



Informatica®

9.6.1 HotFix 2

Guide de démarrage du profilage

© 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 © Unicode, 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.

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 SORTIE, 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/javaservice/>, <http://www.postgresql.org/about/licence.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> et <https://github.com/yokato/libgeohash/blob/master/LICENSE>.

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 des 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; et RE44,478), des brevets internationaux ainsi que par d'autres brevets en attente.

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: 2018-07-05

Sommaire

Préface.....	7
Ressources Informatica.	7
Portail Mon support Informatica.	7
Documentation Informatica.	7
Site Web Informatica.	7
Bibliothèque de procédures Informatica.	8
Base de connaissances Informatica.	8
Canal YouTube du support Informatica.	8
Informatica Marketplace.	8
Informatica Velocity.	8
Support client international Informatica.	8
 Chapitre 1: Présentation de la mise en route.....	 10
Présentation du profilage.	10
Introduction à Informatica Analyst.	11
Interface utilisateur d'Informatica Analyst.	12
Interface utilisateur d'Informatica Developer.	14
Page d'accueil Informatica Developer.	15
Feuilles d'aide.	15
Tâches d'Informatica Developer.	16
Scénario du tutoriel.	16
Structure du tutoriel.	16
Prérequis du tutoriel.	17
Tutoriel de Informatica Analyst.	17
Outil Informatica Developer.	18
 Partie I: Débuter avec Informatica Analyst.....	 19
 Chapitre 2: Leçon 1. Configuration de Informatica Analyst.....	 20
Configuration de Informatica Analyst - Présentation.	20
Tâche 1. Se connecter à Informatica Analyst.	21
Tâche 2. Créer un projet.	21
Tâche 3. Créer un dossier.	21
Configuration de Informatica Analyst - Résumé.	22
 Chapitre 3: Leçon 2. Création d'objets de données.....	 23
Création d'objets de données - Présentation.	23
Tâche 1. Créer l'objet de données de fichier plat.	24
Tâche 2. Afficher les propriétés de l'objet de données.	24
Création d'objets de données - Résumé.	25

Chapitre 4: Leçon 3. Création de profils rapides.	26
Création de profils rapides - Présentation.	26
Tâche 1. Créer et exécuter un profil rapide.	27
Résultats de profil de colonne.	27
Tâche 2. Afficher les résultats de profil.	28
Création de profils rapides - Résumé.	28
 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

Préface

Le *Guide de démarrage* Informatica s'adresse aux développeurs et analystes responsables de la qualité des données et des services de données. Il fournit un tutoriel 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.

Ce site contient des informations sur les produits et les groupes d'utilisateurs, des bulletins d'information, un lien vers la 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>.

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 profilage, 10](#)
- [Introduction à Informatica Analyst, 11](#)
- [Interface utilisateur d'Informatica Developer, 14](#)
- [Scénario du tutoriel, 16](#)
- [Structure du tutoriel, 16](#)

Présentation du profilage

Les profils s'utilisent pour rechercher le contenu, la qualité et la structure des sources de données d'une application, d'un schéma ou d'une entreprise. Le contenu de la source de données comprend les fréquences de valeurs et les types de données. La structure de sources de données comprend les clés et les dépendances fonctionnelles.

Vous pouvez créer et exécuter des profils comme partie du processus de découverte. Un profil est un objet de référentiel qui recherche et analyse toutes les irrégularités de données dans les sources de données de l'entreprise et les problèmes de données cachés qui mettent en péril les projets de données. L'exécution d'un profil sur une source de données quelconque de l'entreprise vous aide à bien comprendre les points forts et les points faibles de ses données et métadonnées.

Vous pouvez utiliser les outils Analyst et Developer pour analyser les données source et les métadonnées. Les analystes et les développeurs peuvent utiliser ces outils pour collaborer, identifier les problèmes de qualité de données et analyser les relations de données. Selon votre fonction, vous pouvez utiliser les fonctionnalités de l'outil Analyst ou de l'outil Developer. Le degré de profilage que vous pouvez effectuer varie selon l'outil que vous utilisez.

Vous pouvez effectuer les tâches suivantes dans l'outil Developer et l'outil Analyst :

- Effectuer le profilage de colonnes. Ce processus comprend la découverte du nombre de valeurs uniques, de valeurs Null et de modèles de données dans une colonne.
- Effectuez la découverte de domaines de données. Vous pouvez découvrir des caractéristiques de données critiques au sein d'une entreprise.
- Conserver les résultats de profil, y compris les types de données, les domaines de données, les clés primaires et les clés étrangères.
- Créer des fiches d'évaluation pour contrôler la qualité des données.
- Afficher le lignage des fiches d'évaluation pour chaque score de fiche d'évaluation et groupe de scores.

- Créer et affecter des balises aux objets de données.
- Rechercher la signification d'un nom d'objet comme un terme métier dans le bureau de glossaire métier. Par exemple, vous pouvez rechercher la signification d'un nom de colonne ou d'un nom de profil pour comprendre sa spécification dans l'entreprise et son implémentation actuelle.

Vous pouvez effectuer les tâches suivantes dans l'outil Developer :

- Découvrir le degré de jointures potentielles entre deux colonnes de données dans une source de données.
- Déterminer le pourcentage de chevauchement des données entre des paires de colonnes d'une source de données ou de plusieurs sources de données.
- Comparer les résultats du profilage de colonnes.
- Générer un objet de mappage depuis un profil.
- Découvrir des clés principale dans une source de données.
- Découvrir des clés étrangères dans un ensemble d'une ou plusieurs sources de données.
- Découvrir la dépendance fonctionnelle entre les colonnes d'une source de données.
- Exécuter des tâches de découverte de données sur un grand nombre de sources de données sur plusieurs connexions. Les tâches de découverte des données incluent le profil de colonne, l'inférence des relations de clé principale à clé étrangère, la découverte des domaines de données et la génération d'un résumé graphique consolidé des relations de données.

Vous pouvez effectuer les tâches suivantes dans l'outil Analyst :

- Effectuer la découverte des données d'entreprise sur un grand nombre de sources de données sur plusieurs connexions. Vous pouvez afficher un résumé consolidé des résultats de découverte de métadonnées de colonne et de domaines de données.
- Effectuer une recherche de découverte pour trouver où les données et métadonnées existent dans l'entreprise. Vous pouvez rechercher des ressources spécifiques, telles que des objets de données, des règles et des profils. La découverte de recherche trouve les ressources et identifie les relations avec d'autres ressources des bases de données et des schémas de l'entreprise.

Introduction à Informatica Analyst

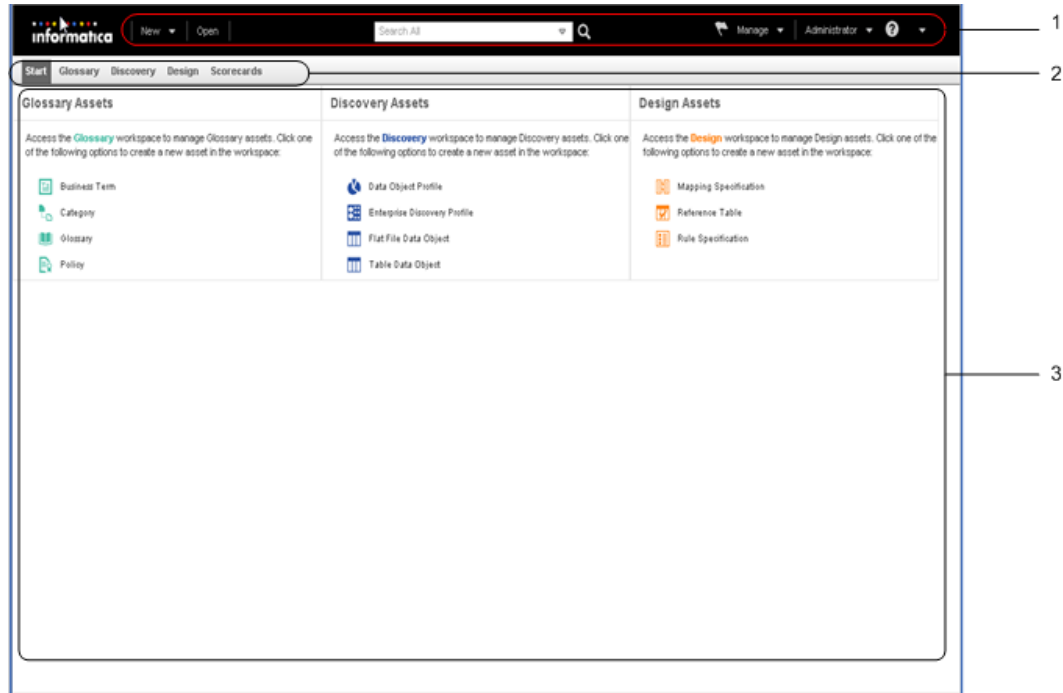
Informatica Analyst est un client d'application basé sur le Web que les analystes peuvent utiliser pour analyser, nettoyer, normaliser, effectuer le profil et évaluer les données dans une entreprise.

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.

Interface utilisateur d'Informatica Analyst

L'interface utilisateur de l'outil Analyst comporte plusieurs espaces de travail que vous utilisez pour créer des solutions d'intégration de données, de qualité des données et de fédération des données. Un espace de travail est une page Web à laquelle vous pouvez accéder en fonction de la licence et des privilèges.

L'image suivante montre l'interface utilisateur de l'outil Analyst :



1. Zone d'en-tête
2. Onglets des espaces de travail
3. Panneau d'accès aux espaces de travail

L'outil Analyst s'ouvre dans l'espace de travail **Démarrer**. L'espace de travail **Démarrer** répertorie les espaces de travail pour lesquels vous disposez d'une licence et du privilège d'utilisation via des panneaux d'accès aux espaces de travail.

L'outil Analyst affiche les menus d'en-tête suivants :

Nouveau

Permet de créer des ressources dans les espaces de travail **Glossaire**, **Découvrir** et **Conception**.

Ouvrir

Permet d'ouvrir l'espace de travail **Bibliothèque**.

Notifications d'alerte

Permet d'afficher les notifications pour les ressources de glossaire.

Gérer

Ouvrir des espaces de travail temporaires et des **Notifications**. Vous pouvez ouvrir les espaces de travail **Connexions**, **Domaines de données**, **Statut de la tâche**, **Projets** et **Sécurité du glossaire métier**.

Nom d'utilisateur

Permet de définir les préférences de l'utilisateur pour modifier le mot de passe et se déconnecter de l'outil Analyst.

Aide

Permet d'accéder à l'aide de l'espace de travail en cours.

L'outil Analyst contient les espaces de travail permanents suivants :

Démarrer

Permet d'accéder à d'autres espaces de travail pour lesquels vous disposez d'une licence d'accès via les panneaux d'accès aux espaces de travail. Si vous disposez de la licence permettant d'effectuer la gestion d'exceptions, vos tâches s'affichent dans le panneau **Mes tâches** de l'espace de travail.

Glossaire

Contient les définitions et les descriptions des concepts métier importants pour votre organisation. Vous pouvez créer et gérer les termes, les catégories, les glossaires et les stratégies métier.

Découverte

Permet d'analyser la qualité des données et des métadonnées des systèmes source. Vous pouvez créer et gérer les profils des objets de données, les profils des découvertes d'entreprise, les objets de données fichier plat ainsi que les objets de données table.

Conception

Permet de concevoir la logique métier qui aide les analystes et les développeurs à collaborer. Vous pouvez créer et gérer les spécifications de mappage, les tables de référence et les définitions de règles.

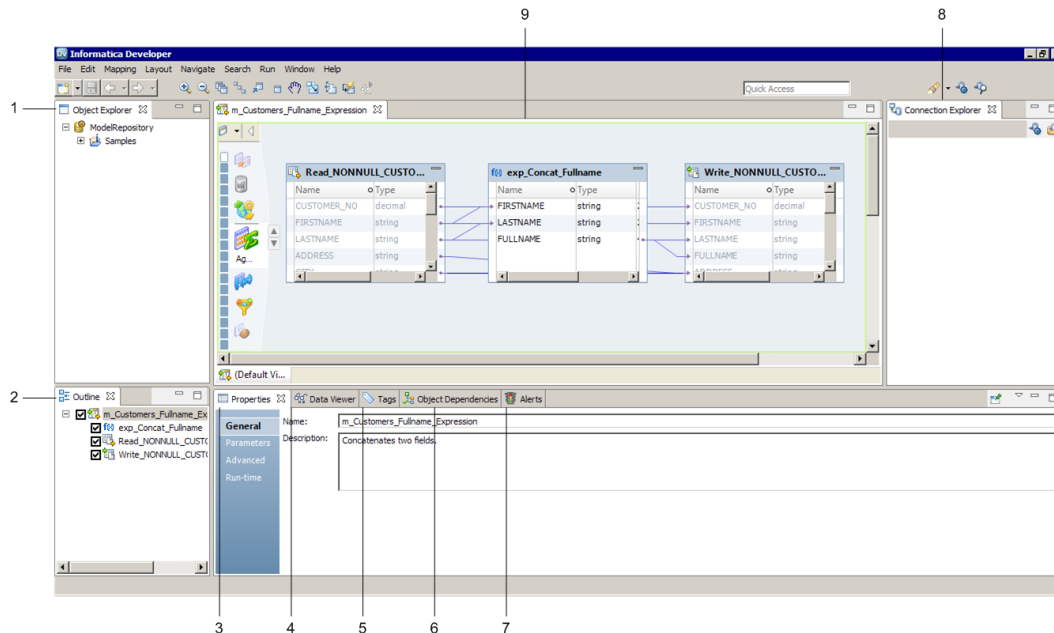
Fiches d'évaluation

Permettent d'ouvrir, de modifier et d'exécuter des fiches d'évaluation que vous avez créées à partir des résultats de profil. Vous pouvez ajouter des scores, développer des colonnes et afficher des graphes de tendance pour une fiche d'évaluation.

Interface utilisateur d'Informatica Developer

L'interface utilisateur de l'outil Developer comporte un espace de travail composé de plusieurs vues que vous utilisez pour créer des solutions d'intégration de données et de qualité des données de fédération de données .

L'image suivante présente l'interface utilisateur de l'outil Developer :



1. Vue Explorateur d'objets
2. Vue Structure
3. Vue Propriétés
4. Vue Visionneuse de données
5. Vue Balises
6. Vue Dépendances d'objets
7. Affichage des alertes
8. Vue Explorateur de connexion
9. Éditeur

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. Vous pouvez également prévisualiser des messages de service Web ou exécuter une requête SQL depuis la vue Visionneuse de données.

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**.

Tâches d'Informatica Developer

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 des 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.
- 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 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.

Scénario du tutoriel

HypoStores Corporation est une société de vente au détail 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 classe 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 souhaite effectuer le profil des données de Boston et de Los Angeles avant qu'elles intègrent les ensembles de données. Les opérations de profil identifient les problèmes de qualité des données que HypoStores peut réparer avant l'intégration.

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.

Prérequis du tutoriel

Avant de commencer les leçons du tutoriel, le domaine Informatica doit être actif avec au moins un nœud configuré. Vous devez vérifier que le service Analyst, le service de référentiel modèle et le service d'intégration de données sont en cours d'exécution.

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
- Customer_Order.xsd
- LA_customers.csv
- orders.csv

Tutoriel de Informatica Analyst

Dans ce tutoriel, un analyste se connecte à l'outil Analyst, crée des projets et des dossiers, des profils et des règles, des fiches d'évaluation sur des données et des tables de référence.

Le tableau suivant décrit les leçons que vous pouvez effectuer :

Leçon	Description
Leçon 1. Configuration d'Informatica Analyst	Connectez-vous à l'outil Analyst et créez un projet et un dossier pour les leçons du tutoriel.
Leçon 2. Création d'objets de données	Importez un fichier plat comme objet de données et prévisualisez les données.
Leçon 3. Création de profils rapides	Création d'un profil rapide pour obtenir rapidement une idée de la 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.
Leçon 5. Création de règles d'expression	Créez des règles d'expression pour modifier et effectuer le profil des valeurs de colonne.
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.

Leçon	Description
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.
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.

Outil Informatica Developer

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

Les utilisateurs du profilage utilisent l'outil Developer pour créer et exécuter des profils qui analysent le contenu et la structure de données.

Exécutez les leçons suivantes dans le tutoriel :

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.

Partie I : Débuter avec Informatica Analyst

Cette partie contient les chapitres suivants :

- [Leçon 1. Configuration de Informatica Analyst, 20](#)
- [Leçon 2. Création d'objets de données, 23](#)
- [Leçon 3. Création de profils rapides, 26](#)
- [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, 20](#)
- [Tâche 1. Se connecter à Informatica Analyst, 21](#)
- [Tâche 2. Créer un projet, 21](#)
- [Tâche 3. Créer un dossier, 21](#)
- [Configuration de Informatica Analyst - Résumé, 22](#)

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, 23](#)
- [Tâche 1. Créer l'objet de données de fichier plat, 24](#)
- [Tâche 2. Afficher les propriétés de l'objet de données, 24](#)
- [Création d'objets de données - Résumé, 25](#)

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 l'objet de données de fichier plat

Dans cette tâche, vous utilisez l'assistant **Ajouter un fichier plat** pour créer un objet de données fichier plat à partir du fichier de données LA_Customers.

1. Dans le navigateur **Projets**, sélectionnez le dossier Clients dans votre projet de tutoriel.
Remarque: Vous devez sélectionner le projet ou dossier dans lequel vous voulez créer l'objet de données de fichier plat pour pouvoir le créer.
2. Cliquez sur **Nouveau > Objet de données de fichier plat**.
L'assistant **Ajouter un fichier plat** s'affiche dans l'espace de travail **Découverte**.
3. Sélectionnez **Parcourir et charger** et cliquez sur **Choisir le fichier**.
4. Allez à l'emplacement de LA_Customers.csv et cliquez sur **Ouvrir**.
5. Cliquez sur **Suivant**.
La page **Choisissez le type d'importation** affiche les options **Délimité** et **Largeur fixe**. L'option par défaut est **Délimité**.
6. Cliquez sur **Suivant**.
7. Dans **Spécifier les délimiteurs et qualificateurs de texte utilisés dans les données**, sélectionnez **Guillemets doubles** comme qualificateur de texte.
8. Sous **Spécifiez les lignes à importer**, sélectionnez **Importer depuis la première ligne** pour importer les noms de colonne depuis la première ligne qui n'est pas vide.
9. Cliquez sur **Suivant**.
Le panneau **Attributs des colonnes** affiche le nom, le type de données, la précision, l'échelle et le format de chaque colonne.
10. Cliquez sur **Suivant**.
Le champ **Nom** affiche **LA_Customers**.
11. Éventuellement, modifiez le nom du fichier et ajoutez une description.
12. Dans la section **Dossiers**, sélectionnez le dossier **Clients**.
13. Cliquez sur **Terminer**.
L'objet de données LA_Customers s'affiche dans l'espace de travail **Découverte** avec les données de prévisualisation des données source.

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, 26](#)
- [Tâche 1. Créer et exécuter un profil rapide, 27](#)
- [Tâche 2. Afficher les résultats de profil, 28](#)
- [Création de profils rapides - Résumé, 28](#)

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**.

Résultats de profil de colonne

Les résultats de profil de colonne comprennent le nom, les valeurs uniques, les valeurs Null et les types de données.

Le tableau suivant décrit les informations qui s'affichent pour chaque colonne d'un profil :

Propriété	Description
Nom	Nom de la colonne dans le profil.
Valeurs uniques	Nombre de valeurs uniques dans la colonne.
% Unique	Pourcentage des valeurs uniques dans la colonne.
NULL	Nombre de valeurs Null dans la colonne.
% Null	Pourcentage des valeurs de colonne qui sont Null.
Type de données	Type de données tiré des valeurs de la colonne. L'outil Analyst peut tirer les types de données suivants à partir des valeurs de colonne : <ul style="list-style-type: none">- String- Varchar- Décimal- Entier- Null [-]

Propriété	Description
% déduits	Pourcentage des valeurs qui correspondent au type de données déduit par l'outil Analyst.
Type de données consigné	Type de données déclaré pour la colonne du projet profilé.
Valeur minimale	Valeur minimale dans la colonne.
Valeur maximale	Valeur maximale dans la colonne.
Dernière exécution de profil	Date et heure où vous avez exécuté le profil pour la dernière fois.
Développement	Si sélectionné, permet le développement des données en direct pour la colonne.

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 de développement pour un profilage plus rapide. Configurez les options d'échantillonnage pour sélectionner les lignes d'échantillon dans le fichier plat. Configurez les options de développement pour développer les enregistrements dans les résultats de profil et développer les lignes dans les données sources ou données temporaires.

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 sélectionnez l'objet de données et les colonnes d'objet de données que vous voulez profiler, configurez les options d'échantillonnage et configurez les options de développement.

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 développeur 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 débiter 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](#)

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.