](#)
- [Tâche 1. Créer et exécuter un profil rapide, 28](#)
- [Tâche 2. Afficher les résultats de profil, 28](#)
- [Création de profils rapides - Résumé, 29](#)

Création de profils rapides - Présentation

Un profil représente l'analyse de la qualité des données basée sur le contenu et la structure des données. Un profil rapide est un profil que vous créez avec des options par défaut. Utilisez un profil rapide pour obtenir des résultats de profil sans configurer toutes les colonnes et options d'un profil.

Créez et exécutez un profil rapide pour analyser la qualité des données lorsque vous démarrez un projet de qualité des données. Lorsque vous créez un objet de profil rapide, vous sélectionnez l'objet de données et les colonnes d'objet de données que vous voulez analyser. Un profil rapide ignore la colonne de profil et la configuration des options. L'outil Analyst effectue le profilage sur le fichier plat temporaire pour l'objet de données de fichier plat.

Scénario

HypoStores souhaite intégrer les données de ses nouveaux bureaux de Los Angeles dans son entrepôt de données. Avant de pouvoir intégrer les données dans l'entrepôt de données, elles doivent être nettoyées. Vous êtes l'analyste responsable de l'évaluation de la qualité des données et de la transmission des informations au programmeur responsable du nettoyage des données. Vous voulez afficher rapidement les résultats de profil et vous faire une première idée de la qualité des données.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

1. Créer et exécuter un profil rapide pour l'objet de données fichier plat LA_Customers.
2. Afficher les résultats de profil.

Prérequis

Avant de commencer cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 et 2 de ce tutoriel.

Timing

Prévoyez 5 à 10 minutes pour cette leçon.

Tâche 1. Créer et exécuter un profil rapide

Dans cette tâche, vous créez un profil rapide pour toutes les colonnes de l'objet de données et utilisez les options d'échantillonnage et de développement par défaut.

1. Cliquez sur **Nouveau > Profil d'objet de données**.
L'assistant **Nouveau profil** s'affiche dans l'espace de travail **Découverte**. L'option **Profilage de colonnes** est sélectionnée par défaut.
2. Cliquez sur **Suivant**.
3. Dans la section **Sources**, accédez à **LA_Customers**.
4. Cochez la case en regard de **LA_Customers**.
5. Cliquez sur **Enregistrer et exécuter** pour créer et exécuter le profil.
L'outil Analyst crée le profil et les résultats de profil s'affichent dans l'espace de travail **Découverte**.

Tâche 2. Afficher les résultats de profil

Dans cette tâche, vous utilisez l'onglet **Profilage de colonne** du profil LA_Customers afin d'avoir un aperçu rapide des résultats de profil.

1. Dans l'espace de travail **Découverte**, cliquez sur l'en-tête pour que la colonne **NULL** trie les valeurs.
Notez que les colonnes Address2, Address3 et City2 ont 100 % de valeurs Null.
Dans la Leçon 4, vous créez un profil personnalisé pour exclure ces colonnes.
2. Dans la colonne **Nom** du volet de gauche, cliquez sur **Nom complet**. Les valeurs de la colonne s'affichent dans l'onglet **Valeurs**.
Notez que les prénoms et noms de famille ne s'affichent pas dans des colonnes séparées.
Dans la Leçon 5, vous créez une règle pour séparer les prénoms et noms de famille dans des colonnes distinctes.
3. Dans la colonne **Nom** du volet de gauche, cliquez sur **CustomerTier**.
Notez que les valeurs de CustomerTier ne sont pas cohérentes.
Dans la Leçon 6, vous créez une fiche d'évaluation pour évaluer les valeurs CustomerTier. Dans la Leçon 7, vous créez une table de référence qu'un développeur peut utiliser pour normaliser les valeurs CustomerTier.

4. Dans la colonne **Nom** du volet de gauche, cliquez sur **État**, puis cliquez sur l'onglet **Modèles** dans le volet de droite.

Notez que 483 colonnes ont une forme XX, ce qui indique les valeurs valides. Dix-sept valeurs ne sont pas valides car elles ne correspondent pas à la forme valide.

Dans la Leçon 6, vous créez une fiche d'évaluation pour évaluer les valeurs État.

Création de profils rapides - Résumé

Dans cette leçon, vous avez appris qu'un profil rapide affiche les résultats de profil sans configurer toutes les colonnes et options d'échantillonnage de lignes pour un profil. Vous avez appris à créer et exécuter un profil rapide pour analyser la qualité des données lorsque vous démarrez un projet de qualité des données. Vous avez aussi appris que l'outil Analyst effectue le profilage sur le fichier plat temporaire pour l'objet de données de fichier plat.

Vous avez créé un profil rapide et analysé les résultats de profil. Vous avez obtenu de plus amples informations sur les colonnes du profil, y compris les valeurs null et types de données. Vous avez aussi utilisé les valeurs de colonne et formes pour identifier les problèmes de qualité des données.

Après avoir analysé les résultats d'un profil rapide, vous pouvez effectuer les tâches suivantes :

- Créer un profil personnalisé pour exclure les colonnes du profil et inclure seulement les colonnes qui vous intéressent.
- Créer une règle d'expression pour créer des colonnes virtuelles et affectuer leur profil.
- Créer une table de référence pour inclure les valeurs valides d'une colonne.

CHAPITRE 5

Leçon 4. Création de profils personnalisés

Ce chapitre comprend les rubriques suivantes :

- [Création de profils personnalisés -Présentation, 30](#)
- [Tâche 1. Créer un profil personnalisé, 31](#)
- [Tâche 2. Exécuter le profil, 32](#)
- [Tâche 3. Développer les résultats de profil, 32](#)
- [Création de profils personnalisés - Résumé, 32](#)

Création de profils personnalisés -Présentation

Un profil représente l'analyse de la qualité des données basée sur le contenu et la structure des données. Un profil personnalisé est un profil que vous créez lorsque vous voulez configurer les colonnes, les options d'échantillonnage et d'exploration pour un profilage plus rapide. Configurez les options d'échantillonnage pour sélectionner les lignes d'exemple dans la source. Configurez les options d'exploration pour explorer les lignes de données dans les données source ou intermédiaires. Vous pouvez choisir d'exécuter le profil dans un environnement natif ou Hive.

Vous créez et exécutez un profil pour analyser la qualité des données lorsque vous démarrez un projet de qualité des données. Lorsque vous créez un objet de profil, vous commencez par sélectionner l'objet de données et les colonnes d'objet de données sur lesquelles vous voulez exécuter un profil.

Scénario

HypoStores a besoin d'intégrer les données de ses nouveaux bureaux de Los Angeles dans son entrepôt de données. HypoStores souhaite accéder à la qualité des données de niveau des clients dans le fichier de données clients de Los Angeles. Vous êtes l'analyste responsable de l'évaluation de la qualité des données et de la transmission des informations au programmeur responsable du nettoyage des données.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

1. Créer un profil personnalisé pour l'objet de données de fichier plat et exclure les colonnes avec valeur null.
2. Exécuter le profil pour analyser le contenu et la structure de la colonne CustomerTier.
3. Développer les lignes pour les résultats de profil.

Prérequis

Avant de commencer cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1, 2 et 3 de ce tutoriel.

Timing

Prévoyez 5 à 10 minutes pour cette leçon.

Tâche 1. Créer un profil personnalisé

Dans cette tâche, vous utilisez l'assistant **Nouveau profil** pour créer un profil personnalisé. Lorsque vous créez un profil, vous sélectionnez l'objet de données et les colonnes sur lesquelles vous voulez exécuter un profil. Vous configurez également les options d'échantillonnage et les options d'exploration.

1. Cliquez sur **Nouveau > Profil d'objet de données**.

L'assistant **Nouveau profil** s'affiche. Le profilage de colonne est sélectionné par défaut.

2. Cliquez sur **Suivant**.

3. Dans la section **Sources**, allez dans l'objet de données LA_Customers.

La section **Colonnes** affiche les colonnes de l'objet de données.

4. Sélectionnez la case en regard de **LA_Customers**.

5. Cliquez sur **Suivant**.

6. Dans le champ Nom, entrez **Profile_LA_Customers_Custom**.

7. Dans la section **Dossiers**, sélectionnez le dossier **Clients** dans le projet Tutoriel.

La section **Profils** affiche Profile_LA_Customers.

8. Cliquez sur **Suivant**.

9. Dans la section **Colonnes**, effacez les colonnes Address2, Address3 et City2 .

10. Dans la section **Options d'échantillonnage**, vérifiez que l'option **Toutes les lignes** est sélectionnée.

11. Vérifiez que l'option **Exclure l'inférence du type de données pour les colonnes avec un type de données approuvé** est sélectionnée. Ce paramètre exclut les colonnes contenant un type de données approuvé de l'inférence de type de données de l'exécution de profil.

12. Dans la section **Options de développement**, vérifiez que **Développer les données en direct** est sélectionné.

13. Cliquez sur **Suivant**.

14. Définissez éventuellement un filtre pour le profil.

15. Cliquez sur **Suivant**.

16. Pour exécuter le profil dans l'environnement Hive, sélectionnez **Hive**, puis sélectionnez une connexion Hive. La connexion Hive permet au service d'intégration de données de communiquer avec la grappe Hadoop pour forcer l'exécution du profil depuis le service d'intégration de données vers la grappe Hadoop.

17. Cliquez sur **Enregistrer**.

L'outil Analyst crée le profil et l'affiche le profil dans l'espace de travail **Découverte**. Vous devez exécuter le profil pour afficher les résultats.

Tâche 2. Exécuter le profil

Dans cette tâche, vous exécutez un profil pour effectuer le profilage sur l'objet de données et affichez les résultats de profil. Analyst Tool effectue le profilage sur le fichier plat temporaire pour l'objet de données fichier plat.

1. Vérifiez que vous êtes dans l'espace de travail **Découverte**.
Vous pouvez voir le profil **Profile_LA_Customers_Custom** dans l'espace de travail.
2. Cliquez sur **Actions > Exécuter le profil**.
L'outil Analyst effectue le profilage sur l'objet de données et affiche les résultats du profil.

Tâche 3. Développer les résultats de profil

Dans cette tâche, vous développez les valeurs de la colonne CustomerTier pour afficher les lignes sous-jacentes dans l'objet de données du profil.

1. Vérifiez que vous êtes dans l'espace de travail **Découverte**.
Vous pouvez voir le profil **Profile_LA_Customers_Custom** dans l'espace de travail.
2. Dans le panneau **Profilage de colonne**, sélectionnez la colonne **CustomerTier**.
Les valeurs de la colonne s'affichent dans l'onglet **Valeurs** du volet de droite.
3. Utilisez la touche Maj pour sélectionner les valeurs Diamond, Ruby, Emerald et Bronze.
4. Cliquez avec le bouton droit et sélectionnez **Développement**.
Les lignes de la colonne ayant une valeur Diamond, Ruby, Emerald et Bronze s'affichent dans le panneau **Développement**. Le panneau **Développement** affiche les 100 premières lignes de la colonne sélectionnée. Le titre du panneau **Développement** affiche la logique utilisée pour la colonne source.

Création de profils personnalisés - Résumé

Dans cette leçon, vous avez appris comment configurer les colonnes profilées et configurer les options d'échantillonnage et de développement. Vous avez appris que vous pouvez développer pour afficher les lignes sous-jacentes pour les valeurs des colonnes et que vous pouvez configurer les colonnes incluses lorsque vous affichez les valeurs de colonnes.

Vous avez créé un profil personnalisé qui incluait la colonne CustomerTier, exécuté le profil et développé les lignes sous-jacentes pour la colonne CustomerTier dans les résultats.

Utilisez l'objet de profil personnalisé pour créer une règle d'expression dans la leçon 5.

CHAPITRE 6

Leçon 5. Création de règles d'expression

Ce chapitre comprend les rubriques suivantes :

- [Création de règles d'expression - Présentation, 33](#)
- [Tâche 1. Créer des règles d'expression et exécuter le profil, 34](#)
- [Tâche 2. Afficher la sortie de la règle d'expression, 35](#)
- [Tâche 3. Éditer les règles d'expression, 35](#)
- [Création de règles d'expression - Résumé, 36](#)

Création de règles d'expression - Présentation

Les règles d'expression utilisent des fonctions d'expression et des colonnes source pour définir la logique de la règle. Vous pouvez créer des règles d'expression et les ajouter à un profil dans l'outil Analyst. Une règle d'expression peut être associée à un ou plusieurs profils.

La sortie d'une règle d'expression est une colonne virtuelle dans le profil. L'outil Analyst effectue le profil de la colonne virtuelle lorsque vous exécutez le profil.

Vous pouvez utiliser des règles d'expression pour valider les colonnes ou créer des colonnes source supplémentaires basées sur la valeur des colonnes source.

Scénario

HypoStores souhaite intégrer les données de ses nouveaux bureaux de Los Angeles dans son entrepôt de données. HypoStores souhaite analyser les noms des clients et les séparer en prénom et nom de famille. HypoStores souhaite utiliser des règles d'expression pour analyser une colonne qui contient les prénoms et noms de famille en deux colonnes virtuelles séparées puis effectuer leur profil. HypoStores souhaite aussi rendre les règles disponibles à d'autres analystes qui ont besoin d'analyser la sortie de ces règles.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

1. Créer des règles d'expression pour séparer la colonne FullName en colonnes prénom et nom de famille. Vous créez une règle qui sépare le prénom du nom de famille. Vous créez une autre règle qui sépare le nom de famille du prénom. Vous créez ces règles pour le profil Profile_LA_Customers_Custom.
2. Exécuter le profil et afficher la sortie des règles dans le profil.

3. Éditer les règles pour les rendre inutilisables pour d'autres utilisateurs de l'outil Analyst.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 à 4.

Timing

Prévoyez 10 à 15 minutes pour cette leçon.

Tâche 1. Créer des règles d'expression et exécuter le profil

Dans cette tâche, vous créez deux règles d'expression pour analyser la colonne FullName en deux colonnes virtuelles séparées nommées FirstName et LastName. Les noms des règles sont FirstName et LastName.

1. Vérifiez que vous êtes dans l'espace de travail **Découverte**.
Vous pouvez voir les résultats de profil du profil **Profile_LA_Customers_Custom**.
2. Dans le menu **Actions**, sélectionnez **Modifier > Règles de profilage de colonnes**.
La boîte de dialogue **Éditer le profil** s'affiche.
3. Cliquez sur **+** pour créer une règle.
4. Sélectionnez **Créer une règle**.
5. Cliquez sur **Suivant**.
6. Dans le champ **Nom**, entrez **FirstName**.
7. Dans la section **Expression**, entrez l'expression suivante pour séparer le prénom de la colonne **Nom** :

```
SUBSTR(FullName,1, INSTR(FullName, ' ', -1, 1) - 1)
```
8. Cliquez sur **Valider**.
9. Cliquez sur **OK**.
10. Cliquez sur **OK** pour revenir au panneau **Règles de profilage de colonnes**.
11. Cliquez sur **Enregistrer**.
L'outil Analyst crée la règle.
12. Faites défiler vers le bas du panneau **Profilage de colonne** pour afficher les détails de règle.
13. Répétez les étapes précédentes pour créer une règle nommée **LastName** et entrez l'expression suivante pour séparer le nom de famille de la colonne **Nom** :

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

Tâche 2. Afficher la sortie de la règle d'expression

Dans cette tâche, vous affichez la sortie des règles d'expression qui séparent les prénoms et noms de famille après avoir exécuté un profil.

1. Dans le menu **Actions**, sélectionnez **Exécuter le profil**.
L'outil Analyst exécute le profil et affiche les résultats de profil.
2. Dans le menu **Actions**, sélectionnez **Profilage de colonne**.
3. Dans la section **Colonnes**, cochez la case en regard de **Nom** dans la barre d'outils pour effacer toutes les colonnes.
Vous pouvez voir que l'une des colonnes est sélectionnée par défaut, car vous devez sélectionner au moins une colonne dans la section **Colonnes**.
4. Sélectionnez la colonne **FullName** et les règles **FirstName** et **LastName**.
5. Effacez toute autre colonne que vous voyez sélectionnée.
6. Cliquez sur **Enregistrer et exécuter**.
7. Dans les résultats de profil, cliquez sur la règle **FirstName**.
Les valeurs s'affichent dans l'onglet **Valeurs** dans le volet de droite.
8. Sélectionnez une valeur dans l'onglet **Valeurs**.
9. Cliquez avec le bouton droit et sélectionnez **Développement**.
Les valeurs de la colonne FullName et les règles FirstName et LastName s'affichent dans le panneau **Développement** avec d'autres valeurs de colonne. Notez que l'outil Analyst a séparé la colonne FullName en prénom et nom.

Tâche 3. Éditer les règles d'expression

Dans cette tâche, vous rendez les règles d'expression réutilisables et disponibles à tous les utilisateurs de l'outil Analyst.

1. Dans le panneau **Profilage de colonne**, vérifiez que la règle **FirstName** est sélectionnée.
2. Dans le menu **Actions**, sélectionnez **Modifier > Règles de profilage de colonnes**.
La boîte de dialogue **Modifier le profil** s'affiche.
3. Vérifiez que la règle est sélectionnée.
La règle d'informations s'affiche dans le panneau inférieur.
4. Cliquez sur l'icône **Modifier**.
5. Sélectionnez **Enregistrer comme règle réutilisable dans**.
Par défaut, l'outil Analyst enregistre la règle dans le profil et dossier en cours.
6. Cliquez sur **OK**.
7. Sélectionnez la règle **LastName** et répétez les étapes 3 à 6.
8. Cliquez sur **Enregistrer**.
Tout utilisateur de l'outil Analyst peut utiliser les règles FirstName et LastName pour partager une colonne avec prénoms et noms de famille dans des colonnes séparées.

Création de règles d'expression - Résumé

Dans cette leçon, vous avez appris que les règles d'expression utilisent des fonctions d'expression et des colonnes source pour définir la logique de la règle. Vous avez appris que la sortie d'une règle d'expression est une colonne virtuelle dans le profil. L'outil Analyst inclut la colonne virtuelle lorsque vous exécutez le profil.

Vous avez créé deux règles d'expression, les avez ajoutées à un profil, et avez exécuté le profil. Vous avez affiché la sortie des règles et les avez mises à disposition de tous les utilisateurs de l'outil Analyst.

CHAPITRE 7

Leçon 6. Création et exécution de fiches d'évaluation

Ce chapitre comprend les rubriques suivantes :

- [Création et de exécution de fiches d'évaluation - Présentation, 37](#)
- [Tâche 1. Créer une fiche d'évaluation à partir des résultats de profil, 38](#)
- [Tâche 2. Exécuter la fiche d'évaluation, 39](#)
- [Tâche 3. Afficher la fiche d'évaluation, 39](#)
- [Tâche 4. Éditer la fiche d'évaluation, 40](#)
- [Tâche 5. Configurer les seuils, 40](#)
- [Tâche 6. Afficher les graphes de tendance de score, 40](#)
- [Création et de exécution de fiches d'évaluation - Résumé, 41](#)

Création et de exécution de fiches d'évaluation - Présentation

Une fiche d'évaluation est la représentation graphique des valeurs valides d'une colonne ou la sortie d'une règle dans les résultats de profil. Utilisez les fiches d'évaluation pour mesurer et contrôler la progression de la qualité des données dans le temps.

Pour créer une fiche d'évaluation, ajoutez-lui depuis le profil des colonnes de mesures, affectez les poids aux mesures et configurez les seuils de scores. Pour exécuter une fiche d'évaluation, sélectionnez les valeurs valides de la mesure et exécutez la fiche d'évaluation pour afficher les scores des mesures.

Les fiches d'évaluation affichent la fréquence des valeurs dans les colonnes d'un profil sous forme de score. Les scores reflètent le pourcentage de valeurs valides d'une mesure.

Scénario

HypoStores souhaite intégrer les données de ses nouveaux bureaux de Los Angeles dans son entrepôt de données. Avant qu'ils ne fusionnent les données, ils veulent s'assurer que les données de différents niveaux et statuts de clients soient analysées à des fins de qualité. Vous êtes l'analyste responsable du contrôle de la progression et des analyses de la qualité des données. Vous voulez créer une fiche d'évaluation à partir des colonnes niveau de client et état du profil, configurer les seuils pour la qualité des données, et afficher les graphes de tendance de score pour déterminer l'amélioration des scores dans le temps.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

1. Créer une fiche d'évaluation depuis les résultats du profil Profile_LA_Customers_Custom pour afficher les scores des colonnes CustomerTier et Etat.
2. Exécuter la fiche d'évaluation pour générer les scores pour les colonnes CustomerTier et Etat.
3. Afficher la fiche d'évaluation pour voir les scores de chaque colonne.
4. Modifier la fiche d'évaluation pour spécifier différentes valeurs valides pour les scores.
5. Configurer les seuils de score et exécuter la fiche d'évaluation.
6. Afficher les graphes de tendance de score pour déterminer l'amélioration des scores dans le temps.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 à 5 de ce tutoriel.

Timing

Réservez 15 minutes pour effectuer les tâches de cette leçon.

Tâche 1. Créer une fiche d'évaluation à partir des résultats de profil

Dans cette tâche, vous créez une fiche d'évaluation à partir du profil Profile_LA_Customers_Custom pour évaluer les valeurs de colonne CustomerTier et État.

1. Vérifiez que vous êtes dans l'espace de travail **Découverte**. Vous pouvez voir le profil **Profile_LA_Customers_Custom** dans l'espace de travail.
2. Dans le menu **Actions**, sélectionnez **Ajouter à > Fiche d'évaluation**.
L'assistant **Ajouter à la fiche d'évaluation** s'ouvre. L'option **Nouvelle fiche d'évaluation** est sélectionnée par défaut.
3. Cliquez sur **Suivant**.
4. Saisissez sc_LA_Customer comme nom de fiche d'évaluation et accédez au dossier Clients pour l'emplacement de la fiche d'évaluation.
5. Cliquez sur **Suivant**.
6. Sélectionnez les colonnes **CustomerTier** et **État** pour les ajouter à la fiche d'évaluation.
7. Cliquez sur **Suivant**.
8. Vérifiez que dans la section **Mesures**, la mesure CustomerTier est sélectionnée.
9. Dans la section **Score utilisant : des valeurs**, sélectionnez toutes les valeurs et cliquez sur le bouton **Ajouter tout** pour déplacer les valeurs vers la section **Valeurs valides**.
Utilisez la touche **Maj** pour sélectionner plusieurs valeurs.
10. Dans la section **Mesures**, sélectionnez la mesure **État** et sélectionnez les valeurs qui ont des codes d'état à deux lettres dans la section **Score utilisant : des valeurs**.
11. Cliquez sur le bouton **Ajouter** pour déplacer les valeurs dans la section **Valeurs valides**.

Vous pouvez voir le nombre total de valeurs valides et la valeur valide du pourcentage en haut de la section.

12. Pour chaque mesure de la section **Mesures**, acceptez les paramètres par défaut des seuils de scores de la section **Seuils des mesures**.
13. Cliquez sur **Suivant**.
14. Éventuellement, sélectionnez un groupe de mesures auquel ajouter les mesures. Par défaut, l'outil Analyst ajoute les mesures au groupe de mesures **Par défaut**.
15. Cliquez sur **Suivant**.
16. Dans la section **Mesures**, double-cliquez sur la colonne **Poids** pour la mesure **CustomerTier**.
Lorsque vous exécutez une fiche d'évaluation, l'outil Analyst calcule la moyenne pondérée de chaque groupe de mesures en fonction du score de mesure et du poids que vous avez affecté à chaque mesure.
17. Entrez un poids pour les mesures CustomerTier et État.
18. Cliquez sur **Enregistrer** pour créer la fiche d'évaluation.
La fiche d'évaluation s'affiche dans l'espace de travail **Fiches d'évaluation**.

Tâche 2. Exécuter la fiche d'évaluation

Dans cette tâche, vous exécutez la fiche d'évaluation sc_LA_Customer afin de générer les scores pour les colonnes CustomerTier et Etat.

1. Vérifiez que vous êtes dans l'espace de travail **Fiches d'évaluation**. Vous pouvez voir la fiche d'évaluation sc_LA_Customer.
2. Cliquez sur **Actions > Exécuter la fiche d'évaluation**.
La boîte de dialogue **Exécuter la fiche d'évaluation** s'affiche.
3. Vérifiez les paramètres dans la boîte de dialogue et cliquez sur **Exécuter**.
L'espace de travail **Fiches d'évaluation** affiche les scores pour les colonnes CustomerTier et État.

Tâche 3. Afficher la fiche d'évaluation

Dans cette tâche, vous voyez la fiche d'évaluation sc_LA_Customer pour afficher les scores des colonnes CustomerTier et Etat.

1. Sélectionnez la colonne **État** qui contient le score État que vous voulez afficher.
Dans la section **Mesures**, vous pouvez afficher le nom de la mesure, le nombre total de lignes, le nombre de lignes qui ne sont pas valides et le pourcentage du score. Vous pouvez voir le score affiché sous la forme d'un graphique à barres, la tendance du score, l'objet de données du score, la source et le type de source du score.
2. Dans le menu **Actions**, sélectionnez **Développement**.
Les scores qui ne sont pas valides pour la colonne État s'affichent dans la section **Lignes non valides**.
3. Sélectionnez **Valider les lignes** pour afficher les scores qui sont valides pour la colonne État.
4. Répétez les étapes précédentes pour la colonne CustomerTier.
Tous les scores de la colonne CustomerTier sont valides.

Tâche 4. Éditer la fiche d'évaluation

Dans cette tâche, vous allez modifier la fiche d'évaluation sc_LA_Customer pour indiquer que la valeur Ruby n'est pas valide pour le score CustomerTier.

1. Dans le menu **Actions**, sélectionnez **Modifier > Mesures**.
La boîte de dialogue **Modifier la fiche d'évaluation** s'ouvre.
2. Dans la section **Mesures**, sélectionnez **CustomerTier**.
3. Dans la section **Score utilisant : des valeurs**, déplacez **Ruby** de la section **Valeurs valides** vers la section **Valeurs disponibles**.
Acceptez les paramètres par défaut dans la section **Seuils des mesures**.
4. Cliquez sur **Enregistrer et exécuter** pour enregistrer les modifications apportées à la fiche d'évaluation et l'exécuter.
5. Affichez à nouveau le score de CustomerTier.
Le score de CustomerTier passe à un pourcentage de 81,4.

Tâche 5. Configurer les seuils

Dans cette tâche, vous configurez les seuils du score Etat dans la fiche d'évaluation sc_LA_Customer afin de déterminer les plages acceptables pour les données de la colonne Etat. Les valeurs avec un code à deux lettres, comme par exemple CA, sont acceptables, et les codes avec plus de deux lettres, comme par exemple Calif, ne sont pas acceptables.

1. Dans le menu **Actions**, sélectionnez **Modifier > Mesures**.
La boîte de dialogue **Modifier la fiche d'évaluation** s'ouvre.
2. Dans la section **Mesures**, sélectionnez **État**.
3. Dans la section **Seuils des mesures**, entrez les plages suivantes pour les scores Satisfaisant et Inacceptable : 90 à 100 % Satisfaisant ; 0 à 50 % Inacceptable ; 51% à 89 % Acceptable.
Les seuils représentent les limites inférieures des plages Bon et Acceptable.
4. Cliquez sur **Enregistrer et exécuter** pour enregistrer les modifications apportées à la fiche d'évaluation et l'exécuter.
Dans le panneau **Fiche d'évaluation**, affichez les modifications du pourcentage et le score affiché en barre pour le score État.

Tâche 6. Afficher les graphes de tendance de score

Dans cette tâche, vous affichez le graphe de tendance pour le score Etat. Vous pouvez afficher les graphiques de tendance pour visualiser les scores dans le temps.

1. Dans le panneau **Fiche d'évaluation**, sélectionnez **État**.
2. Dans le menu **Actions**, sélectionnez **Afficher le graphe de tendance** ou cliquez sur la flèche dans la colonne **Tendance de score**.

La boîte de dialogue **Détail de graphe de tendance**. Vous pouvez voir les seuils **Bon**, **Acceptable**, et **Inacceptable** pour le score. Les seuils changent à chaque fois que vous exécutez la fiche d'évaluation après avoir modifié les valeurs des scores dans la fiche d'évaluation.

3. Pointez vers un cercle quelconque dans le graphe pour afficher les valeurs valides dans la section **Valeurs valides** en bas du graphe.
4. Cliquez sur **Fermer** pour revenir à la fiche d'évaluation.

Création et de exécution de fiches d'évaluation - Résumé

Dans cette leçon, vous avez appris que vous pouvez créer une fiche d'évaluation à partir des résultats d'un profil. Une fiche d'évaluation contient les colonnes d'un profil. Vous avez appris que vous pouvez exécuter une fiche d'évaluation pour générer des scores pour les colonnes. Vous avez édité une fiche d'évaluation pour configurer les valeurs valides et définir les seuils des scores. Vous avez également appris comment afficher le graphe de tendance de score.

Vous avez créé une fiche d'évaluation à partir des colonnes CustomerTier et Etat d'un profil pour analyser la qualité de leurs données. Vous avez exécuté la fiche d'évaluation pour générer des scores de chaque colonne. Vous avez modifié la fiche d'évaluation pour spécifier différentes valeurs valides pour les scores. Vous avez configuré les seuils pour un score et affiché le graphe de tendance de score.

CHAPITRE 8

Leçon 7. Création de tables de référence à partir des colonnes de profil

Ce chapitre comprend les rubriques suivantes :

- [Création de tables de référence à partir des colonnes de profil - Présentation, 42](#)
- [Tâche 1. Créer une table de référence à partir des colonnes de profil, 43](#)
- [Tâche 2. Éditer la table de référence, 44](#)
- [Création de tables de référence à partir des colonnes de profil - Résumé, 44](#)

Création de tables de référence à partir des colonnes de profil - Présentation

Une table de référence contient des données de référence que vous pouvez utiliser pour normaliser les données source. Les données de référence peuvent inclure des valeurs valides et standard. Créez des tables de référence pour établir des relations entre des valeurs de données source et les valeurs valides et standard.

Vous pouvez créer une table de référence à partir des résultats d'un profil. Après avoir créé une table de référence, vous pouvez l'éditer afin d'ajouter des colonnes ou des lignes et ajouter ou éditer les valeurs standard et valides. Vous pouvez afficher les modifications d'une table de référence dans un suivi d'audit.

Scénario

HypoStores souhaite effectuer le profil des données pour détecter les anomalies et normaliser les données avec des valeurs valides. Vous êtes l'analyste responsable de la normalisation des valeurs valides dans les données. Vous voulez créer une table de référence basée sur les valeurs valides des colonnes du profil.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

1. Créer une table de référence depuis la colonne CustomerTier du profil Profile_LA_Customers_Custom en sélectionnant les valeurs valides pour les colonnes.
2. Éditer la table de référence pour configurer différentes valeurs valides pour les colonnes.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 à 6 de ce tutoriel.

Timing

Réservez 15 minutes pour effectuer les tâches de cette leçon.

Tâche 1. Créer une table de référence à partir des colonnes de profil

Dans cette tâche, vous créez une table de référence et ajoutez la colonne CustomerTier depuis le profil Profile_LA_Customers_Custom vers la table de référence.

1. Vérifiez que vous êtes dans l'espace de travail **Découverte**.
Vous pouvez voir le profil Profile_LA_Customers_Custom dans l'espace de travail.
2. Dans le panneau **Profilage de colonne**, sélectionnez la colonne CustomerTier que vous voulez ajouter à une table de référence.
Vous pouvez développer les fréquences des valeurs et formes pour la colonne CustomerTier afin de vérifier les enregistrements qui ont des valeurs de catégorie client non standard.
3. Dans l'onglet **Valeurs** du volet de droite, sélectionnez les valeurs de niveau client valides que vous voulez ajouter. Utilisez les touches CTRL ou MAJ pour sélectionner les valeurs suivantes : Diamond, Gold, Silver, Bronze, Emerald.
4. Dans le menu **Actions**, sélectionnez **Ajouter à > Table de référence**.
5. Sélectionnez **Créer une table de référence**.
6. Cliquez sur **Suivant**.
7. Dans le champ **Nom de la table**, entrez **Reftab_CustTier_HypoStores**.
8. Saisissez une description et définissez 0 comme valeur par défaut.
L'outil Analyst utilise la valeur par défaut pour tout enregistrement de la table qui ne contient pas de valeur.
9. Cliquez sur **Suivant**.
10. Dans le panneau **Attributs des colonnes**, configurez les propriétés de colonne suivantes pour la colonne CustomerTier :

Propriété	Description
Nom	CustomerTier
Type de données	String
Précision	10
Description	Valeurs de niveau client de référence

11. Éventuellement, choisissez de créer une colonne de description pour les lignes de la table de référence. Saisissez le nom et la précision pour la colonne.
12. Vérifiez les valeurs de la colonne CustomerTier dans la section **Prévisualisation**.
13. Cliquez sur **Suivant**.
La nom de la table de référence Reftab_CustomerTier_HypoStores s'affiche. Vous pouvez entrer une description facultative.
14. Dans le panneau **Enregistrer dans**, sélectionnez le projet de tutoriel dans lequel vous voulez créer la table de référence.
Le panneau **Tables de référence** : répertorie les tables de référence dans l'emplacement que vous sélectionnez.
15. Entrez une note d'audit facultative.
16. Cliquez sur **Terminer**.

Tâche 2. Éditer la table de référence

Dans cette tâche, vous éditez la table Reftab_CustomerTier_HypoStores afin d'ajouter des valeurs alternatives pour les niveaux clients.

1. Dans le navigateur **Bibliothèque**, recherchez la table de référence Reftab_CustomerTier_HypoStores.
2. Cliquez sur la table de référence Reftab_CustomerTier_HypoStores.
La table de référence s'ouvre dans l'espace de travail **Conception**.
3. Pour éditer une ligne, sélectionnez-la et cliquez sur **Actions > Modifier** ou cliquez sur l'icône **Modifier**.
Le boîte de dialogue **Modifier la ligne** s'affiche. (Facultatif) Sélectionnez plusieurs lignes pour ajouter la même valeur alternative à chaque ligne.
4. Saisissez les valeurs alternatives suivantes pour les lignes Diamond, Emerald, Gold, Silver et Bronze : 1, 2, 3, 4, 5.
Saisissez une note d'audit facultative.
5. Cliquez sur **Appliquer** pour appliquer les changements.
6. Cliquez sur **Fermer**.
Les valeurs de la table de référence modifiées s'affichent dans l'espace de travail **Conception**.

Création de tables de référence à partir des colonnes de profil - Résumé

Dans cette leçon, vous avez appris à créer des tables de référence à partir des résultats d'un profil afin de configurer des valeurs valides pour les données source.

Vous avez créé une table de référence à partir d'une colonne de profil en sélectionnant les valeurs valides pour les colonnes. Vous avez modifié la table de référence afin de configurer différentes valeurs valides pour les colonnes.

CHAPITRE 9

Leçon 8. Création de tables de référence

Ce chapitre comprend les rubriques suivantes :

- [Création de tables de référence - Présentation, 45](#)
- [Tâche 1. Créer une table de référence, 46](#)
- [Création de tables de référence - Résumé, 46](#)

Création de tables de référence - Présentation

Une table de référence contient des données de référence que vous pouvez utiliser pour normaliser les données source. Les données de référence peuvent inclure des valeurs valides et standard. Créez des tables de référence pour établir des relations entre les valeurs de données source et les valeurs standard et valides.

Vous pouvez créer manuellement une table de référence en utilisant l'éditeur de table de référence. Utilisez la table de référence pour définir et normaliser les données source. Vous pouvez partager la table de référence avec un développeur pour utilisation dans les transformations Normalisation et Rechercher de l'outil Developer.

Scénario

HypoStores souhaite normaliser les données avec des valeurs valides. Vous êtes l'analyste responsable de la normalisation des valeurs valides dans les données. Vous voulez créer une table de référence pour définir des codes de niveau client standard qui référencent les données clients de Los Angeles. Vous pouvez ensuite partager la table de référence avec un développeur.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Créer une table de référence en utilisant l'éditeur de table de référence afin de définir les codes de niveau client standard qui référencent les données clients de Los Angeles.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 et 2 de ce tutoriel.

Timing

Réservez 10 minutes pour effectuer les tâches de cette leçon.

Tâche 1. Créer une table de référence

Dans cette tâche, vous allez créer la table de référence Reftab_CustomerTier_Codes pour normaliser les valeurs valides pour les données de niveau client.

1. Cliquez sur **Nouveau > Table de référence**.
L'assistant **Nouvelle table de référence** s'affiche.
2. Sélectionnez **Utilisez l'éditeur de table de référence**.
3. Cliquez sur **Suivant**.
4. Pour chaque colonne que vous voulez inclure dans la table de référence, cliquez sur l'icône **Ajouter une colonne** et configurez les propriétés de colonne pour chaque colonne.
Ajoutez les noms de colonne suivants : CustomerID, CustomerTier, et Statut. Vous pouvez réorganiser ou effacer des colonnes.
5. Entrez une description facultative et définissez la valeur par défaut à **0**.
L'outil Analyst utilise la valeur par défaut pour tout enregistrement de la table qui ne contient pas de valeur.
6. Cliquez sur **Suivant**.
7. Dans le champ **Nom**, entrez **Reftab_CustomerTier_Codes**.
8. Dans la section **Dossiers**, sélectionnez le dossier **Clients** dans le projet Tutoriel.
9. Cliquez sur **Terminer**.
La table de référence s'affiche dans l'espace de travail **Conception**.
10. Dans le menu **Actions**, sélectionnez **Ajouter une ligne** pour remplir chaque colonne de table de référence avec les quatre valeurs suivantes :
CustomerID = LA1, LA2, LA3, LA4
CustomerTier = 1, 2, 3, 4.
Statut = Actif, Inactif

Création de tables de référence - Résumé

Dans cette leçon, vous avez appris comment créer des tables de référence à l'aide de l'éditeur de table de référence afin de créer des valeurs standard valides à utiliser avec les données source.

Vous avez créé une table de référence à l'aide de l'éditeur de table de référence afin de normaliser les valeurs de niveau clients pour les données clients de Los Angeles.

Partie II : Débuter avec Informatica Developer

Cette partie contient les chapitres suivants :

- [Leçon 1. Configuration de Informatica Developer, 48](#)
- [Leçon 2 : Importation d'objets de données physiques, 52](#)
- [Leçon 3. Profilage des données, 56](#)
- [Leçon 4. Analyse des données, 61](#)
- [Leçon 5. Normalisation des données , 68](#)
- [Leçon 6. Validation des données d'adresse, 74](#)

CHAPITRE 10

Leçon 1. Configuration de Informatica Developer

Ce chapitre comprend les rubriques suivantes :

- [Configuration de Informatica Developer - Présentation, 48](#)
- [Tâche 1. Démarrer Informatica Developer, 49](#)
- [Tâche 2. Ajouter un domaine, 49](#)
- [Tâche 3. Ajouter un référentiel modèle, 50](#)
- [Tâche 4. Créer un projet, 50](#)
- [Tâche 5. Créer un dossier, 50](#)
- [Tâche 6. Sélectionner un service Data Integration Service par défaut., 51](#)
- [Configuration de Informatica Developer - Résumé, 51](#)

Configuration de Informatica Developer - Présentation

Avant de commencer les leçons de ce tutoriel, vous devez démarrer et configurer l'outil Developer. Pour configurer l'outil Developer, vous ajoutez un domaine. Vous ajoutez un référentiel modèle qui est dans le domaine et vous créez un projet et un dossier pour enregistrer votre travail. Vous sélectionnez également un service Data Integration Service par défaut.

Le domaine Informatica est une collection de nœuds et de services qui définissent l'environnement Informatica. Les services du domaine incluent Model Repository Service et Data Integration Service.

Model Repository Service gère le référentiel modèle. Le référentiel modèle est une base de données relationnelles qui stocke les métadonnées des projets que vous créez dans l'outil Developer. Un projet stocke des objets que vous créez dans l'outil Developer. Un projet peut aussi contenir des dossiers qui stockent les objets apparentés, tels que les objets qui font partie des mêmes spécifications d'entreprise.

Data Integration Service effectue les tâches d'intégration des données dans l'outil Developer.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Démarrer l'outil Developer et accéder à l'espace de travail de l'outil Developer.

- Ajouter un domaine dans l'outil Developer.
- Ajouter un référentiel modèle de façon à pouvoir créer un projet.
- Créer un projet pour stocker les objets définis dans l'outil Developer.
- Créer un dossier en mesure de stocker des objets apparentés dans le projet.
- Sélectionner un service Data Integration Service par défaut pour exécuter les tâches d'intégration de données.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez installé l'outil Developer.
- Vous disposez d'un nom de domaine, d'un nom d'hôte et d'un numéro de port pour la connexion au domaine. Vous pouvez obtenir ces informations auprès de l'administrateur du domaine.
- Un administrateur de domaine a configuré un service de référentiel modèle dans l'outil Administrator.
- Vous disposez d'un nom d'utilisateur et d'un mot de passe pour accéder à Model Repository Service. Vous pouvez obtenir ces informations auprès de l'administrateur de domaine.
- Un administrateur de domaine a configuré un service Data Integration Service.
- Data Integration Service est en cours d'exécution.

Timing

Réservez 5 à 10 minutes pour effectuer les tâches de cette leçon.

Tâche 1. Démarrer Informatica Developer

Démarrez l'outil Developer pour commencer le tutoriel.

1. Démarrez l'outil Developer.
La page **Accueil** de l'outil Developer apparaît.
2. Cliquez sur le bouton **Espace de travail**.
L'espace de travail de l'outil Developer s'affiche.

Tâche 2. Ajouter un domaine

Dans cette tâche, vous allez ajouter un domaine dans l'outil Developer pour accéder à un référentiel modèle.

1. Cliquez sur **Fenêtre > Préférences**.
La boîte de dialogue **Préférences** s'affiche.
2. Sélectionnez **Informatica > Domaines**.
3. Cliquez sur **Ajouter**.
La boîte de dialogue **Nouveau domaine** s'affiche.
4. Entrez le nom du domaine, le nom d'hôte et le numéro de port.
5. Cliquez sur **Terminer**.

6. Cliquez sur **OK**.

Tâche 3. Ajouter un référentiel modèle

Dans cette tâche, vous allez ajouter le référentiel modèle que vous voulez utiliser pour stocker les projets et les dossiers.

1. Cliquez sur **Fichier > Connexion au référentiel**.
La boîte de dialogue **Connexion au référentiel** s'ouvre.
2. Cliquez sur **Parcourir** pour sélectionner un service Model Repository Service.
3. Cliquez sur **OK**.
4. Cliquez sur **Suivant**.
5. Entrez votre nom d'utilisateur et votre mot de passe.
6. Cliquez sur **Terminer**.

Le référentiel modèle apparaît dans la vue **Explorateur d'objets**.

Tâche 4. Créer un projet

Dans cette tâche, vous allez créer un projet pour stocker des objets définis dans l'outil Developer. Vous pouvez créer un projet pour tous les tutoriels de ce guide.

1. Dans la vue **Explorateur d'objets**, sélectionnez un service Model Repository Service.
2. Cliquez sur **Fichier > Nouveau > Projet**.
La boîte de dialogue **Nouveau projet** s'affiche.
3. Entrez votre nom avec le préfixe « Tutoriel_ » comme nom du projet.
4. Cliquez sur **Terminer**.

Le projet s'affiche sous le service Model Repository Service dans la vue **Explorateur d'objets**.

Tâche 5. Créer un dossier

Dans cette tâche, vous allez créer un dossier pour stocker les objets apparentés. Vous pouvez créer un dossier pour tous les tutoriels de ce guide.

1. Dans la vue **Explorateur d'objets**, sélectionnez le projet auquel vous voulez ajouter le dossier.
2. Cliquez sur **Fichier > Nouveau > Dossier**.
3. Entrez le nom du dossier.
4. Cliquez sur **Terminer**.

L'outil Developer ajoute le dossier sous le projet dans la vue **Explorateur d'objets**. Développez le projet pour voir le dossier.

Tâche 6. Sélectionner un service Data Integration Service par défaut.

Dans cette tâche, vous allez sélectionner un service Data Integration Service par défaut pour pouvoir exécuter les mappages et prévisualiser les données.

1. Cliquez sur **Fenêtre > Préférences**.
La boîte de dialogue **Préférences** s'affiche.
2. Sélectionnez **Informatica > Data Integration Service**.
3. Développez le domaine.
4. Sélectionnez un service Data Integration Service.
5. Cliquez sur **Définir par défaut**.
6. Cliquez sur **OK**.

Configuration de Informatica Developer - Résumé

Dans cette leçon, vous avez appris que le domaine Informatica inclut Model Repository Service et Data Integration Service. Model Repository Service gère le référentiel modèle. Un référentiel modèle contient des projets et des dossiers. Data Integration Service effectue les tâches d'intégration de données.

Vous avez démarré l'outil Developer et vous l'avez configuré. Vous avez ajouté un domaine à l'outil Developer, ajouté un référentiel modèle et créé un projet et un dossier. Vous avez également sélectionné un service Data Integration Service par défaut.

Maintenant, vous pouvez utiliser l'outil Developer pour suivre les autres leçons de ce tutoriel.

CHAPITRE 11

Leçon 2 : Importation d'objets de données physiques

Ce chapitre comprend les rubriques suivantes :

- [Importation d'objets de données physiques - Présentation, 52](#)
- [Tâche 1. Importer l'objet de données de fichier plat Boston_Customers, 53](#)
- [Tâche 2. Importer l'objet de données de fichier plat LA_Customers, 53](#)
- [Tâche 3. Importer l'objet de données de fichier plat All_Customers, 54](#)
- [Importation d'objets de données physiques - Résumé, 55](#)

Importation d'objets de données physiques - Présentation

Un objet de données physique est une représentation des données d'un fichier plat ou d'une table de base de données relationnelles. Vous pouvez importer un fichier plat ou une table de base de données relationnelles en tant qu'objet de données physique à utiliser comme source ou cible lors d'un mappage.

Scénario

HypoStores Corporation stocke les données clients des bureaux de Los Angeles et de Boston dans des fichiers plats. Vous voulez travailler avec ces données clients dans l'outil Developer. Pour ce faire, vous devez importer chaque fichier plat en tant qu'objet de données physique.

Objectifs

Dans cette leçon, vous importez des fichiers plats en tant qu'objets de données physiques. Vous définissez aussi le répertoire de fichiers source pour que Data Integration Service puisse lire les données source dans le répertoire approprié.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez effectué la leçon 1 de ce tutoriel.

Timing

Réservez 10 à 15 minutes pour effectuer les tâches de cette leçon.

Tâche 1. Importer l'objet de données de fichier plat Boston_Customers

Dans cette tâche, vous importez un objet de données physique à partir d'un fichier qui contient les données clients du bureau de Boston.

1. Dans la vue **Explorateur d'objets**, sélectionnez le projet de tutoriel.
2. Cliquez sur **Fichier > Nouveau > Objet de données**.
La boîte de dialogue **Nouveau** s'affiche.
3. Sélectionnez **Objets de données physiques > Objet de données de fichier plat** et cliquez sur **Suivant**.
La boîte de dialogue **Nouvel objet de données de fichier plat** s'ouvre.
4. Sélectionnez **Créer à partir d'un fichier plat existant**.
5. Cliquez sur **Parcourir** et allez à Boston_Customer.csv dans le répertoire suivant : <Répertoire d'installation Informatique>\clients\DeveloperClient\Tutorials
6. Cliquez sur **Ouvrir**.
L'assistant nomme l'objet de données « Boston_Customers ».
7. Cliquez sur **Suivant**.
8. Vérifiez que la page de code est MS Windows Latin 1 (ANSI), sur-ensemble de Latin 1.
9. Vérifiez que le format est délimité.
10. Cliquez sur **Suivant**.
11. Vérifiez que le séparateur est défini sur la virgule.
12. Sélectionnez **Importer les noms de colonne à partir de la première ligne**.
13. Cliquez sur **Terminer**.
L'objet de données physique Boston_Customers s'affiche dans les **Objets de données physiques** du projet de tutoriel.
14. Cliquez sur la vue **Lecture** et sélectionnez la transformation Sortie.
15. Cliquez sur l'onglet **Moteur d'exécution** dans la vue **Propriétés**.
16. Définissez le Répertoire du fichier source sur le répertoire suivant de la machine Data Integration Service : <Répertoire d'installation Informatique>\server\Tutorials
17. Cliquez sur **Fichier > Enregistrer**.

Tâche 2. Importer l'objet de données de fichier plat LA_Customers

Dans cette tâche, vous allez importer un objet de données physique à partir d'un fichier qui contient les données clients du bureau de Los Angeles.

1. Dans la vue **Explorateur d'objets**, sélectionnez le projet de tutoriel.
2. Cliquez sur **Fichier > Nouveau > Objet de données**.
La boîte de dialogue **Nouveau** s'affiche.
3. Sélectionnez **Objets de données physiques > Objet de données de fichier plat** et cliquez sur **Suivant**.

- La boîte de dialogue **Nouvel objet de données de fichier plat** s'ouvre.
4. Sélectionnez **Créer à partir d'un fichier plat existant**.
 5. Cliquez sur **Parcourir** et accédez à LA_Customer.csv dans le répertoire suivant : <Répertoire d'installation Informatica>\clients\DeveloperClient\Tutorials
 6. Cliquez sur **Ouvrir**.
L'assistant nomme l'objet de données « LA_Customers ».
 7. Cliquez sur **Suivant**.
 8. Vérifiez que la page de code est MS Windows Latin 1 (ANSI), sur-ensemble de Latin 1.
 9. Vérifiez que le format est délimité.
 10. Cliquez sur **Suivant**.
 11. Vérifiez que le séparateur est défini sur la virgule.
 12. Sélectionnez **Importer les noms de colonne à partir de la première ligne**.
 13. Cliquez sur **Terminer**.
L'objet de données physique LA_Customers s'affiche dans les **Objets de données physiques** du projet de tutoriel.
 14. Cliquez sur la vue **Lecture** et sélectionnez la transformation Sortie.
 15. Cliquez sur l'onglet **Moteur d'exécution** dans la vue **Propriétés**.
 16. Définissez le Répertoire du fichier source sur le répertoire suivant de la machine Data Integration Service : <Répertoire d'installation Informatica>\server\Tutorials
 17. Cliquez sur **Fichier > Enregistrer**.

Tâche 3. Importer l'objet de données de fichier plat All_Customers

Dans cette tâche, vous importez un objet de données physique à partir d'un fichier plat qui combine les données de commandes des clients des bureaux de Los Angeles et de Boston.

1. Dans la vue **Explorateur d'objets**, sélectionnez le projet du tutoriel.
2. Cliquez sur **Fichier > Nouveau > Objet de données**.
La boîte de dialogue **Nouveau** s'affiche.
3. Sélectionnez **Objets de données physiques > Objet de données de fichier plat** et cliquez sur **Suivant**.
La boîte de dialogue **Nouvel objet de données de fichier plat** s'ouvre.
4. Sélectionnez **Créer à partir d'un fichier plat existant**.
5. Cliquez sur **Parcourir** et accédez à All_Customer.csv dans le répertoire suivant : <Répertoire d'installation Informatica>\clients\DeveloperClient\Tutorials
6. Cliquez sur **Ouvrir**.
L'assistant nomme l'objet de données « All_Customers ».
7. Cliquez sur **Suivant**.
8. Vérifiez que la page de code est MS Windows Latin 1 (ANSI), sur-ensemble de Latin 1.
9. Vérifiez que le format est délimité.

10. Cliquez sur **Suivant**.
11. Vérifiez que le séparateur est défini sur la virgule.
12. Sélectionnez **Importer les noms de colonne à partir de la première ligne**.
13. Cliquez sur **Terminer**.
L'objet de données physique All_Customers s'affiche dans les **Objets de données physiques** du projet de tutoriel.
14. Cliquez sur la vue **Lecture** et sélectionnez la transformation Sortie.
15. Cliquez sur l'onglet **Moteur d'exécution** dans la vue **Propriétés**.
16. Définissez le Répertoire du fichier source sur le répertoire suivant de Data Integration Service :
<Répertoire d'installation Informatica>\server\Tutorials
17. Cliquez sur **Fichier > Enregistrer**.

Importation d'objets de données physiques - Résumé

Dans cette leçon, vous avez appris que les objets de données physiques sont des représentations de données basées sur des fichiers plats ou sur une table de base de données relationnelles.

Vous avez créé des objets de données physiques à partir des fichiers plats. Vous avez également défini le répertoire de fichiers source pour que Data Integration Service puisse lire les données source dans le répertoire approprié.

Vous utilisez les objets de données comme sources de mappage dans les leçons de qualité des données.

CHAPITRE 12

Leçon 3. Profilage des données

Ce chapitre comprend les rubriques suivantes :

- [Profilage des données - Présentation, 56](#)
- [Tâche 1. Effectuer une analyse de jointure sur deux sources de données, 57](#)
- [Tâche 2. Afficher les résultats de l'analyse de jointure, 58](#)
- [Tâche 3. Exécuter un profil sur une source de données, 59](#)
- [Tâche 4. Afficher les résultats de profilage de colonne, 59](#)
- [Profilage des données Résumé, 60](#)

Profilage des données - Présentation

Un profil est un ensemble de métadonnées qui décrit le contenu et la structure d'un ensemble de données.

Le profilage de données constitue souvent la première étape d'un projet. Vous pouvez exécuter un profil pour évaluer la structure des données et vérifier que les colonnes de données sont remplies avec les types d'informations attendus. Si un profil révèle des problèmes dans les données, vous pouvez définir les étapes pour résoudre ces problèmes dans votre projet. Par exemple, si un profil révèle qu'une colonne contient des valeurs plus longues que prévu, vous pouvez concevoir des processus de qualité des données pour supprimer ou corriger les valeurs à problème.

Un profil qui analyse la qualité des données des colonnes sélectionnées est appelé un profil de colonne.

Remarque: Vous pouvez également utiliser l'outil Developer pour trouver la clé principale, la clé étrangère et les relations de dépendance fonctionnelle, ainsi que pour analyser les conditions de jointures sur les colonnes de données.

Un profil de colonne fournit les informations suivantes sur les données :

- Le nombre de valeurs uniques et null dans chaque colonne, exprimé sous forme de nombre et de pourcentage.
- Les formes des données de chaque colonne, et les fréquences auxquelles ces valeurs apparaissent.
- Les statistiques sur les valeurs de colonnes, comme par exemple les longueurs maximum et minimum des valeurs ainsi que la première et la dernière valeur dans chaque colonne.
- Pour les profils d'analyse de jointure, le degré de chevauchement entre les deux colonnes de données, affiché sous forme de diagramme Venn et de valeur de pourcentage. Utilisez les profils d'analyse de jointure pour identifier les problèmes éventuels avec les conditions de jointure de colonne.

Vous pouvez à tout moment exécuter un profil de colonne dans un projet afin de mesurer la qualité des données et de vérifier que les modifications des données correspondent à vos objectifs pour le projet. Vous

pouvez exécuter un profil de colonne sur une transformation dans un mappage pour indiquer l'effet que la transformation aura sur les données.

Scénario

HypoStores souhaite vérifier que les données clients ne contiennent pas d'erreurs, d'incohérences ou d'informations en double. Avant qu'HypoStores ne conçoive les processus pour atteindre les objectifs de qualité des données, elle doit mesurer la qualité de ses fichiers de données source et confirmer que les données sont prêtes à être traitées.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Effectuer une analyse de jointure sur la source de données `Boston_Customers` et la source de données `LA_Customers`.
- Afficher les résultats d'une analyse de jointure pour déterminer si vous pouvez fusionner les données des deux bureaux avec succès.
- Exécuter un profil de colonne sur la source de données `All_Customers`.
- Afficher les résultats de profilage de colonne pour respecter les valeurs et formes contenues dans les données.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 et 2 de ce tutoriel.

Temps requis

- Prévoyez 20 minutes pour cette leçon.

Tâche 1. Effectuer une analyse de jointure sur deux sources de données

Dans cette tâche, vous effectuez une analyse de jointure sur les sources de données `Boston_Customers` et `LA_Customers` pour afficher les conditions de jointure.

1. Sélectionnez le dossier du tutoriel et cliquez sur **Fichier > Nouveau > Profil**.
2. Sélectionnez **Profil de découverte d'entreprise**.
3. Cliquez sur **Suivant**.
4. Dans le champ **Nom**, saisissez **Tutorial_Profile**.
5. Cliquez sur **Terminer**.

Le profil `Tutorial_Profile` s'affiche dans l'explorateur d'objets.

6. Faites glisser les sources de données **`Boston_Customers`** et **`LA_Customers`** vers l'éditeur sur la droite.

Astuce: Maintenez la touche Maj enfoncée pour sélectionner plusieurs objets de données.

7. Cliquez avec le bouton droit de la souris sur un nom d'objet de données et sélectionnez **Profil de jointure**.

L'assistant **Nouveau profil de jointure** s'affiche.

8. Dans le champ **Nom**, saisissez **JoinAnalysis**.

9. Vérifiez que Boston_Customers et LA_Customers s'affichent comme des objets de données et cliquez sur **Suivant**.
10. Vérifiez que la colonne **CustomerID** est sélectionnée dans les deux sources de données. Faites défiler le volet de l'assistant pour afficher les colonnes dans les deux ensembles de données. Cliquez sur **Suivant**.
11. Cliquez sur **Ajouter** pour ajouter des conditions de jointure. La boîte de dialogue **Condition de jointure** s'affiche.
12. Dans la section **Colonnes**, cliquez sur **Ajouter une ligne**.
13. Double-cliquez sur la première ligne dans la colonne de gauche et sélectionnez **CustomerID**.
14. Double-cliquez sur la première ligne dans la colonne de droite et sélectionnez **CustomerID**.
15. Cliquez sur **OK**, puis cliquez sur **Terminer**.
16. Si les outils Developer vous invitent à enregistrer les modifications, cliquez sur **Oui**. L'outil Developer exécute le profil.

Remarque: Ne fermez pas le profil. Vous afficherez les résultats de profil dans la tâche suivante.

Tâche 2. Afficher les résultats de l'analyse de jointure

Dans cette tâche, vous affichez les résultats de l'analyse de jointure dans la vue Résultats de jointure du profil JoinAnalysis.

1. Cliquez sur l'onglet **JoinAnalysis** dans l'éditeur.
2. Dans la section **Résultats de jointure**, cliquez sur la première ligne. La section **Détails** affiche un diagramme de Venn et la clé de couleur qui détaille les résultats de l'analyse de jointure.
3. Vérifiez que la colonne **Lignes de jointure** affiche zéro pour le nombre de lignes qui contiennent une jointure. Cette valeur indique que les champs CustomerID n'ont pas de doublons. Vous pouvez fusionner sans problème les deux sources de données.
4. Pour afficher les valeurs CustomerID de l'objet de données LA_Customers, double-cliquez sur le cercle nommé **LA_Customers** dans le diagramme de Venn.

Astuce: Double-cliquez sur les cercles dans le diagramme de Venn pour afficher les lignes de données. Si des cercles se chevauchent dans le diagramme de Venn, double-cliquez sur l'intersection pour afficher les valeurs de données communes aux deux ensembles de données.

La Visionneuse de données affiche les valeurs CustomerID contenues dans l'objet de données LA_Customers.

Tâche 3. Exécuter un profil sur une source de données

Dans cette tâche, vous exécutez un profil sur la source de données All_Customers pour afficher le contenu et la structure des données.

1. Dans la vue **Explorateur d'objets**, développez les objets de données du projet de tutoriel.
2. Sélectionnez la source de données **All_Customers**.
3. Cliquez sur **Fichier > Nouveau > Profil**.
La boîte de dialogue **Nouveau** s'affiche.
4. Sélectionnez **Profil**.
5. Cliquez sur **Suivant**.
6. Dans le champ **Nom**, saisissez **All_Customers**.
7. Cliquez sur **Terminer**.

Le profil All_Customers s'ouvre dans l'éditeur et le profil s'exécute.

Tâche 4. Afficher les résultats de profilage de colonne

Dans cette tâche, vous affichez les résultats de profilage de colonnes pour l'objet de données All_Customers et examinez les valeurs et modèles contenus dans les données.

1. Cliquez sur **Fenêtre > Afficher la vue > Progression** pour afficher la progression du profil All_Customers.
La vue Progression s'ouvre.
2. Quand la vue Progression indique que le profil All_Customers a terminé son exécution, cliquez sur la vue **Résultats** dans l'éditeur.
3. Dans la section **Profilage de colonnes**, cliquez sur la colonne **CustomerTier**.
La section **Détails** affiche toutes les valeurs contenues dans la colonne CustomerTier et affiche des informations sur la fréquence d'apparition des valeurs dans l'ensemble de données.
4. Dans la section **Détails**, double-cliquez sur **Ruby**.
La Visionneuse de données s'exécute et affiche les enregistrements où la colonne CustomerTier contient la valeur **Ruby**.
5. Dans la section **Profilage de colonnes**, cliquez sur la colonne **OrderAmount**.
6. Dans la section **Détails**, cliquez sur la liste **Afficher** et sélectionnez **Modèles**.
La section **Détails** affiche les modèles trouvés dans la colonne OrderAmount. La chaîne **9(5)** de la colonne Modèle fait référence aux enregistrements contenant des quantités de commande à cinq chiffres. La chaîne **9(4)** fait référence aux enregistrements contenant des quantités de commande à quatre chiffres.
7. Dans la colonne **Modèle**, double-cliquez sur la chaîne **9(4)**.
La Visionneuse de données s'exécute et affiche les enregistrements où la colonne OrderAmount contient une quantité de commande à quatre chiffres.

8. Dans la section **Détails**, cliquez sur la liste **Afficher** et sélectionnez **Statistiques**.

La section **Détails** affiche les statistiques de la colonne `OrderAmount`, y compris la valeur moyenne, l'écart-type, les longueurs maximale et minimale, les cinq valeurs les plus fréquentes, et les cinq valeurs les moins fréquentes.

Profilage des données Résumé

Dans cette leçon, vous avez appris qu'un profil fournit les informations sur le contenu et la structure des données.

Vous avez appris que vous pouvez effectuer une analyse de jointure sur deux objets de données et afficher le degré de chevauchement entre les objets de données. Vous avez aussi appris que vous pouvez exécuter un profil de colonne sur un objet de données et afficher les valeurs, patrons, et statistiques en relation avec chaque colonne dans l'objet de données.

Vous avez créé le profil `JoinAnalysis` pour déterminer si les données de l'objet de données `Boston_Customers` peuvent se fusionner avec les données de l'objet de données `LA_Customers`. Vous avez affiché les résultats de ce profil et déterminé que toutes les valeurs dans la colonne `CustomerID` sont uniques et que vous pouvez fusionner les objets de données avec succès.

Vous avez créé le profil `All_Customers` et exécuté un profil de colonne sur l'objet de données `All_Customers`. Vous avez vu les résultats de ce profil pour découvrir les valeurs, patrons, et statistiques pour les colonnes dans l'objet de données `All_Customers`. Enfin, vous avez exécuté la Visionneuse de données pour afficher les lignes contenant des valeurs et patrons spécifiques, ce qui vous permet de vérifier la qualité des données.

CHAPITRE 13

Leçon 4. Analyse des données

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'analyse des données, 61](#)
- [Tâche 1. Création d'un objet de données cible, 62](#)
- [Tâche 2. Création d'un mappage pour analyser les données, 64](#)
- [Tâche 3. Exécution d'un profil sur la transformation Analyseur, 66](#)
- [Tâche 4. Exécution du mappage, 66](#)
- [Tâche 5. Affichage de la sortie du mappage, 66](#)
- [Résumé de l'analyse des données, 67](#)

Présentation de l'analyse des données

Vous analysez les données afin d'identifier un ou plusieurs éléments de données dans un champ d'entrée et pour écrire chaque élément dans un champ de sortie différent.

L'analyse vous permet d'avoir plus de contrôle sur les informations de chaque colonne. Par exemple, prenons un champ de données contenant le nom complet d'une personne, `Michel Martin`. Vous pouvez utiliser la transformation Analyseur pour séparer le nom complet en colonnes de données distinctes pour le prénom et le nom. Après avoir analysé les données dans de nouvelles colonnes, vous pouvez créer des opérations personnalisées de qualité des données pour chaque colonne.

Vous pouvez configurer la transformation Analyseur afin d'utiliser des *jeux de jetons* pour analyser les colonnes de données en chaînes de composant. Un jeu de jetons identifie les éléments de données tels que les mots, les codes ZIP, les numéros de téléphone et les numéros de sécurité sociale.

Vous pouvez également utiliser la transformation Analyseur pour analyser les données correspondant aux entrées de la table de référence ou aux expressions régulières que vous entrez.

Scénario

HypoStores souhaite que le format des fichiers de données clients du bureau de Los Angeles corresponde au format des fichiers de données du bureau de Boston. Les données client du bureau de Los Angeles stockent les noms des clients dans une colonne `FullName`, alors que les données clients du bureau de Boston stockent les noms des clients dans des colonnes `FirstName` et `LastName` distinctes. HypoStores doit décomposer les données de la colonne `FullName` du bureau de Los Angeles en prénoms et en noms afin que le format des données de Los Angeles corresponde au format des données de Boston.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Créer et configurer un objet de données `LA_Customers_tgt` qui contiendra des données analysées.
- Créer un mappage afin de décomposer la colonne `FullName` en colonnes `FirstName` et `LastName` distinctes.
- Ajouter l'objet de données `LA_Customers` au mappage à connecter à la source de données.
- Ajouter l'objet de données `LA_Customers_tgt` au mappage pour créer un objet de données cible.
- Ajouter une transformation `Analyseur` au mappage et la configurer pour qu'elle utilise un jeu de jetons afin de décomposer les noms complets en prénoms et noms.
- Exécuter un profil sur la transformation `Analyseur` pour examiner les données avant de générer la source des données cible.
- Exécuter le mappage pour générer les noms décomposés.
- Exécuter la Visionneuse de données pour afficher la sortie du mappage.

Prérequis

Avant de commencer cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 et 2 de ce tutoriel.

Timing

Réservez 20 minutes pour effectuer les tâches de cette leçon.

Tâche 1. Création d'un objet de données cible

Dans cette tâche, vous créez un objet de données `LA_Customers_tgt` sur lequel vous pouvez écrire des noms analysés.

Pour créer un objet de données cible, procédez comme suit :

1. Créez un objet de données `LA_Customers_tgt` basé sur le fichier `LA_Customers.csv`.
2. Configurez les options de lecture et d'écriture pour l'objet de données, y compris les noms et emplacements de fichiers.
3. Ajoutez les colonnes `Firstname` et `Lastname` à l'objet de données `LA_Customers_tgt`.

Étape 1. Création d'un objet de données `LA_Customers_tgt`

Dans cette étape, vous créez un objet de données `LA_Customers_tgt` basé sur le fichier `LA_Customers.csv`.

1. Cliquez sur **Fichier > Nouveau > Objet de données**.
La fenêtre **Nouveau** s'ouvre.
2. Sélectionnez **Objet de données de fichier plat** et cliquez sur **Suivant**.
3. Vérifiez que **Créer à partir d'un fichier plat existant** est sélectionné.
4. Cliquez sur **Parcourir** et allez à `LA_Customers.csv` dans le répertoire suivant : `<Répertoire d'installation Informatique>\clients\DeveloperClient\Tutorials`
5. Cliquez sur **Ouvrir**.

6. Dans le champ **Nom**, entrez `LA_Customers_tgt`.
7. Cliquez sur **Suivant**.
8. Cliquez sur **Suivant**.
9. Dans la section **Options de prévisualisation**, sélectionnez **Importer les noms de colonnes de la première ligne** et cliquez sur **Suivant**.
10. Cliquez sur **Terminer**.
L'objet de données `LA_Customers_tgt` s'affiche dans l'éditeur.

Étape 2. Configuration des options Lire et Écrire

Dans cette étape, vous configurez les options de lecture et d'écriture pour l'objet de données `LA_Customers_tgt`, y compris les noms et emplacements des fichiers.

1. Vérifiez que l'objet de données `LA_Customers_tgt` est ouvert dans l'éditeur.
2. Dans l'éditeur, sélectionnez la vue **Lire**.
3. Cliquez sur **Fenêtre > Afficher la vue > Propriétés**.
4. Dans la vue **Propriétés**, sélectionnez la vue **Exécution**.
5. Dans la colonne **Valeur**, double-cliquez sur le nom du fichier source et entrez `LA_Customers_tgt.csv`.
6. Dans la colonne **Valeur**, double-cliquez pour surligner le répertoire du fichier source.
7. Cliquez avec le bouton droit de la souris et sélectionnez **Copier**.
8. Dans l'éditeur, sélectionnez la vue **Écrire**.
9. Dans la vue **Propriétés**, sélectionnez la vue **Exécution**.
10. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Répertoire du fichier de sortie**.
11. Cliquez avec le bouton droit de la souris et sélectionnez **Coller** pour coller l'emplacement du répertoire que vous avez copié depuis la vue **Lire**.
12. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Options d'en-tête** et sélectionnez `Noms des champs de sortie`.
13. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Nom du fichier de sortie** et entrez `LA_Customers_tgt.csv`.
14. Cliquez sur **Fichier > Enregistrer** pour enregistrer l'objet de données.

Étape 3. Ajout de colonnes à l'objet de données

Dans cette étape, vous ajoutez des colonnes `Firstname` et `Lastname` à l'objet de données `LA_Customers_tgt`.

1. Dans la vue Explorateur d'objets, allez aux objets de données de votre projet de tutoriel.
2. Double-cliquez sur l'objet de données `LA_Customers_tgt`.
L'objet de données `LA_Customers_tgt` s'ouvre dans l'éditeur.
3. Vérifiez que la vue **Aperçu** est sélectionnée.
4. Sélectionnez la colonne **FullName** et cliquez sur le bouton **Nouvelle** pour ajouter une colonne.
Une colonne nommée `FullName1` s'affiche.
5. Renommez la colonne `Firstname`. Cliquez sur le champ **Précision** et entrez « 30 ».
6. Sélectionnez la colonne `Firstname` et cliquez sur le bouton **Nouvelle** pour ajouter une colonne.
Une colonne nommée `Firstname1` s'affiche.

7. Renommez la colonne `Lastname`. Cliquez sur le champ **Précision** et entrez « 30 ».
8. Cliquez sur **Fichier > Enregistrer** pour enregistrer l'objet de données.

Tâche 2. Création d'un mappage pour analyser les données

Dans cette tâche, vous créez un mappage et le configurez pour utiliser des objets de données et une transformation Analyseur.

Pour créer un mappage afin d'analyser les données, procédez comme suit :

1. Créez un mappage.
2. Ajoutez les objets de données source et cible au mappage.
3. Ajouter une transformation Analyseur au mappage.
4. Configurez la transformation Analyseur pour analyser la colonne source contenant le nom complet du client dans des colonnes cibles distinctes contenant le prénom et le nom.

Étape 1. Création d'un mappage

Dans cette étape, vous créez et nommez le mappage.

1. Dans la vue Explorateur d'objets, sélectionnez votre projet de tutoriel.
 2. Cliquez sur **Fichier > Nouveau > Mappage**.
- La fenêtre **Nouveau mappage** s'ouvre.
3. Dans le champ **Nom**, entrez `ParserMapping`.
 4. Cliquez sur **Terminer**.

Le mappage s'ouvre dans l'éditeur.

Étape 2. Ajout d'objets de données au mappage

Dans cette étape, vous ajoutez l'objet de données `LA_Customers` et l'objet de données `LA_Customers_tgt` au mappage.

1. Dans la vue Explorateur d'objets, allez aux objets de données de votre projet de tutoriel.
 2. Sélectionnez l'objet de données `LA_Customers` et faites-le glisser dans l'éditeur.
- La fenêtre **Ajouter l'objet de données physique au mappage** s'ouvre.
3. Vérifiez que **Lire** est sélectionné et cliquez sur **OK**.
- L'objet de données s'affiche dans l'éditeur.
4. Dans la vue Explorateur d'objets, allez aux objets de données de votre projet de tutoriel.
 5. Sélectionnez l'objet de données `Clients_LA` et faites-le glisser dans l'éditeur.
- La fenêtre **Ajouter l'objet de données physique au mappage** s'ouvre.
6. Sélectionnez **Écrire** et cliquez sur **OK**.
- L'objet de données s'affiche dans l'éditeur.

7. Sélectionnez les ports `CustomerID`, `CustomerTier` et `FullName` dans l'objet de données `LA_Customers`. Faites glisser les ports dans le port `CustomerID` de l'objet de données `LA_Customers_tgt`.

Astuce: Maintenez la touche **CTRL** enfoncée pour sélectionner plusieurs ports.

Les ports de l'objet de données `LA_Customers` se connectent aux ports correspondants dans l'objet de données `LA_Customers_tgt`.

Étape 3. Ajout d'une transformation Analyseur au mappage

Dans cette étape, vous ajoutez une transformation Analyseur au mappage `ParserMapping`.

1. Sélectionnez l'éditeur contenant le mappage `ParserMapping`.
2. Dans la palette de transformation, sélectionnez la transformation Analyseur.
3. Cliquez sur l'éditeur.
La fenêtre **Nouvelle transformation Analyseur** s'ouvre.
4. Vérifiez que **Analyseur de jetons** est sélectionné et cliquez sur **Terminer**.
La transformation Analyseur s'affiche dans l'éditeur.
5. Sélectionnez le port `FullName` dans l'objet de données `LA_Customers` et faites-le glisser dans le groupe **Entrée** de la transformation Analyseur.
Le port `FullName` s'affiche dans la transformation Analyseur et il est connecté au port `FullName` de l'objet de données.

Étape 4. Configuration de la transformation Analyseur

Dans cette étape, vous configurez la transformation Analyseur afin de décomposer la colonne contenant le nom complet du client dans des colonnes distinctes contenant le prénom et le nom.

1. Sélectionnez l'éditeur contenant le mappage `ParserMapping`.
2. Cliquez sur la transformation Analyseur.
3. Cliquez sur **Fenêtre > Afficher la vue > Propriétés**.
4. Dans la vue Propriétés, sélectionnez la vue **Stratégies**.
5. Cliquez sur **Nouveau**. L'assistant Nouvelle stratégie s'affiche.
6. Cliquez sur la flèche de sélection dans la colonne Entrées et sélectionnez le port `FullName`.
7. Sélectionnez le délimiteur espace [`\s`].
8. Cliquez sur **Suivant**.
9. Sélectionnez l'opération **Analyser à l'aide d'un jeu de jetons** et cliquez sur **Suivant**.
10. Sélectionnez **Jeux de jetons fixes (Une seule sortie uniquement)** et sélectionnez le jeu de jetons **Non défini**.
11. Cliquez sur le champ **Sorties** et sélectionnez **Nouvelle**.
12. Dans la boîte de dialogue **Sorties d'opération**, remplacez le nom de sortie par `Undefined_Output`.
13. Cliquez sur **Terminer**.
14. Dans la transformation Analyseur, cliquez sur le port `Undefined_Output` et faites-le glisser dans le port `FirstName` de l'objet de données `LA_customers_tgt`.
Une connexion entre les ports s'affiche.

15. Dans la transformation **Analyseur**, cliquez sur le port `OverflowField` et faites-le glisser dans le port `LastName` de l'objet de données `LA_customers_tgt`.
Une connexion entre les ports s'affiche.
16. Cliquez sur **Fichier > Enregistrer** pour enregistrer le mappage.

Tâche 3. Exécution d'un profil sur la transformation Analyseur

Dans cette tâche, vous exécutez un profil sur la transformation **Analyseur** afin de vérifier qu'elle est correctement configurée pour analyser le nom complet.

1. Sélectionnez l'éditeur contenant le mappage `ParserMapping`.
2. Cliquez avec le bouton droit de la souris sur la transformation **Analyseur** et sélectionnez **Profil immédiat**.
Le profil s'exécute et s'ouvre dans l'éditeur.
3. Dans l'éditeur, cliquez sur la vue **Résultats** pour afficher le résultat de l'opération de profilage.
4. Sélectionnez la colonne `Undefined_output` pour afficher les informations relatives à la colonne de la section **Détails**.
Les valeurs contenues dans la colonne `Undefined_output` s'affichent dans la section **Détails**, ainsi que la fréquence et les statistiques en pourcentage pour chaque valeur.
5. Affichez les données et vérifiez que seuls les prénoms s'affichent dans la colonne `Undefined_output`.

Tâche 4. Exécution du mappage

Dans cette tâche, vous exécutez le mappage pour créer la sortie du mappage.

1. Sélectionnez l'éditeur contenant le mappage `ParserMapping`.
2. Cliquez sur **Exécuter > Exécuter le mappage**.
Le mappage s'exécute et écrit la sortie sur le fichier `LA_Customers_tgt.csv`.

Tâche 5. Affichage de la sortie du mappage

Dans cette tâche, vous exécutez la Visionneuse de données pour afficher la sortie du mappage.

1. Dans la vue Explorateur d'objets, recherchez l'objet de données `LA_Customers_tgt` dans votre projet de tutoriel et double-cliquez sur l'objet de données.
L'objet de données s'ouvre dans l'éditeur.
2. Cliquez sur **Fenêtre > Afficher la vue > Visionneuse de données**.
La vue Visionneuse de données s'ouvre.
3. Dans la Visionneuse de données, cliquez sur **Exécuter**.

La Visionneuse de données s'exécute et affiche les données.

4. Vérifiez que les colonnes `FirstName` et `LastName` affichent des données correctement analysées.

Résumé de l'analyse des données

Dans cette leçon, vous avez appris que l'analyse des données identifie les éléments de données dans un champ d'entrée et écrit chaque élément dans une nouvelle colonne.

Vous avez appris que la transformation `Analyseur` sert à analyser les données. Vous avez également appris que vous pouvez créer un profil pour une transformation dans un mappage afin d'analyser la sortie de cette transformation. Enfin, vous avez appris que vous pouvez afficher la sortie du mappage à l'aide de la Visionneuse de données.

Vous avez créé et configuré l'objet de données `LA_Customers_tgt` pour contenir la sortie analysée. Vous avez créé un mappage pour analyser les données. Dans ce mappage, vous avez configuré une transformation `Analyseur` avec un jeu de jetons afin d'analyser les prénoms et les noms de la colonne `FullName` du fichier clients de Los Angeles. Vous avez configuré le mappage afin d'écrire les données analysées dans les colonnes `Firstname` et `Lastname` de l'objet de données `LA_Customers_tgt`. Vous avez également exécuté un profil pour afficher la sortie de la transformation avant l'exécution du mappage. Enfin, vous avez exécuté le mappage et utilisé la Visionneuse de données pour afficher les nouvelles colonnes de données dans l'objet de données `LA_Customers_tgt`.

CHAPITRE 14

Leçon 5. Normalisation des données

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la normalisation des données, 68](#)
- [Tâche 1. Création d'un objet de données cible, 69](#)
- [Tâche 2. Création d'un mappage pour normaliser les données, 70](#)
- [Tâche 3. Exécution du mappage, 73](#)
- [Tâche 4. Affichage de la sortie du mappage, 73](#)
- [Résumé de la normalisation des données, 73](#)

Présentation de la normalisation des données

La normalisation des données améliore la qualité des données en supprimant les erreurs et les incohérences dans ces dernières.

Pour améliorer la qualité des données, normalisez les données contenant les types de valeur suivants :

- Valeurs incorrectes
- Valeurs avec des informations correctes dans un format incorrect
- Valeurs à partir desquelles vous souhaitez extraire de nouvelles informations

Utilisez la transformation de normalisation pour rechercher ces valeurs dans les données. Vous pouvez sélectionner l'un des types d'opération de recherche suivants :

- **Texte.** Recherche des chaînes personnalisées que vous entrez. Supprimez ces chaînes ou remplacez-les par du texte personnalisé.
- **Table de référence.** Recherche des chaînes contenues dans une table de référence que vous sélectionnez. Supprimez ces chaînes ou remplacez-les par des entrées de table de référence ou du texte personnalisé.

Par exemple, vous pouvez configurer la transformation de normalisation pour normaliser les données d'adresse contenant les chaînes personnalisées `Rue` et `R` à l'aide de la chaîne de remplacement `R..` La transformation de normalisation remplace les termes de recherche par le terme `R.` et écrit le résultat dans une nouvelle colonne de données.

Scénario

HypoStores doit normaliser les données d'adresse de ses clients afin d'harmoniser les termes qu'elles utilisent. Les données d'adresse de l'objet de données `All_Customers` contiennent des entrées dont le format n'est pas harmonisé pour les termes communs comme `Rue`, `Boulevard`, `Avenue`, `Allée` et `Parc` d'activités.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Créer et configurer un objet de données `All_Customers_Stdz_tgt` pour qu'il contienne des données normalisées.
- Créer un mappage pour normaliser les termes d'adresse `Rue`, `Boulevard`, `Avenue`, `Allée` et `Parc` d'activités dans un format cohérent.
- Ajouter l'objet de données `All_Customers` au mappage à connecter aux données source.
- Ajouter l'objet de données `All_Customers_Stdz_tgt` au mappage pour créer un objet de données cible.
- Ajouter une transformation `Normalisation` au mappage et la configurer pour normaliser les termes d'adresse.
- Exécuter le mappage pour générer des données d'adresse normalisées.
- Exécuter la Visionneuse de données pour afficher la sortie du mappage.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 et 2 de ce tutoriel.

Timing

Prévoyez 15 minutes pour cette leçon.

Tâche 1. Création d'un objet de données cible

Dans cette tâche, vous créez un objet de données `All_Customers_Stdz_tgt` sur lequel vous pouvez écrire les données normalisées.

Pour créer un objet de données cible, procédez comme suit :

1. Créez un objet de données `All_Customers_Stdz_tgt` basé sur le fichier `All_Customers.csv`.
2. Configurez les options de lecture et d'écriture pour l'objet de données, y compris les noms et emplacements des fichiers.

Étape 1. Création d'un objet de données `All_Customers_Stdz_tgt`

Dans cette étape, vous créez un objet de données `All_Customers_Stdz_tgt` basé sur le fichier `All_Customers.csv`.

1. Cliquez sur **Fichier > Nouveau > Objet de données**.
La fenêtre **Nouveau** s'ouvre.
2. Sélectionnez **Objet de données de fichier plat** et cliquez sur **Suivant**.
3. Vérifiez que **Créer à partir d'un fichier plat existant** est sélectionné.

4. Cliquez sur **Parcourir** et allez à `All_Customers.csv` dans le répertoire suivant : <Répertoire d'installation Informatica>\clients\DeveloperClient\Tutorials
 5. Cliquez sur **Ouvrir**.
 6. Dans le champ `Nom`, entrez `All_Customers_Stdz_tgt`.
 7. Cliquez sur **Suivant**.
 8. Cliquez sur **Suivant**.
 9. Dans la section **Options de prévisualisation**, sélectionnez **Importer les noms de colonnes de la première ligne** et cliquez sur **Suivant**.
 10. Cliquez sur **Terminer**.
- L'objet de données `All_Customers_Stdz_tgt` s'affiche dans l'éditeur.

Étape 2. Configuration des options Lire et Écrire

Dans cette étape, vous configurez les options de lecture et d'écriture pour l'objet de données `All_Customers_Stdz_tgt`, y compris les noms et les emplacements des fichiers.

1. Vérifiez que l'objet de données `All_Customers_Stdz_tgt` est ouvert dans l'éditeur.
2. Dans l'éditeur, sélectionnez la vue **Lire**.
3. Cliquez sur **Fenêtre > Afficher la vue > Propriétés**.
4. Dans la vue **Propriétés**, sélectionnez la vue **Exécution**.
5. Dans la colonne **Valeur**, double-cliquez sur le nom et le type du fichier `All_Customers_Stdz_tgt.csv`.
6. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Répertoire du fichier source**.
7. Cliquez avec le bouton droit de la souris sur le nom surligné et sélectionnez **Copier**.
8. Dans l'éditeur, sélectionnez la vue **Écrire**.
9. Dans la vue **Propriétés**, sélectionnez la vue **Exécution**.
10. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Répertoire du fichier de sortie**.
11. Cliquez avec le bouton droit de la souris et sélectionnez **Coller** pour coller l'emplacement du répertoire que vous avez copié à partir de la vue **Lire**.
12. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Options d'en-tête** et sélectionnez `Noms des champs de sortie`.
13. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Nom du fichier de sortie** et entrez `All_Customers_Stdz_tgt.csv`.
14. Cliquez sur **Fichier > Enregistrer** pour enregistrer l'objet de données.

Tâche 2. Création d'un mappage pour normaliser les données

Dans cette tâche, vous créez un mappage et le configurez pour utiliser des objets de données et une transformation de normalisation.

Pour créer un mappage afin de normaliser les données, procédez comme suit :

1. Créez un mappage.

2. Ajoutez des objets de données source et cible au mappage.
3. Ajouter une transformation Normalisation au mappage.
4. Configurez la transformation Normalisation pour normaliser les termes d'adresse communs dans des formats cohérents.

Étape 1. Création d'un mappage

Dans cette étape, vous créez et nommez le mappage.

1. Dans la vue **Explorateur d'objets**, sélectionnez votre projet de tutoriel.
2. Cliquez sur **Fichier > Nouveau > Mappage**.
La fenêtre **Nouveau mappage** s'ouvre.
3. Dans le champ **Nom**, entrez `StandardizerMapping`.
4. Cliquez sur **Terminer**.
Le mappage s'ouvre dans l'éditeur.

Étape 2. Ajout d'objets de données au mappage

Dans cette étape, vous ajoutez l'objet de données `All_Customers` et l'objet de données `All_Customers_Stdz_tgt` au mappage.

1. Dans la vue **Explorateur d'objets**, allez aux objets de données de votre projet de tutoriel.
2. Sélectionnez l'objet de données `All_Customers` et faites le glisser dans l'éditeur.
La fenêtre **Ajouter l'objet de données physique au mappage** s'ouvre.
3. Vérifiez que **Lire** est sélectionné et cliquez sur **OK**.
L'objet de données s'affiche dans l'éditeur.
4. Dans la vue **Explorateur d'objets**, allez aux objets de données dans votre projet de tutoriel.
5. Sélectionnez l'objet de données `All_Customers_Stdz_tgt` et faites-le glisser dans l'éditeur.
La fenêtre **Ajouter l'objet de données physique au mappage** s'ouvre.
6. Sélectionnez **Écrire** et cliquez sur **OK**.
L'objet de données s'affiche dans l'éditeur.
7. Sélectionnez tous les ports dans l'objet de données `All_Customers`. Faites glisser les ports dans le port `CustomerID` dans l'objet de données `All_Customers_Stdz_tgt`.
Astuce: Maintenez la touche **Maj.** enfoncée pour sélectionner plusieurs ports. Pour sélectionner tous les ports, vous devrez peut-être faire défiler la liste vers le bas.
Les ports de l'objet de données `All_Customers` se connectent aux ports correspondants dans l'objet de données `All_Customers_Stdz_tgt`.

Étape 3. Ajout d'une transformation Normalisation au mappage

Dans cette étape, vous ajoutez une transformation de normalisation afin de normaliser les chaînes des données d'adresses.

1. Sélectionnez l'éditeur contenant le mappage `StandardizerMapping`.
2. Dans la palette de transformation, sélectionnez la transformation de normalisation.
3. Cliquez sur l'éditeur.

Une transformation de normalisation nommée `NewStandardizer` s'affiche dans le mappage.

4. Pour renommer la transformation de normalisation, double-cliquez sur la barre de titre de la transformation et entrez `AddressStandardizer`.
5. Sélectionnez le port `Address1` dans l'objet de données `All_Customers` et faites-le glisser dans le groupe **Entrée** de la transformation de normalisation.

Un port nommé `Address1` s'affiche dans le groupe d'entrée. Le port se connecte au port `Address1` de l'objet de données `All_Customers`.

Remarque: Vous ajoutez un port de sortie à la transformation lorsque vous configurez une stratégie de normalisation.

Étape 4. Configuration de la transformation Normalisation

Dans cette étape, vous configurez la transformation de normalisation pour normaliser les termes d'adresse dans les données source.

Remarque: Vous définirez cinq opérations de normalisation dans cette tâche. Chaque opération remplace une chaîne de la colonne d'entrée par une nouvelle chaîne.

1. Sélectionnez l'éditeur contenant le mappage `StandardizerMapping`.
2. Cliquez sur la transformation Normalisation.
3. Cliquez sur **Fenêtre > Afficher la vue > Propriétés**.
4. Dans la vue Propriétés, sélectionnez **Stratégies**.
5. Cliquez sur **Nouvelle**. L'assistant Nouvelle stratégie s'affiche.
6. Cliquez sur la flèche de sélection dans la colonne Entrées et sélectionnez le port d'entrée `Address1`.
Le champ Sorties affiche `Address1` comme port de sortie.
7. Sélectionnez les délimiteurs espace et virgule [`\s`] et [`,`]. Sélectionnez éventuellement les options pour supprimer les espaces en fin de mot.
8. Cliquez sur **Suivant**.
9. Sélectionnez l'opération **Remplacer les chaînes personnalisées** et cliquez sur **Suivant**.
10. Sous **Propriétés**, cliquez sur **Nouvelle**.
11. Modifiez les champs **Chaînes personnalisées** et **Remplacer par** pour qu'ils contiennent la première paire de chaînes de la table suivante :

Chaînes personnalisées	Remplacer par
RUE	R.
BOULEVARD	BD
AVENUE	AV.
ALLÉE	AL.
PARC D'ACTIVITÉS	PA

12. Répétez les étapes 9 à 12 pour définir les opérations de normalisation pour toutes les chaînes de la table.

13. Faites glisser le port de sortie `Address1` vers le port `Address1` de l'objet de données `All_Customers_Stdz_tgt`.
14. Cliquez sur **Fichier > Enregistrer** pour enregistrer le mappage.

Tâche 3. Exécution du mappage

Dans cette tâche, vous exécutez le mappage pour écrire des adresses normalisées sur l'objet de données de sortie.

1. Sélectionnez l'éditeur contenant le mappage `StandardizerMapping`.
2. Cliquez sur **Exécuter > Exécuter le mappage**.

Le mappage s'exécute et écrit la sortie sur le fichier `All_Customers_Stdz_tgt.csv`.

Tâche 4. Affichage de la sortie du mappage

Dans cette tâche, vous exécutez la visionneuse de données pour afficher la sortie du mappage et vérifier que les données d'adresses sont correctement normalisées.

1. Dans la vue **Explorateur d'objets**, recherchez l'objet de données `All_Customers_Stdz_tgt` dans votre projet de tutoriel et double-cliquez sur l'objet de données.
L'objet de données s'ouvre dans l'éditeur.
2. Cliquez sur **Fenêtre > Afficher la vue > Visionneuse de données**.
La vue Visionneuse de données s'ouvre.
3. Dans la Visionneuse de données, cliquez sur **Exécuter**.
La Visionneuse de données affiche la sortie du mappage.
4. Vérifiez que la colonne `Address1` affiche correctement les données normalisées. Par exemple, toutes les occurrences de la chaîne `RUE` doivent être remplacées par la chaîne `R`.

Résumé de la normalisation des données

Dans cette leçon, vous avez appris que vous pouvez normaliser des données pour supprimer les erreurs et incohérences dans ces dernières.

Vous avez appris que vous pouvez utiliser une transformation de normalisation pour normaliser les chaînes dans une colonne d'entrée. Vous avez également appris que vous pouvez afficher la sortie du mappage à l'aide de la Visionneuse de données.

Vous avez créé et configuré l'objet de données `All_Customers_Stdz_tgt` pour qu'il contienne la sortie normalisée. Vous avez créé un mappage pour normaliser les données. Dans ce mappage, vous avez configuré une transformation de normalisation pour normaliser la colonne `Address1` de l'objet de données `All_Customers`. Vous avez configuré le mappage pour écrire la sortie normalisée sur l'objet de données `All_Customers_Stdz_tgt`. Enfin, vous avez exécuté le mappage et utilisé la Visionneuse de données pour afficher les données normalisées dans l'objet de données `All_Customers_Stdz_tgt`.

CHAPITRE 15

Leçon 6. Validation des données d'adresse

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la validation des données d'adresse, 74](#)
- [Tâche 1. Création d'un objet de données cible , 75](#)
- [Tâche 2. Création d'un mappage pour valider les adresses, 77](#)
- [Tâche 3. Configuration de la transformation du validateur d'adresses, 78](#)
- [Tâche 4. Exécution du mappage, 82](#)
- [Tâche 5. Affichage de la sortie du mappage, 82](#)
- [Résumé de la validation des données d'adresse, 85](#)

Présentation de la validation des données d'adresse

La validation d'adresses est le processus d'évaluation et d'amélioration de la qualité des adresses postales. Elle évalue la qualité des adresses en comparant les adresses d'entrée avec un jeu de données de référence d'adresses valides. Elle améliore la qualité des adresses en identifiant les valeurs d'adresse incorrectes et en utilisant le jeu de données de référence pour créer des champs contenant des valeurs correctes.

Une adresse est valide lorsqu'elle est livrable. Une adresse peut être correctement formatée et contenir les informations de rue, ville et code postal réelles, mais si les données ne produisent pas une adresse livrable, alors celle-ci n'est pas valide. L'outil Developer utilise des jeux de données de référence d'adresses pour vérifier que les adresses d'entrée peuvent être livrées. Informatica fournit des jeux de données de référence d'adresses.

Un jeu de données de référence d'adresses contient des données décrivant toutes les adresses livrables d'un pays. Le processus de validation d'adresses recherche le jeu de données d'adresse qui ressemble le plus aux données de l'adresse d'entrée. Lorsque le processus trouve une correspondance proche dans le jeu de données de référence, il écrit les nouvelles valeurs pour toute valeur de données incorrecte ou incomplète. Le processus crée un jeu de codes alphanumérique décrivant le type de correspondance entre l'adresse d'entrée et les adresses de référence. Il peut également restructurer les adresses et ajouter des informations absentes de l'adresse d'entrée, comme par exemple le suffixe du code ZIP à quatre chiffres pour une adresse aux États-Unis.

Utilisez la transformation du validateur d'adresses pour créer des processus de validation d'adresses dans l'outil Developer. Cette transformation multigroupe contient un jeu de ports d'entrée et de sortie prédéfinis correspondant à tous les champs possibles dans une adresse d'entrée. Lorsque vous configurez une

transformation de l'outil de validation des adresses, vous sélectionnez le jeu de données de référence par défaut et vous créez une structure d'adresse d'entrée et de sortie à l'aide des ports de transformation. Dans cette leçon, vous configurez la transformation pour valider les données d'adresses aux États-Unis.

Scénario

HypoStores doit corriger et compléter les données d'adresse pour s'assurer que ses campagnes de publicité directe et les autres courriers destinés aux consommateurs parviennent à ses clients. Corriger et compléter les données d'adresse réduit également le coût des opérations de publipostage pour l'organisation. De plus, les données clients d'HypoStores doivent inclure les adresses dans un format imprimable suffisamment flexible pour inclure des adresses de longueurs différentes.

Pour satisfaire ces besoins professionnels, l'équipe ICC HypoStores crée un mappage de validation d'adresses dans l'outil Developer.

Objectifs

Dans cette leçon, vous allez effectuer les tâches suivantes :

- Créer un objet de données cible qui contiendra les champs d'adresse validés et les codes de correspondance.
- Créer un mappage avec un objet de données source, un objet de données cible et une transformation du validateur d'adresses.
- Configurer la transformation du validateur d'adresses pour valider les données d'adresse de vos clients.
- Exécuter le mappage pour valider les données d'adresse et examiner les sorties du code de correspondance pour vérifier la validité des données d'adresse.

Prérequis

Avant de débiter cette leçon, vérifiez les prérequis suivants :

- Vous avez suivi les leçons 1 et 2 de ce tutoriel.
- Les données de référence d'adresses aux États-Unis sont installées dans le domaine et enregistrées à l'aide de l'outil Administrator. Contactez votre administrateur Informatica pour vérifier que les données d'adresses aux États-Unis sont installées sur votre système. Les données de référence sont installées à l'aide du programme d'installation de contenu Data Quality.

Timing

Prévoyez 25 minutes pour cette leçon.

Tâche 1. Création d'un objet de données cible

Dans cette tâche, vous créez un objet de données cible, configurez les options d'écriture et ajoutez les ports.

Pour créer et configurer l'objet de données cible, procédez comme suit :

1. Créez un objet de données All_Customers_av_tgt basé sur le fichier All_Customers.csv.
2. Configurez les options de lecture et d'écriture de l'objet de données, y compris les noms et les emplacements de fichiers.
3. Ajoutez des ports à l'objet de données pour recevoir les valeurs du code de correspondance généré par la transformation du validateur d'adresses.

Étape 1. Créez l'objet de données All_Customers_av_tgt

Dans cette étape, vous créez un objet de données `All_Customers_av_tgt` basé sur le fichier `All_Customers.csv`.

1. Cliquez sur **Fichier > Nouveau > Objet de données**.
La fenêtre **Nouveau** s'ouvre.
2. Sélectionnez **Objet de données de fichier plat** et cliquez sur **Suivant**.
3. Vérifiez que **Créer à partir d'un fichier plat existant** est sélectionné. Cliquez sur **Parcourir** à côté de cette sélection, recherchez le fichier `All_Customers.csv` et cliquez sur **Ouvrir**.
4. Dans le champ **Nom**, entrez `All_Customers_av_tgt`.
5. Cliquez sur **Suivant**.
6. Cliquez sur **Suivant**.
7. Dans la section **Options de prévisualisation**, sélectionnez **Importer les noms de colonnes de la première ligne** et cliquez sur **Suivant**.
8. Cliquez sur **Terminer**.
L'objet de données `All_Customers_av_tgt` s'affiche dans l'éditeur.

Étape 2. Configuration des options Lire et Écrire

Dans cette étape, vous configurez les options de lecture et d'écriture pour l'objet de données `All_Customers_av_tgt`, y compris le nom et l'emplacement du fichier cible.

1. Vérifiez que l'objet de données `All_Customers_av_tgt` est ouvert dans l'éditeur.
2. Dans l'éditeur, sélectionnez la vue **Lire**.
3. Sélectionnez **Fenêtre > Afficher la vue > Propriétés**.
4. Dans la vue **Propriétés**, sélectionnez la vue **Exécution**.
5. Dans la colonne **Valeur**, double-cliquez sur le nom et type de fichier source `All_Customers_av_tgt.csv`.
6. Dans la colonne **Valeur**, double-cliquez pour surligner le chemin du répertoire du fichier source.
7. Cliquez avec le bouton droit de la souris sur le nom et le chemin surlignés et sélectionnez **Copier**.
8. Dans l'éditeur, sélectionnez la vue **Écrire**.
9. Dans la vue **Propriétés**, sélectionnez la vue **Exécution**.
10. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Répertoire du fichier de sortie**.
11. Cliquez avec le bouton droit de la souris sur cette entrée et sélectionnez **Coller** pour ajouter le chemin que vous avez copié à partir de la vue **Lire**.
12. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Options d'en-tête** et sélectionnez `Noms des champs de sortie`.
13. Dans la colonne **Valeur**, double-cliquez sur l'entrée **Nom du fichier de sortie** et entrez `All_Customers_av_tgt.csv`.
14. Sélectionnez **Fichier > Enregistrer** pour enregistrer l'objet de données.

Étape 3. Ajout de ports à l'objet de données

Dans cette étape, vous ajoutez deux ports à l'objet de données `All_Customers_av_tgt` pour que la transformation du validateur d'adresses puisse écrire les valeurs du code de correspondance sur le fichier cible. Nommez les ports `MailabilityScore` et `MatchCode`.

La valeur `MailabilityScore` décrit si l'adresse d'entrée est livrable. La valeur `MatchCode` décrit le type de correspondance que la transformation effectue entre l'adresse d'entrée et les adresses des données de référence.

1. Dans la vue **Explorateur d'objets**, allez aux objets de données de votre projet de tutoriel.
2. Double-cliquez sur l'objet de données `All_Customers_av_tgt`.
L'objet de données `All_Customers_av_tgt` s'ouvre dans l'éditeur.
3. Vérifiez que **Aperçu** est sélectionné.
4. Sélectionnez le port final dans la liste des ports. Ce port est nommé `MiscDate`.
5. Cliquez sur **Nouveau**.
Un port nommé `MiscDate1` s'affiche.
6. Renommez le port `MiscDate1` `MailabilityScore`.
7. Sélectionnez le port `MailabilityScore`.
8. Cliquez sur **Nouveau**.
Un port nommé `MailabilityScore1` s'affiche.
9. Renommez le port `MailabilityScore1` `MatchCode`.
10. Cliquez sur **Fichier > Enregistrer** pour enregistrer l'objet de données.

Tâche 2. Création d'un mappage pour valider les adresses

Dans cette tâche, vous créez un mappage et ajoutez des objets de données et une transformation du validateur d'adresses.

Pour créer le mappage et ajouter les objets dont vous avez besoin, procédez comme suit :

1. Créez un objet de mappage.
2. Ajoutez des objets de données source et cible au mappage.
3. Ajouter une transformation du validateur d'adresses au mappage.

Étape 1. Création d'un mappage

Dans cette étape, vous créez et nommez le mappage.

1. Dans la vue **Explorateur d'objets**, sélectionnez votre projet de tutoriel.
2. Sélectionnez **Fichier > Nouveau > Mappage**.
La fenêtre **Nouveau mappage** s'ouvre.
3. Dans le champ **Nom**, entrez `ValidateAddresses`.

4. Cliquez sur **Terminer**.
Le mappage s'ouvre dans l'éditeur.

Étape 2. Ajout d'objets de données au mappage

Dans cette étape, vous ajoutez les objets de données source et cible au mappage.

`All_Customers` est l'objet de données source du mappage. La transformation du validateur d'adresses lit les données de cet objet. `All_Customers_av_tgt` est l'objet de données cible du mappage. Cet objet lit les données depuis la transformation du validateur d'adresses

1. Dans la vue **Explorateur d'objets**, allez aux objets de données de votre projet de tutoriel.
2. Sélectionnez l'objet de données `All_Customers` et faites-le glisser dans l'éditeur.
La fenêtre **Ajouter l'objet de données physique au mappage** s'ouvre.
3. Vérifiez que **Lire** est sélectionné et cliquez sur **OK**.
L'objet de données s'affiche dans l'éditeur.
4. Dans la vue **Explorateur d'objets**, allez aux objets de données de votre projet de tutoriel.
5. Sélectionnez l'objet de données `All_Customers_av_tgt` et faites-le glisser dans l'éditeur.
La fenêtre **Ajouter l'objet de données physique au mappage** s'ouvre.
6. Sélectionnez **Écrire** et cliquez sur **OK**.
L'objet de données s'affiche dans l'éditeur.
7. Cliquez sur **Enregistrer**.

Étape 3. Ajout d'une transformation du validateur d'adresses au mappage

Dans cette étape, vous ajoutez une transformation du validateur d'adresses au mappage contenant la source et les objets de données.

Lorsque cette étape est terminée, vous pouvez configurer la transformation et connecter ses ports aux objets de données.

1. Sélectionnez l'éditeur contenant le mappage `ValidateAddresses`.
2. Dans la palette de transformation, sélectionnez la transformation du validateur d'adresses.
3. Cliquez sur l'éditeur.
La transformation du validateur d'adresses s'affiche dans l'éditeur.

Tâche 3. Configuration de la transformation du validateur d'adresses

Dans cette tâche, vous configurez la transformation du validateur d'adresses pour lire et valider les adresses à partir de la source de données `All_Customers`.

Remarque: La transformation du validateur d'adresses contient une série de ports d'entrée et de sortie prédéfinis. Sélectionnez les ports dont vous avez besoin et connectez-les aux objets dans le mappage.

Pour configurer la transformation, procédez comme suit :

1. Sélectionnez le jeu de données de référence d'adresses par défaut.
2. Configurez les ports de transformation et connectez la transformation au mappage.
3. Connectez les ports source inutilisés à la cible de données.

Étape 1. Définition du jeu de données de référence d'adresses par défaut

Dans cette étape, vous sélectionnez le jeu de données de référence par défaut. Les fichiers de données de référence d'adresses sont définis par pays, vous sélectionnez donc un nom de pays comme jeu de données par défaut. La transformation du validateur d'adresses utilise les données de référence relatives à ce pays s'il ne peut pas déterminer le pays à utiliser à partir des valeurs des données d'entrée.

1. Sélectionnez la transformation du validateur d'adresses dans l'éditeur.
2. Sous **Propriétés**, cliquez sur **Paramètres généraux**.
3. Dans le menu **Pays par défaut**, sélectionnez **États-Unis**.

Étape 2. Configuration des ports d'entrée de la transformation du validateur d'adresses

Dans cette étape, vous sélectionnez les ports d'entrée de la transformation et les connectez à l'objet de données `All_Customers_av`.

La transformation du validateur d'adresses contient plusieurs groupes de ports d'entrée prédéfinis. Sélectionnez les ports d'entrée correspondant aux champs de votre adresse d'entrée et ajoutez-les à la transformation.

Maintenez la touche Ctrl enfoncée lorsque vous sélectionnez les ports dans les étapes suivantes afin de sélectionner plusieurs ports en une seule opération.

1. Sélectionnez la transformation du validateur d'adresses dans l'éditeur.
2. Sous **Propriétés**, cliquez sur **Modèles**.
3. Développez le groupe de ports **Modèle de base**.
4. Développez le groupe de ports d'entrée **Hybride** et sélectionnez les ports suivants :

Nom de port	Description
Ligne d'adresse de livraison 1	Données d'adresse relatives à la rue, comme par exemple le nom de la rue et le numéro du bâtiment.
Localité complète 1	Nom de la ville.
Code postal 1	Code postal ou code ZIP.
Province 1	Nom de la province/état ou département/région.
Nom du pays	Nom ou abréviation du pays.

Remarque: Maintenez la touche Ctrl enfoncée pour sélectionner plusieurs ports en une seule opération.

5. Dans la barre d'outils située au dessus de la liste des noms de ports, cliquez sur **Ajouter le port à la transformation**.

Cette barre d'outils est visible lorsque vous sélectionnez **Modèles**.

Les ports sélectionnés s'affichent dans la transformation de l'éditeur de mappage.

6. Connectez les ports source à la transformation du validateur d'adresses comme suit :

Port source	Port de transformation du validateur d'adresses
Address1	Ligne d'adresse de livraison 1
Ville	Localité complète 1
ZIP	Code postal 1
État/Province/Département/Région	Province 1
Pays	Nom du pays

Étape 3. Configuration de la transformation du validateur d'adresses

Dans cette étape, vous sélectionnez les ports de sortie de la transformation et connectez ces ports à l'objet de données `All_Customers_av_tgt`.

La transformation du validateur d'adresses contient plusieurs groupes de ports de sortie prédéfinis. Sélectionnez les ports définissant la structure de l'adresse dont vous avez besoin et ajoutez ces ports à la transformation.

Vous pouvez également sélectionner des ports contenant des informations sur le type de validation effectuée pour chaque adresse.

1. Sélectionnez la transformation du validateur d'adresses dans l'éditeur de mappage.
2. Sous **Propriétés**, cliquez sur **Modèles**.
3. Développez le groupe de port **Modèle de base**.
4. Développez le groupe de ports de sortie **Éléments d'adresse** et sélectionnez le port suivant :

Nom de port	Description
Rue complète 1	Données d'adresse relatives à la rue, comme par exemple le nom de la rue et le numéro du bâtiment.

5. Développez le groupe de ports de sortie **Éléments de dernière ligne** et sélectionnez les ports suivants :

Nom de port	Description
Localité complète 1	Nom de la ville.
Code postal 1	Code postal ou code ZIP.
Province abréviation 1	Identifiant de la province/état/département/région.

Remarque: Maintenez la touche Ctrl enfoncée pour sélectionner plusieurs ports en une seule opération.

6. Développez le groupe de ports de sortie **Pays** et sélectionnez le port suivant :

Nom de port	Description
Nom du pays 1	Nom du pays.

7. Développez le groupe de ports de sortie **Info Statut** et sélectionnez les ports suivants :

Nom de port	Description
Score d'envoi	Score représentant les probabilités de réussite de la livraison postale.
Code de correspondance	Code représentant le degré de similarité entre l'adresse d'entrée et les données de référence.

8. Dans la barre d'outils située au dessus de la liste des noms de port, cliquez sur **Ajouter le port à la transformation**.

Cette barre d'outils s'affiche lorsque vous sélectionnez **Modèles**.

9. Connectez les ports de la transformation du validateur d'adresses aux ports `All_Customers_av_tgt` comme suit :

Port de la transformation du validateur d'adresses	Port cible
Rue complète 1	Address1
Localité complète 1	Ville
Code postal 1	ZIP
Province abréviation 1	Département/Région
Nom du pays 1	Pays
Score d'envoi	MailabilityScore
Code de correspondance	MatchCode

Étape 4. Connexion des ports inutilisés de la source de données à la cible des données

Dans cette étape, vous connectez les ports inutilisés sur la source de données `All_Customers` à la cible de données.

- Connectez les ports inutilisés sur la source de données aux ports de même noms sur la cible de données.

Tâche 4. Exécution du mappage

Dans cette tâche, vous exécutez le mappage pour créer la sortie du mappage.

1. Sélectionnez l'éditeur contenant le mappage `ValidateAddresses`.
2. Sélectionnez **Exécuter** > **Exécuter le mappage**.

Le mappage s'exécute et écrit la sortie sur le fichier `All_Customers_av_tgt.csv`.

Tâche 5. Affichage de la sortie du mappage

Dans cette tâche, vous exécutez la Visionneuse de données pour afficher la sortie du mappage. Vérifiez la qualité de vos adresses validées en examinant les valeurs écrites dans les colonnes Score d'envoi et Code de correspondance de l'objet de données cible.

La valeur du code de correspondance est un code alphanumérique représentant le type de validation que le mappage a effectué sur l'adresse.

La valeur du Score d'envoi est une valeur à un chiffre qui résume si l'adresse est livrable.

1. Dans la vue Explorateur d'objets, recherchez l'objet de données `All_Customers_av_tgt` de votre projet de tutoriel et double-cliquez sur l'objet de données.

L'objet de données s'ouvre dans l'éditeur.

2. Sélectionnez **Fenêtre** > **Afficher la vue** > **Visionneuse de données**.

La Visionneuse de données s'ouvre.

3. Dans la Visionneuse de données, cliquez sur **Exécuter**.

La Visionneuse de données affiche la sortie du mappage.

4. Faites défiler les résultats du mappage afin d'afficher les colonnes Score d'envoi et Code de correspondance.

5. Vérifiez les valeurs de la colonne Score d'envoi.

Les scores peuvent être compris entre 0 à 5. Les adresses possédant les scores les plus élevés sont plus susceptibles d'être livrées.

6. Vérifiez les valeurs de la colonne Code de correspondance.

Le code de correspondance est un code alphanumérique. Le caractère alphabétique indique le type de validation que la transformation a effectué et le chiffre indique la qualité de l'adresse finale.

La table suivante décrit les valeurs du code de correspondance :

Code	Description
A1	La recherche du code d'adresse a renvoyé une adresse partielle ou complète pour le code entré.
A0	La recherche du code d'adresse n'a renvoyé aucune adresse pour le code entré.
C4	Corrigé. Tous les éléments pertinents de types postaux sont vérifiés.

Code	Description
C3	Corrigé. Certains éléments ne peuvent pas être vérifiés.
C2	Corrigé, mais l'état de livraison est incertain en raison de l'absence des données de référence.
C1	Corrigé, mais le statut de livraison est incertain, car la normalisation utilisateur a introduit des erreurs.
I4	Les données ne peuvent pas être entièrement corrigées, mais il existe une correspondance unique avec une adresse dans les données de référence.
I3	Les données ne peuvent pas être entièrement corrigées et il existe plusieurs correspondances avec des adresses dans les données de référence.
I2	Les données ne peuvent pas être corrigées. Le mode Lot renvoie des suggestions d'adresses partielles.
I1	Les données ne peuvent pas être corrigées. Le mode Lot ne peut pas suggérer d'adresse.
N7	Erreur de validation. La validation n'a pas eu lieu, car la validation sur une seule ligne n'est pas déverrouillée.
N6	Erreur de validation. La validation n'a pas eu lieu, car la validation sur une seule ligne n'est pas prise en charge pour le pays de destination.
N5	Erreur de validation. La validation n'a pas eu lieu, car la base de données de référence est obsolète.
N4	Erreur de validation. La validation n'a pas eu lieu, car les données de référence sont altérées ou leur format est incorrect.
N3	Erreur de validation. La validation n'a pas eu lieu, car les données de pays ne peuvent pas être déverrouillées.
N2	Erreur de validation. La validation n'a pas eu lieu, car la base de données de référence requise n'est pas disponible.
N1	Erreur de validation. La validation n'a pas eu lieu, car le pays n'est pas reconnu ou pris en charge.
Q3	Mode Liste de suggestion. La validation des adresses peut récupérer une ou plusieurs adresses complètes à partir des données de référence d'adresse qui correspondent à l'adresse d'entrée.
Q2	Mode Liste de suggestion. La validation des adresses peut combiner les éléments d'adresse d'entrée et des éléments à partir des données de référence d'adresse pour créer une adresse complète.
Q1	Mode Liste de suggestion. La validation des adresses ne peut pas suggérer d'adresse complète. Pour générer une suggestion d'adresse complète, ajoutez des données à l'adresse d'entrée.

Code	Description
Q0	Mode Liste de suggestion. Les données d'entrée sont insuffisantes pour générer une suggestion.
RB	Pays reconnu d'après l'abréviation. Reconnaît les codes de pays ISO à deux et trois caractères. Peut également reconnaître les abréviations courantes telles que « GER » pour l'Allemagne.
RA	Pays reconnu depuis le paramètre Forcer le pays dans la transformation.
R9	Pays reconnu depuis le paramètre Pays par défaut dans la transformation.
R8	Pays reconnu à partir de son nom.
R7	Pays reconnu à partir de son nom, mais la transformation a identifié des erreurs dans les données de pays.
R6	Pays reconnu à partir des données de territoire.
R5	Pays reconnu à partir des données de province.
R4	Pays reconnu à partir des données de ville principale.
R3	Pays reconnu à partir du format d'adresse.
R2	Pays reconnu à partir d'un script.
R1	Pays non reconnu, car plusieurs correspondances sont disponibles.
R0	Pays non reconnu.
S4	Mode Analyser. L'adresse a été parfaitement analysée.
S3	Mode Analyser. L'adresse a été analysée avec plusieurs résultats.
S1	Mode Analyser. Une non-correspondance au niveau du format d'entrée a généré une erreur d'analyse.
V4	Vérifié. Les données d'entrée sont correctes. La validation des adresses a vérifié tous les éléments pertinents de types postaux et les entrées correspondent parfaitement.
V3	Vérifié. Les données d'entrée sont correctes, mais une partie ou la totalité des éléments a été normalisée, ou l'entrée contient des noms ou des exonymes obsolètes.
V2	Vérifié. Les données d'entrée sont correctes, mais certains éléments ne peuvent pas être vérifiés en raison de données de référence incomplètes.
V1	Vérifié. Les données d'entrée sont correctes, mais la normalisation utilisateur a eu un impact négatif sur le caractère livrable. Par exemple, le code postal est trop court.

Résumé de la validation des données d'adresse

Dans cette leçon, vous avez appris que la validation d'adresses compare les données d'adresse d'entrée avec les données de référence et renvoie la version la plus précise possible de l'adresse.

Vous avez appris que le processus de validation d'adresses renvoie également les informations du statut relatives à la qualité de chaque adresse.

Vous avez appris que les utilisateurs de l'outil Administrator exécutent le programme d'installation de contenu Data Quality pour installer les données de référence d'adresses.

Vous avez également appris que la transformation de l'outil de validation des adresses est une transformation multigroupe et que vous sélectionnez les ports d'entrée et de sortie pour la transformation parmi les groupes de ports. Les ports d'entrée que vous sélectionnez déterminent le contenu de l'adresse qui est validé. Les ports de sortie déterminent le contenu de l'enregistrement de l'adresse finale.

ANNEXE A

Forum Aux Questions (FAQ)

Cette annexe comprend les rubriques suivantes :

- [FAQ Informatica Analyst, 86](#)
- [Foire Aux Questions \(FAQ\) Informatica Developer , 86](#)

FAQ Informatica Analyst

Consultez les FAQ pour répondre aux questions que vous vous posez sur Informatica Analyst.

Puis-je accéder aux outils Administrator, Developer et Analyst à partir d'un seul compte ?

Oui. Vous pouvez autoriser un utilisateur à accéder aux trois outils. Il n'est pas nécessaire de créer des comptes d'utilisateur différents pour chaque application client.

Où les données de référence sont-elles stockées ?

Vous pouvez utiliser les outils Developer et Analyst pour créer et partager les objets de données de référence. Le référentiel modèle stocke les métadonnées de l'objet de données de référence. La base de données des données de référence stocke les valeurs de données de table de référence. Configurez la base de données de référence sur le service de gestion du contenu.

Foire Aux Questions (FAQ) Informatica Developer

Consultez la foire aux questions pour répondre aux questions que vous vous posez à propos d'Informatica Developer.

Quelle est la différence entre une source et une cible dans PowerCenter et un objet de données physique dans Developer Tool ?

Dans PowerCenter, vous créez une définition source à inclure comme mappage source. Vous créez une définition cible à inclure comme mappage cible. Dans Developer Tool, vous créez un objet de données physique que vous pouvez utiliser comme mappage source ou cible.

Quelle est la différence entre un mappage dans Developer Tool et un mappage dans PowerCenter ?

Un mappage PowerCenter indique comment déplacer les données entre les sources et les cibles. Un mappage de Developer Tool indique comment déplacer les données entre l'entrée et la sortie du mappage.

Un mappage PowerCenter doit inclure une ou plusieurs définitions source, qualifiants source et définitions cible. Un mappage PowerCenter peut aussi inclure des raccourcis, des transformations et des mapplets.

Un mappage de Developer Tool doit inclure une entrée et une sortie de mappage. Un mappage de Developer Tool peut aussi inclure des transformations et des mapplets.

Developer Tool dispose des types de mappage suivants :

- Mappage déplaçant les données entre les sources et les cibles. Ce type de mappage diffère du mappage de PowerCenter seulement par le fait qu'il ne peut pas utiliser de raccourcis et qu'il n'utilise pas de qualifiant source.
- Mappage d'objet de données logique. Mappage dans un modèle d'objet de données logique. Un mappage d'objet de données logique peut contenir un objet de données logique comme entrée de mappage et un objet de données comme sortie de mappage. Il peut également contenir un ou plusieurs objets de données physiques comme entrée de mappage et un objet de données logique comme mappage de sortie.
- Mappage de table virtuelle. Mappage dans un service de données SQL. Il contient un objet de données comme entrée de mappage et une table virtuelle comme sortie de mappage.
- Mappage de procédure stockée virtuelle. Définit un ensemble de logique métier dans un service de données SQL. Il contient une transformation de paramètre d'entrée ou objet de données physique comme entrée de mappage et une transformation de paramètre de sortie ou objet de données physique comme sortie de mappage.

Quelle est la différence entre un mapplet dans PowerCenter et un mapplet dans Developer Tool ?

Un mapplet dans PowerCenter et dans Developer Tool est un objet réutilisable qui contient un ensemble de transformations. Vous pouvez réutiliser la logique de transformation dans plusieurs mappages.

Un mapplet PowerCenter peut contenir des définitions source ou des transformations d'entrée comme entrée de mapplet. Il doit contenir des transformations de sortie comme sortie de mapplet.

Un mapplet de Developer Tool peut contenir des objets de données ou des transformations d'entrée comme entrée de mapplet. Il peut contenir des objets de données ou des transformations de sortie comme sortie de mapplet. Un mappage dans Developer Tool peut aussi inclure les fonctions suivantes :

- Vous pouvez valider un mapplet en tant que règle. Vous utilisez une règle dans un profil.
- Un mapplet peut contenir d'autres mapplets.

Quelle est la différence entre un mapplet et une règle ?

Vous pouvez valider un mapplet en tant que règle. Une règle est une logique métier qui définit les conditions appliquées aux données source lorsque vous exécutez un profil. Vous pouvez valider un mapplet en tant que règle lorsque le mapplet répond aux spécifications suivantes :

- Il contient une transformation d'entrée et de sortie.
- Le mapplet ne contient pas de transformations actives.
- Elle ne spécifie pas de cardinalité entre les groupes d'entrée.