



Informatica®

Informatica®

10.5.3

Guide de Catalog Administrator

© Copyright Informatica LLC 2015, 2022

Ce logiciel et la documentation associée sont fournis uniquement sous un accord de licence séparé contenant des restrictions d'utilisation et de divulgation. Il est interdit de reproduire ou de transmettre sous quelle que forme et par quel que moyen que ce soit (électronique, photocopie, enregistrement ou autre) tout ou partie de ce document sans le consentement préalable d'Informatica LLC.

Informatica et le logo Informatica sont des marques ou des marques déposées d'Informatica LLC aux États-Unis et dans de nombreux autres pays. La liste actuelle des marques de commerce de Informatica est disponible sur le Web à l'adresse <https://www.informatica.com/trademarks.html>. Les autres noms de société ou de produit peuvent être des marques de commerce ou des marques déposées de leurs détenteurs respectifs.

U.S. GOVERNMENT RIGHTS Les programmes, les logiciels, les bases de données et les documents connexes et les données techniques fournis aux clients du gouvernement américain sont des « logiciels commerciaux » ou des « données techniques commerciales », conformément au règlement fédéral sur les acquisitions et aux règlements supplémentaires propres à l'Agence. En tant que tel, l'utilisation, la duplication, la divulgation, la modification et l'adaptation sont assujetties aux restrictions et aux conditions de licence énoncées dans le contrat gouvernemental applicable et, dans la mesure applicable par les termes du contrat gouvernemental, les droits additionnels énoncés dans la réglementation FAR 52.227-19, licence de logiciel d'ordinateur commercial.

Soumis à vos droits de retrait, le logiciel transmettra automatiquement certaines informations à Informatica (aux États-Unis) concernant l'environnement informatique et réseau dans lequel le Logiciel est déployé et les statistiques du système et d'utilisation des données du déploiement. Cette transmission est considérée comme faisant partie des Services selon la politique de confidentialité d'Informatica et Informatica utilisera et traitera par ailleurs ces informations conformément à la politique de confidentialité d'Informatica disponible sur <https://www.informatica.com/in/privacy-policy.html>. Il est possible de désactiver la collecte d'utilisation dans l'outil Administrator tool.

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 © University of Toronto. Tous droits réservés. Copyright © Daniel Veillard. Tous droits réservés. Copyright © Unicon, Inc. Copyright IBM Corp. Tous droits réservés. Copyright © MicroQuill Software Publishing, Inc. Tous droits réservés. Copyright © PassMark Software Pty Ltd. Tous droits réservés. Copyright © LogiXML, Inc. Tous droits réservés. Copyright © 2003-2010 Lorenzi Davide. Tous droits réservés. Copyright © Red Hat, Inc. Tous droits réservés. Copyright © The Board of Trustees of the Leland Stanford Junior University. Tous droits réservés. Copyright © EMC Corporation. Tous droits réservés. Copyright © Flexera Software. Tous droits réservés. Copyright © Jinfonet Software. Tous droits réservés. Copyright © Apple Inc. Tous droits réservés. Copyright © Telerik Inc. Tous droits réservés. Copyright © BEA Systems. Tous droits réservés. Copyright © PDFlib GmbH. Tous droits réservés. Copyright © Orientation in Objects GmbH. Tous droits réservés. Copyright © Tanuki Software, Ltd. Tous droits réservés. Copyright © Ricebridge. Tous droits réservés. Copyright © Sencha, Inc. Tous droits réservés. Copyright © Scalable Systems, Inc. Tous droits réservés. Copyright © jQWidgets. Tous droits réservés. Copyright © Tableau Software, Inc. Tous droits réservés. Copyright © MaxMind, Inc. Tous droits réservés. Copyright © TMatte Software s.r.o. Tous droits réservés. Copyright © MapR Technologies Inc. Tous droits réservés. Copyright © Amazon Corporate LLC. Tous droits réservés. Copyright © Highsoft. Tous droits réservés. Copyright © Python Software Foundation. Tous droits réservés. Copyright © BeOpen.com. Tous droits réservés. Copyright © CNRI. Tous droits réservés.

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

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

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

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

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

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

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

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

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

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

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

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

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

Ce produit comprend des logiciels sous licence dont les conditions se trouvent aux adresses : <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib/?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqldbLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>, <http://antlr.org/license.html>, <http://aopalliance.sourceforge.net/>, <http://www.bouncycastle.org/licence.html>, <http://www.jgraph.com/jgraphdownload.html>, <http://www.jcraft.com/jsch/LICENSE.txt>, http://jotm.objectweb.org/bsd_license.html, <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>, <http://www.slf4j.org/license.html>, <http://nanoxml.sourceforge.net/orig/copyright.html>, <http://www.json.org/license.html>, <http://forge.ow2.org/projects/javaxservice/>, <http://www.postgresql.org/about/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>, <https://github.com/lyokato/libgeohash/blob/master/LICENSE>, <https://github.com/hjiang/jsonxx/blob/master/LICENSE>, <https://code.google.com/p/lz4/>, <https://github.com/jedisct1/libsodium/blob/master/LICENSE>, <http://one-jar.sourceforge.net/index.php?page=documents&file=license>, <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>, <http://www.scala-lang.org/license.html>, <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>, <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>, <https://aws.amazon.com/asl/>, <https://github.com/twbs/bootstrap/blob/master/LICENSE>, <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>, <https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE>, and <https://github.com/apache/hbase/blob/master/LICENSE.txt>.

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.

EXCLUSION DE RESPONSABILITÉ : Informatica LLC 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 LLC 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.

Consultez les brevets applicables à l'adresse <https://www.informatica.com/legal/patents.html>.

Les renseignements contenus dans cette documentation sont sujets à modification sans préavis. Si vous constatez des problèmes liés à la documentation, merci de les signaler par courriel à l'adresse infa_documentation@Informatica.com.

Les produits Informatica sont garantis conformément aux termes et conditions des accords en vertu desquels ils sont fournis. INFORMATICA FOURNIT LES INFORMATIONS DE CE DOCUMENT « EN L'ÉTAT » SANS GARANTIE D'AUCUNE SORTE, EXPRESSE OU IMPLICITE, NOTAMMENT AUCUNE GARANTIE DE QUALITÉ MARCHANDE, D'ADAPTATION À UN USAGE PARTICULIER ET D'ABSENCE DE CONTREFAÇON

Date de publication: 2022-12-20

Sommaire

Préface.....	11
Ressources Informatica.	11
Informatica Network.	11
Base de connaissances Informatica.	11
Documentation Informatica.	12
Matrices de disponibilité des produits Informatica.	12
Informatica Velocity.	12
Informatica Marketplace.	12
Support client international Informatica.	12
 Chapitre 1: Introduction à l'administration du catalogue.....	 13
Présentation d'Enterprise Data Catalog.	13
Architecture Enterprise Unified Metadata.	14
Présentation de l'administration du catalogue.	15
Processus d'administration du catalogue.	15
Accès à Catalog Administrator.	16
Prérequis.	16
Connectez-vous à Catalog Administrator.	16
Modification du mot de passe.	17
 Chapitre 2: Concepts d'Enterprise Data Catalog.....	 18
Présentation des concepts d'Enterprise Data Catalog.	18
Catalogue.	19
Type de ressource.	19
Ressource.	19
Scanner.	19
Planification.	20
Exemple d'entreprise.	20
Domaines de données et groupes de domaines de données.	21
Domaines de données et groupes de domaines de données dans Catalog Administrator.	21
Domaines de données prédéfinis.	22
Domaines de données basés sur des règles.	23
Domaines de données composites.	29
Flux de travail de la découverte de domaines de données composites.	31
Découverte de données.	31
Types de ressources pris en charge pour la découverte de données.	31
Paramètres de connexion au domaine.	33
Paramètres de profil de base.	34
Paramètres de préparation des données du profil de similarité et de fréquence de valeur.	39
Paramètres d'inférence de clé unique.	40

Similarité de colonne.	41
Mode de fonctionnement de la similarité de colonne.	42
Processus de similarité de colonne.	43
Choix des facteurs de propagation des domaines de données intelligents.	44
Exemple d'entreprise.	45
Propagation des termes d'entreprise.	47
Business Glossary Synchronisation.	49
Aperçu, préparation et provisionnement des données.	50
Prérequis.	50
Aperçu des données.	51
Provisionnement de données	53
Préparation des données.	56
Chapitre 3: Utilisation de Catalog Administrator.....	58
Présentation de Catalog Administrator.	58
Espace de travail Présentation.	60
Espace de travail Ressource.	60
Espace de travail Surveillance.	61
Espace de travail Bibliothèque.	61
Espace de travail Domaines de données.	62
Chapitre 4: Gestion des ressources.....	63
Présentation de la gestion des ressources.	63
Ressources et scanners.	64
Ressources et planifications.	64
Ressources et attributs.	64
Ressources de référence.	65
Avantages de l'extraction des actifs de référence.	65
Attribuer des connexions aux ressources de référence.	66
Exemple de cas d'utilisation.	66
Affichage des ressources de référence.	68
Création d'une ressource.	68
Type de ressource.	69
Filtre de profil des données et métadonnées source.	71
Liaison insensible à la casse	72
Activation de la découverte de données.	73
Découverte de domaines de données composites.	80
Activer l'association des termes d'entreprise	81
Activer le provisionnement de données	84
Activer la préparation des données.	84
Modification d'une ressource.	85
Clonage d'une ressource.	85
Exécution d'une analyse sur une ressource.	86

Ressources système.	86
Affichage d'une ressource.	87
Chapitre 5: Gestion de la sécurité des ressources	88
Présentation de la gestion de la sécurité des ressources.	88
Configuration des autorisations par défaut pour les ressources.	89
Configuration des autorisations pour des utilisateurs et des groupes d'utilisateurs spécifiques.	90
Sélection de ressources pour attribuer des autorisations à des utilisateurs ou des groupes d'utilisateurs spécifiques.	91
Application de limitations à une ressource.	92
Attribution des autorisations de configuration aux ressources.	93
Chapitre 6: Gestion des planifications.	94
Présentation de la gestion des planifications.	94
Types de planifications.	94
Planifications réutilisables.	94
Planifications personnalisées.	95
Création d'une planification.	95
Affichage de la liste des planifications.	95
Chapitre 7: Gestion des attributs.	97
Présentation de la gestion des attributs.	97
Attributs système.	97
Attributs personnalisés.	98
Propriétés générales des attributs.	98
Propriétés de la configuration de recherche.	99
Création d'un attribut personnalisé.	99
Modification d'un attribut personnalisé et d'un attribut système.	100
Attribution d'attributs personnalisés aux classes et aux ressources.	100
Chapitre 8: Attribution de connexions.	102
Présentation de l'attribution de connexions.	102
Connexions attribuées automatiquement.	102
Connexions attribuées par l'utilisateur.	103
Gestion des connexions.	103
Attribution de connexions aux ressources de référence.	104
Rapport des liens manquants	104
Colonnes du rapport des liens manquants	105
Génération du rapport des liens manquants	105
Chapitre 9: Configuration des paramètres réutilisables.	106
Présentation de la configuration réutilisable.	106
Propriétés générales de configuration.	106

Propriétés de la connexion du service d'intégration de données.	107
Configuration d'un service d'intégration de données réutilisable.	107
Chapitre 10: Surveillance d'Enterprise Data Catalog.	108
Présentation de la surveillance d'Enterprise Data Catalog.	108
Statuts de tâches.	109
Distribution des tâches.	109
Surveillance par ressource.	110
Surveillance par tâche.	110
Gestion des tâches	111
Application de filtres pour surveiller les tâches.	111
Chapitre 11: Gestion des domaines de données.	113
Présentation de la gestion des domaines de données.	113
Meilleures pratiques pour la création de domaines de données à l'aide de règles.	114
Création d'un domaine de données basé sur des règles.	114
Création d'un groupe de domaines de données.	117
Affichage des domaines de données et des groupes de domaines de données.	117
Filtrage de domaines de données.	118
Filtrage de groupes de domaines de données.	118
Modification d'un domaine de données ou d'un groupe de domaines de données.	118
Autorisations et privilèges.	119
Découverte de domaines de données sur le moteur Spark.	120
Découverte de domaines de données sur le cluster Databricks.	121
Chapitre 12: Gestion des domaines de données composites.	123
Présentation de la gestion des domaines de données composites.	123
Création de domaines de données composites.	123
Affichage des domaines de données composites existants.	124
Filtrage des domaines de données composites.	125
Modification des domaines de données composites existants.	125
Suppression des domaines de données composites existants.	125
Chapitre 13: Gérer les définitions de synonymes.	127
Présentation de la gestion des définitions de synonymes.	127
Chargement de fichiers de définitions de synonymes.	128
Chapitre 14: Gestion de l'organisation Cloud.	129
Présentation de la gestion de l'organisation Cloud	129
Propriétés de l'organisation Cloud	129
Ajout d'une organisation Informatica Intelligent Cloud Services.	130
Modification de l'organisation Informatica Intelligent Cloud Services.	130

Chapitre 15: Présentation de l'intégration de métadonnées personnalisées 131

Présentation de l'intégration de métadonnées personnalisées.	131
Flux de travail d'intégration de métadonnées personnalisées.	132
Téléchargement des fichiers de définition de modèle pour la validation.	133
Création de modèles personnalisés.	134
Mise à jour des modèles personnalisés.	134
Exportation de modèles.	135
Désapprobation de modèles personnalisés.	135
Vue d'ensemble du type de ressource personnalisé.	135
Ressources ETL personnalisées.	136
Création de types de ressources personnalisés.	136
Création de ressources personnalisées.	136
Aperçu de l'ingestion des métadonnées.	138
Exportation du modèle de type de ressource personnalisé.	138
Saisie des détails de l'association.	139
Saisie des détails de la classe.	140
Saisie des détails de la transformation pour les ressources ETL personnalisées.	141
Personnaliser les icônes et configurer les vues de relations pour les ressources personnalisées.	143
Flux de travail.	143
Exportation du canevas de modèle.	144
Création d'un fichier XML de modèle personnalisé.	144
Création du fichier ZIP d'icônes.	144
Création du fichier XML de configuration de vue.	145
Création du modèle personnalisé avec les icônes personnalisés.	154
Lignage de contexte pour les ressources personnalisées.	155
Structure de fichier CSV de ressource de lignage de contexte.	157

Chapitre 16: Utilitaires de sauvegarde de catalogue et de collecte de journaux..... 159

Présentation.	159
Accès aux utilitaires.	160
Exécution d'une sauvegarde de catalogue.	160
Utilitaires de collecte de journaux.	161
Collecter les journaux de service Enterprise Data Catalog.	161
Collecter les journaux de ressource.	162
.	162

Annexe A: Enregistrement de types de données inconnus pour l'exécution de profils..... 163

Présentation.	163
Exemple.	163
Enregistrement d'un type de données inconnu comme type de données pris en charge.	164

Annexe B: Gestion de l'importation à partir de ServiceNow	167
Présentation de la gestion de l'importation à partir de ServiceNow.	167
Importation de connexions à partir de ServiceNow	167
Annexe C: Fichier CSV de ressource de lignage personnalisée	169
Fichier CSV de ressource de lignage personnalisée	169
Annexe D: Agent Enterprise Data Catalog.....	171
Installation d'Enterprise Data Catalog Agent.	171
Installation d'Enterprise Data Catalog Agent en tant que Service Windows.	172
Configuration d'Enterprise Data Catalog Agent pour SSL.	173
Annexe E: Lignage entre les ressources PowerCenter et les fichiers plats..	174
Présentation.	174
Configuration et exécution d'une ressource de fichier plat sur une machine Windows.	175
Affichage du lignage entre les ressources PowerCenter et les fichiers plats.	176
Annexe F: Extraction de métadonnées à partir de ressources inaccessibles et hors ligne.....	179
Présentation.	179
Processus.	181
Composants de l'utilitaire de scanneur.	182
Prérequis et meilleures pratiques.	182
Exécution du script de scanneur	183
Commande de script.	184
generateResourceConfig.	186
encryptPassword.	186
validate.	187
scan.	187
cancel.	188
pause.	189
resume.	190
describeArchive.	190
publishArchive.	191
Structure du fichier d'archive des métadonnées.	193
Annexe G: Enterprise Data CatalogModules complémentaires.....	195
Présentation des modules complémentaires Enterprise Data Catalog.	195
Déployer les modules complémentaires.	196
Connectez-vous aux modules complémentaires Enterprise Data Catalog.	196
Afficher les planifications de ressources et les ressources en double.	196
Accepter les domaines de données.	197

Associer des termes d'entreprise à partir du fichier de synonymes.	198
--	-----

Index.	200
-----------------------	------------

Préface

Pour effectuer des tâches administratives dans Enterprise Data Catalog à l'aide de l'application Catalog Administrator, consultez le *Guide d'Informatica® Catalog Administrator*. Ce guide comprend des informations sur les tâches administratives, notamment sur la création de ressources et de ressources personnalisées, la gestion des planifications et des attributs, et la configuration des profils.

Ressources Informatica

Informatica vous fournit toute une gamme de ressources de produits via Informatica Network et autres portails en ligne. Utilisez ces ressources pour tirer le meilleur parti de vos produits et solutions Informatica, et pour apprendre d'autres utilisateurs et experts en la matière d'Informatica.

Informatica Network

Informatica Network est la passerelle à de nombreuses ressources, y compris la base de connaissances Informatica et le support client international Informatica. Pour accéder à Informatica Network, visitez le site <https://network.informatica.com>.

En tant que membre d'Informatica Network, vous disposez des options suivantes :

- Rechercher les ressources de produits dans la base de connaissances.
- Afficher les informations de disponibilité des produits.
- Créer et vérifier vos dossiers de support.
- Rechercher votre réseau de groupe d'utilisateurs local Informatica et collaborer avec vos pairs.

Base de connaissances Informatica

Utilisez la base de connaissances Informatica pour rechercher des ressources de produits telles que des articles pratiques, des meilleures pratiques, des didacticiels vidéo et des questions fréquemment posées.

Pour effectuer des recherches dans la base de connaissances, visitez le site <https://search.informatica.com>. N'hésitez pas à contacter l'équipe de la base de connaissances Informatica à l'adresse KB_Feedback@informatica.com pour lui faire part de vos questions, commentaires et suggestions concernant la base de connaissances.

Documentation Informatica

Utilisez le portail de documentation Informatica pour explorer une vaste bibliothèque de documentation pour les versions de produits actuelles et récentes. Pour explorer le portail de documentation, visitez le site <https://docs.informatica.com>.

N'hésitez pas à contacter l'équipe Documentation Informatica à l'adresse info_documentation@informatica.com pour lui faire part de vos questions, commentaires ou suggestions concernant la documentation des produits.

Matrices de disponibilité des produits Informatica

Les matrices de disponibilité des produits (PAM) indiquent les versions des systèmes d'exploitation, les bases de données et les types de source et cible de données pris en charge par une version d'un produit. Vous pouvez parcourir les PAM Informatica à l'adresse <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity est un ensemble de conseils et de meilleures pratiques développés par les services professionnels d'Informatica et basés sur les expériences réelles de centaines de projets de gestion des données. Informatica Velocity représente le savoir collectif de consultants d'Informatica qui collaborent avec des organisations du monde entier pour planifier, développer, déployer et gérer des solutions performantes de gestion des données.

Vous trouverez les ressources d'Informatica Velocity à l'adresse <http://velocity.informatica.com>. Si vous avez des questions, des commentaires ou des suggestions sur Informatica Velocity, contactez les services professionnels d'Informatica à l'adresse ips@informatica.com.

Informatica Marketplace

Informatica Marketplace est un forum dans lequel vous pouvez trouver des solutions qui permettent d'augmenter et d'améliorer vos implémentations Informatica. Exploitez les centaines de solutions de développeurs et de partenaires Informatica sur Marketplace pour améliorer votre productivité et accélérer le délai d'implémentation de vos projets. Vous trouverez Informatica Marketplace à l'adresse <https://marketplace.informatica.com>.

Support client international Informatica

Vous pouvez contacter un centre de support international par téléphone ou via le réseau Informatica.

Pour rechercher le numéro de téléphone du support client international Informatica local, visitez le site Web Informatica à l'adresse <https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>.

Pour trouver des ressources de support en ligne sur le réseau Informatica, visitez le site <https://network.informatica.com> et sélectionnez l'option eSupport.

CHAPITRE 1

Introduction à l'administration du catalogue

Ce chapitre comprend les rubriques suivantes :

- [Présentation d'Enterprise Data Catalog, 13](#)
- [Architecture Enterprise Unified Metadata, 14](#)
- [Présentation de l'administration du catalogue, 15](#)
- [Processus d'administration du catalogue, 15](#)
- [Accès à Catalog Administrator, 16](#)

Présentation d'Enterprise Data Catalog

Enterprise Data Catalog regroupe toutes les ressources de données d'une entreprise et présente une vue d'ensemble des ressources de données et de leurs relations. Une ressource de données est un type d'objet de données tel qu'une source de données physique, HDFS ou un référentiel Big Data. Les ressources de données de l'entreprise peuvent exister dans des bases de données relationnelles, des applications conçues à des fins spécifiques, des outils de création de rapport, HDFS et d'autres référentiels Big Data.

Enterprise Data Catalog capture les métadonnées physiques et opérationnelles d'un grand nombre de ressources de données que vous utilisez pour déterminer l'efficacité des données d'entreprise. Les métadonnées sont des données relatives aux données. Les métadonnées contiennent des informations sur la structure des sources de données. Les métadonnées incluent également des informations telles que les modèles de données, les types de données, les relations entre les colonnes et les relations entre plusieurs sources de données.

Enterprise Data Catalog rassemble des informations relatives aux métadonnées à travers l'entreprise. Les métadonnées incluent des statistiques de données de colonne, des domaines de données, des relations d'objet de données et des informations de lignage. Une vue d'ensemble des métadonnées d'entreprise peut vous permettre de prendre des décisions essentielles sur l'intégration des données, la qualité des données et l'administration des données au sein de l'entreprise.

Enterprise Data Catalog traite les questions clés suivantes liées aux métadonnées dans l'entreprise :

- Quel est le contenu d'une ressource de données ?
- Que signifie le contenu d'une ressource de données ?
- Qui est responsable d'une ressource de données spécifique dans l'entreprise ?
- Quelle est la source des données d'une ressource de données ?

- Quelles sont les données sensibles de l'entreprise ?
- Où se trouvent les données sensibles ?

Architecture Enterprise Unified Metadata

L'architecture Enterprise Unified Metadata se compose d'applications, de services et de bases de données. La couche des applications se compose d'applications clientes, telles qu'Enterprise Data Catalog. La couche des services contient des services d'application, tels que le Catalog Service, Data Integration Service et Model Repository Service. Les métadonnées unifiées d'entreprise requièrent le service de catalogue pour extraire les métadonnées des sources de données et gérer les tâches administratives. La couche des bases de données se compose du référentiel modèle et du cluster Informatica pour le stockage et l'analyse des métadonnées. Les sources de données et de métadonnées incluent les référentiels de données source, tels qu'Oracle, Microsoft SQL Server, le référentiel PowerCenter et SAP Business Objects.

Le tableau suivant décrit les composants de l'architecture :

Composant	Description
External Application	Application telle qu'Enterprise Data Catalog que vous utilisez pour découvrir, explorer et relier différents types de métadonnées provenant de sources disparates dans l'entreprise.
Scanner Framework	Framework qui exécute des scanners et gère un registre des scanners disponibles. Un scanner est un composant enfichable qui extrait des métadonnées spécifiques des sources de données externes.
Model Repository Service	Service d'application qui gère le référentiel modèle.
Content Management Service	Service d'application qui gère les données de référence. Il fournit des informations de données de référence au service d'intégration de données et à Informatica Developer. Vous pouvez utiliser Informatica Developer pour importer des domaines de données dans un référentiel modèle.
Data Integration Service	Service d'applications qui effectue des tâches d'intégration de données pour Enterprise Data Catalog et des applications externes.
Catalog Service	Service d'application qui exécute Enterprise Unified Metadata et gère les connexions entre les composants de service et les applications externes.
Model repository	Base de données relationnelle qui stocke la configuration des ressources et les informations du domaine de données.
Metadata Persistence Store	Base de données intermédiaire qui stocke les métadonnées extraites pour une analyse plus poussée.

Composant	Description
Search Index	Informations d'index de recherche Apache Solr. L'index de recherche est basé sur les ressources du référentiel modèle et les ressources du stockage de persistance des métadonnées. Enterprise Unified Metadata utilise les informations indexées pour afficher les résultats de recherche en fonction des métadonnées de ressources et des relations appropriées.
Data sources and metadata sources	Bases de données source ou sources de métadonnées que les métadonnées unifiées d'entreprise analysent pour extraire les métadonnées pertinentes pour une utilisation ultérieure.

Présentation de l'administration du catalogue

Catalog Administrator est un outil d'administration qui vous permet de gérer et de surveiller les ressources, les planifications, les attributs et les connexions.

Vous pouvez utiliser Catalog Administrator pour effectuer les tâches suivantes :

- Gestion des ressources. Créez, modifiez et supprimez des ressources.
- Gestion de la planification. Créez, modifiez et supprimez des planifications.
- Gestion des attributs. Affichez les attributs définis par le système pour les types d'objet de métadonnées. Créez des attributs personnalisés et attribuez-les aux types d'objets de métadonnées, tels que les tables, les vues et les colonnes.
- Gestion des connexions. Affichez les connexions et les schémas attribués automatiquement. Attribuez des schémas et des connexions aux ressources. Annulez l'attribution des connexions attribuées par l'utilisateur.
- Gestion de la configuration de profil. Créez et modifiez des paramètres de définition de profil réutilisables.
- Surveillance des ressources. Surveillez les ressources et les tâches.
- Gestion du domaine de données. Créez et modifiez des domaines de données et des groupes de domaines de données. Attribuez des domaines de données logiques à des groupes de domaines de données.

Processus d'administration du catalogue

Les tâches d'administration incluent la configuration des ressources, l'attribution des planifications et la personnalisation des attributs. Vous devez également surveiller les tâches qui extraient les métadonnées à l'aide des ressources.

Vous pouvez effectuer les tâches suivantes dans le cadre du processus d'administration :

1. Créez des ressources pour chaque type de ressource en fonction du type de source dont vous avez besoin pour extraire des métadonnées.
2. Indiquez si vous voulez extraire les métadonnées source, les métadonnées de profilage ou les deux.
3. Choisissez si vous voulez exécuter les ressources une ou plusieurs fois en fonction d'une planification commune ou personnalisée.

4. Vous pouvez éventuellement attribuer une planification commune ou une planification personnalisée aux ressources.
5. Surveillez les tâches qui extraient des métadonnées de différentes sources.
6. Définissez des domaines de données basés sur un objet de référentiel modèle prédéfini ou défini par l'utilisateur en fonction de la sémantique des données de colonne ou d'un nom de colonne.
7. Dépannez les tâches qui ne s'exécutent pas comme prévu.

Accès à Catalog Administrator

Utilisez Catalog Administrator pour consolider les tâches administratives des ressources, des attributs et des planifications. Lancez Catalog Administrator à partir d'Informatica Administrator. Vous devez connaître le nom d'hôte du nœud de passerelle et le numéro de port d'Informatica Administrator pour vous connecter à Informatica Administrator.

Effectuez les étapes suivantes avant de vous connecter :

1. Lancez Informatica Administrator à l'aide du nœud de passerelle et du numéro de port Informatica Administrator dans l'URL d'Informatica Administrator.
2. Dans Informatica Administrator, configurez un domaine Informatica, des comptes utilisateurs, des connexions de bases de données et des services si vous ne les avez pas créés lors de l'installation. Les services comprennent le service d'intégration de données, le service de référentiel modèle, le service de gestion de contenu, le service de cluster Informatica et le service de catalogue.
3. Dans l'onglet **Services et nœuds** d'Informatica Administrator, sélectionnez le service de catalogue, puis cliquez sur l'URL du service pour lancer Catalog Administrator à partir d'Informatica Administrator.
4. Utilisez les informations d'identification pour vous connecter à Catalog Administrator.

Prérequis

Les composants requis pour lancer Catalog Administrator incluent un domaine Informatica, des informations de connectivité de domaine et un compte utilisateur administrateur.

Vous devez vérifier les conditions préalables suivantes avant de vous connecter à Catalog Administrator :

1. Le domaine Informatica doit être actif.
2. Le domaine Informatica possède un service d'intégration de données, un service de référentiel modèle et un service de catalogue activé.
3. Vous disposez des informations de connectivité du domaine et le compte utilisateur administrateur dans Informatica Administrator.

Connectez-vous à Catalog Administrator

Utilisez Microsoft Internet Explorer ou Google Chrome pour vous connecter à Catalog Administrator.

1. Démarrez Microsoft Internet Explorer ou Google Chrome.
2. Dans le champ Adresse, entrez l'URL de la page de connexion de Catalog Administrator dans le format suivant :

`http://<host>:<port>/ldmadmin`

L'hôte est le nom d'hôte du nœud de passerelle. Le port représente le numéro de port configuré pour le service de catalogue.

3. Dans la page de connexion de Catalog Administrator, entrez le nom d'utilisateur et le mot de passe.
4. Vérifiez que l'option de domaine par défaut L'option **Natif** est sélectionnée. Vous pouvez également sélectionner un domaine LDAP.

Le champ Domaine s'affiche lorsque le domaine Informatica contient un domaine de sécurité LDAP.

5. Cliquez sur **Connexion**.

Modification du mot de passe

Utilisez le menu Administrateur pour modifier le mot de passe.

1. Dans la zone d'en-tête de l'outil Catalog Administrator, cliquez sur **Administrateur > Modifier le mot de passe**.

La page **Modifier le mot de passe** s'affiche.

2. Entrez le mot de passe actuel dans le champ **Mot de passe** et le nouveau mot de passe dans les champs **Nouveau mot de passe** et **Confirmer le nouveau mot de passe**.
3. Cliquez sur **Mettre à jour**.

CHAPITRE 2

Concepts d'Enterprise Data Catalog

Ce chapitre comprend les rubriques suivantes :

- [Présentation des concepts d'Enterprise Data Catalog, 18](#)
- [Catalogue, 19](#)
- [Type de ressource, 19](#)
- [Ressource, 19](#)
- [Scanner, 19](#)
- [Planification, 20](#)
- [Exemple d'entreprise, 20](#)
- [Domaines de données et groupes de domaines de données, 21](#)
- [Domaines de données composites, 29](#)
- [Découverte de données, 31](#)
- [Similarité de colonne, 41](#)
- [Propagation des termes d'entreprise, 47](#)
- [Business Glossary Synchronisation, 49](#)
- [Aperçu, préparation et provisionnement des données, 50](#)

Présentation des concepts d'Enterprise Data Catalog

Enterprise Data Catalog vous permet d'analyser et de comprendre de grands volumes de métadonnées de l'entreprise. Vous pouvez extraire des métadonnées physiques et opérationnelles pour un grand nombre d'objets, organiser les métadonnées en fonction des concepts d'entreprise et afficher les informations de lignage et de relation de chaque objet.

Les concepts clés d'Enterprise Data Catalog incluent le catalogue, la ressource, le type de ressource, le scanner et la planification. Le catalogue stocke toutes les métadonnées extraites des sources. Le type de ressource représente différents systèmes de sources des métadonnées. La ressource est une représentation d'un type de ressource, tel qu'Oracle, SQL Server ou PowerCenter. Les scanners récupèrent les métadonnées et les sauvegardent dans le catalogue. Les planifications déterminent les intervalles auxquels les scanners extraient les métadonnées des systèmes source et les enregistrent dans le catalogue.

Catalogue

Le catalogue représente un inventaire indexé de toutes les ressources de données de l'entreprise que vous configurez dans Catalog Administrator. Enterprise Data Catalog organise toutes les métadonnées d'entreprise dans le catalogue et permet aux utilisateurs d'applications externes de découvrir et de comprendre les données.

Vous pouvez trouver des métadonnées et des informations statistiques, telles que les résultats de profil, les domaines de données et les relations d'actifs de données, à partir du catalogue. Enterprise Data Catalog stocke les métadonnées extraites de sources externes dans la base de données MongoDB.

Type de ressource

Type de source de données externe ou référentiel de métadonnées à partir duquel les scanners extraient des métadonnées. Les sources relationnelles, les sources de veille stratégique et les sources PowerCenter sont des exemples.

Ressource

Une ressource est un objet de catalogue qui représente une source de données externe ou un référentiel de métadonnées à partir duquel les scanners extraient les métadonnées. Une ressource représente une instance d'un type de ressource spécifique. Les opérations de métadonnées de base, telles que l'extraction et le stockage des métadonnées, sont effectuées au niveau des ressources.

Une ressource peut avoir un type de ressource, comme les bases de données relationnelles, la classification Business Glossary et les sources de veille stratégique. Une ressource peut avoir une planification associée. Chaque ressource peut extraire les métadonnées source et les métadonnées de profil des sources de données externes.

Scanner

Composant enfichable d'Enterprise Data Catalog qui extrait des métadonnées spécifiques, telles que des métadonnées source ou des métadonnées de profil, à partir de sources de données externes et les stocke dans le catalogue. Un scanner correspond généralement à un type de ressource unique. Toutefois, il peut y avoir plusieurs scanners pour un type de ressource. Le profilage du scanner et l'analyseur de lignage sont des exemples.

Un scanner effectue une tâche d'analyse sur les sources de métadonnées afin d'extraire les métadonnées dans le catalogue. Si vous disposez de scanners compatibles avec les nouveaux types de ressources, vous pouvez les brancher à Enterprise Data Catalog sans mettre ce dernier à niveau.

Planification

Une planification détermine le moment où les scanners extraient les métadonnées des sources. Vous pouvez avoir des planifications récurrentes quotidiennes, hebdomadaires et mensuelles pour extraire des métadonnées à intervalles réguliers.

Vous pouvez créer les types de planifications suivants :

Planification globale

Planifications réutilisables que vous pouvez joindre à plusieurs ressources.

Planification personnalisée

Planification personnalisée attribuée à une seule ressource.

Exemple d'entreprise

Vous êtes administrateur de catalogue dans une organisation multinationale de vente au détail. Les analystes de données de votre service doivent afficher les métadonnées de différents schémas et tables de base de données dans plusieurs sources pour effectuer une analyse de données avancée. Vous devez également vous assurer que les analystes de données comprennent et approuvent les données qu'ils utilisent. L'organisation peut planifier des audits de sécurité réguliers pour trouver des données sensibles dans les sources de données et les masquer ou les protéger selon les besoins.

L'organisation de vente au détail pour laquelle vous travaillez dispose des systèmes configurés suivants :

- Système de gestion des ressources humaines mis en place sur une base de données Oracle.
- Système de gestion des commandes mis en place sur la même base de données Oracle.
- Entrepôt de données hébergé sur un référentiel Hadoop. L'entrepôt de données a intégré des informations provenant de plusieurs sources de données.
- PowerCenter pour effectuer des tâches d'intégration de données entre des bases de données et des schémas.
- Système de rapports configuré sur une source SAP BusinessObjects.

L'administrateur de l'organisation peut effectuer les tâches suivantes dans Catalog Administrator pour répondre efficacement aux besoins de l'administration de données dans cet exemple :

- Utiliser Catalog Administrator pour créer une ressource Oracle pour le système de gestion des ressources humaines et une autre ressource Oracle pour le système de gestion des commandes. Vous pouvez configurer les paramètres de métadonnées source pour extraire les métadonnées dans le catalogue. Il se peut que vous n'ayez pas besoin de configurer les paramètres de métadonnées de profilage pour ces ressources. Les ressources fournissent la table de base de données et les objets de colonne source requis dans le catalogue pour analyse.
- Créer une ressource Hive pour l'entrepôt Hadoop. La ressource Hive récupère les tables et les colonnes dans le catalogue. En plus de l'extraction des métadonnées source, vous pouvez configurer les paramètres de métadonnées de profilage afin d'avoir des informations relatives à la qualité des données pour une analyse plus poussée.
- Créer une ressource PowerCenter qui correspond aux exigences d'intégration de données. La configuration des ressources fournit les liens entre les objets de données Oracle et les objets Hive.

- Créer une ressource de SAP BusinessObjects et la configurer pour extraire des métadonnées de rapport. La ressource fournit des métadonnées de rapport basées sur les liens entre les objets d'entreprise et les objets Oracle et Hive.
- Définir une planification récurrente pour chaque ressource afin que les scanners extraient les métadonnées source et de profilage des systèmes source à intervalles réguliers.
- Vérifier périodiquement dans Catalog Administrator les tâches qui extraient les métadonnées. Surveillez les tâches afin d'obtenir une vue fonctionnelle d'Enterprise Data Catalog. La surveillance vous permet également d'analyser et d'estimer le type de contenu que les scanners récupèrent dans le catalogue.

Domaines de données et groupes de domaines de données

Un domaine de données est un objet du référentiel modèle défini par l'utilisateur ou prédéfini qui vous permet de découvrir la signification fonctionnelle des données des colonnes ou des noms de colonnes dans une source de données. Le numéro de sécurité sociale, l'état du compte, l'adresse IP et le code UPC sont des exemples de domaines de données. Vous pouvez ajouter un ou plusieurs domaines de données dans un groupe de domaines de données. Les groupes de domaines de données peuvent contenir des domaines de données et non d'autres groupes de domaines de données. PCI, PHI et PII sont des exemples de groupes de domaines de données.

La découverte de domaines de données est le processus de découverte de la signification fonctionnelle des données dans les sources de données en fonction de la sémantique des données. Vous pouvez utiliser les domaines de données pour identifier et comprendre la signification des données source critiques ou non découvertes afin de pouvoir prendre des mesures, comme le masquage des données, en vue de les traiter efficacement. Par exemple, vous pouvez avoir des systèmes de données héritées qui contiennent des numéros de sécurité sociale dans un champ Commentaires. Avant de transférer ces données vers les nouveaux systèmes de données, vous devez les identifier afin de pouvoir prendre les mesures appropriées.

Les groupes de domaines de données vous permettent de classer les domaines de données dans des groupes spécifiques. Par exemple, vous pouvez regrouper les domaines de données `first_name`, `last_name` et `account_number` dans le groupe de domaines de données Informations médicales personnelles (PHI). Lorsque vous créez ou modifiez un groupe de domaines de données, vous pouvez lui ajouter un ou plusieurs domaines de données. Vous pouvez supprimer un domaine de données du groupe de domaine de données. Un domaine de données peut faire partie de plusieurs groupes de domaines de données. Par exemple, le numéro de sécurité sociale peut appartenir aux groupes de domaines de données Industrie des cartes de paiement (PCI) et PII.

Domaines de données et groupes de domaines de données dans Catalog Administrator

Dans Catalog Administrator, vous pouvez afficher des domaines de données et des groupes de domaines de données, créer et modifier des domaines de données basés sur des règles, modifier des domaines de données intelligents et créer et modifier des groupes de domaines de données.

Dans Catalog Administrator, suite à l'activation de la découverte de domaines de données pour une ressource et l'exécution de la ressource, le scanner de profilage utilise les domaines de données pour déduire les données de colonnes correspondantes ou les modèles de noms de colonnes des métadonnées extraites par les ressources. Dans Enterprise Data Catalog, vous pouvez afficher un domaine de données comme un élément. Vous pouvez également afficher les domaines de données pour des éléments de tableau,

de colonne et de champ. Ces domaines de données sont déduits pour l'élément des résultats de profil ou des colonnes similaires. Vous pouvez organiser un domaine de données déduit en fonction de vos besoins.

Dans Catalog Administrator, vous pouvez afficher l'ensemble des domaines de données et des groupes de domaines de données disponibles dans l'espace de travail Bibliothèque. Vous pouvez afficher ou modifier des domaines de données et afficher, modifier ou supprimer des groupes de domaines de données. Vous pouvez utiliser des filtres pour afficher des domaines de données ou des groupes de domaines de données spécifiques.

Pour créer ou afficher des domaines de données et des groupes de domaines de données, vous devez disposer des autorisations et privilèges appropriés. Configurez les privilèges **Gestion des domaines : Admin - Afficher un domaine et un groupe de domaines** et **Gestion des domaines : Admin - Modifier un domaine et un groupe de domaines** dans Informatica Administrator pour un utilisateur afin d'afficher, de créer, de modifier ou de supprimer des domaines de données ou des groupes de domaines de données dans Catalog Administrator. Pour plus d'informations sur les privilèges et les autorisations, consultez le *Guide de référence Informatica Administrator pour Enterprise Data Catalog*.

Dans Enterprise Data Catalog, vous pouvez utiliser les types suivants de domaines de données :

Domaines de données prédéfinis

Ce sont des domaines de données définis par le système. Ils sont importés automatiquement lors de l'installation d'Enterprise Information Catalog.

Domaines de données basés sur des règles

Ce sont des domaines de données définis par l'utilisateur. Vous pouvez utiliser une ou plusieurs règles pour créer un domaine de données de ce type dans Catalog Administrator.

Vous pouvez également créer des domaines de données et des groupes de domaines de données dans Informatica Analyst ou Informatica Developer et les utiliser dans Enterprise Data Catalog. Vous pouvez créer des règles de maplet dans l'outil Developer tool et les utiliser pour créer des domaines de données dans Catalog Administrator. Pour utiliser ces domaines de données, Enterprise Data Catalog lance le scanneur de domaine de données qui est planifié pour s'exécuter automatiquement toutes les 45 minutes. Le scanneur synchronise les domaines de données entre le référentiel modèle et le catalogue.

Domaines de données intelligents

Les domaines de données intelligents sont des domaines de données définis par l'utilisateur. Vous pouvez créer un domaine de données intelligent avec un modèle existant d'une colonne ou d'un champ dans le catalogue. Ce domaine de données ne contient aucune règle. Vous pouvez afficher les domaines de données intelligents dans Catalog Administrator. Pour convertir un domaine de données intelligent en domaine de données basé sur des règles, vous pouvez ajouter une ou plusieurs règles, valeurs de conformité et domaines de données de proximité.

Domaines de données prédéfinis

Les domaines de données et les groupes de domaines de données prédéfinis sont des domaines de données et des groupes de domaines de données définis par le système qui sont disponibles par défaut dans Enterprise Data Catalog. Les domaines de données sont importés automatiquement lors de l'installation d'Enterprise Information Catalog. Les domaines de données prédéfinis utilisent des valeurs de conformité par défaut. Dans Catalog Administrator, vous pouvez modifier les domaines de données prédéfinis afin de remanier les valeurs de conformité par défaut et d'ajouter des domaines de données de proximité.

Les domaines de données prédéfinis utilisent la règle de données, la règle de nom de colonne ou les règles de données et de nom de colonne. AccountNumber, India_NationalID, SSN_General et ZipCode sont des exemples de domaines de données prédéfinis. Après avoir installé Enterprise Data Catalog, vous pouvez choisir d'importer les domaines de données prédéfinis soit dans l'outil Developer, soit en exécutant la commande `infacmd.sh oie importobjects`

Domaines de données basés sur des règles

Vous pouvez utiliser une ou plusieurs règles pour définir un domaine de données basé sur des règles. Dans Catalog Administrator, vous pouvez créer et modifier des domaines de données basés sur des règles. Lorsque vous activez la découverte de domaines de données sur les ressources, Enterprise Data Catalog utilise les domaines de données pour découvrir les données de colonnes correspondantes ou les modèles de noms de colonnes des métadonnées extraites par les ressources. Dans Enterprise Data Catalog, vous pouvez afficher des éléments de domaines de données et des domaines de données dans des éléments de tableau, de colonne et de champ.

Lorsque vous créez ou modifiez un domaine de données, vous pouvez ajouter des règles, choisir des options de résolution de conflit, configurer des valeurs de conformité, et ajouter des domaines de données de proximité et des groupes de domaines de données. Vous pouvez organiser les domaines de données dans Enterprise Data Catalog.

Règles

Vous pouvez utiliser les règles qui ont une sortie unique avec des données de type Integer. Si vous n'utilisez pas de règle dans un domaine de données, Enterprise Data Catalog attribue le domaine de données à des colonnes similaires en fonction du domaine de données attribué à la colonne.

Choisissez l'une des règles suivantes ou les deux afin de définir un modèle de données pour un domaine de données :

- Règle de données. Utilisez les données source qui correspondent aux métadonnées. La règle découvre des colonnes contenant des données qui correspondent à une logique spécifique définie dans la règle.
- Règle de nom de colonne. Utilisez les modèles de nom de colonne qui correspondent aux métadonnées. La règle découvre les colonnes qui correspondent à la logique de nom de colonne définie dans la règle.

Lorsque vous créez ou modifiez un domaine de données, choisissez l'une des règles suivantes comme règle de données ou règle de nom de colonne :

Règle	Description
Table de référence	Choisissez la table de référence dans le référentiel modèle. Lorsque vous choisissez une table de référence, la règle utilise les données de colonne de la table de référence pour découvrir les domaines de données.
Expression régulière	Spécifiez une expression régulière pour qu'elle corresponde aux données de chaîne. Une expression régulière est une formule spécialisée pour les chaînes de texte de correspondance qui suivent un modèle. La syntaxe utilisée pour l'expression régulière doit être conforme aux règles de grammaire et aux jetons décrits ici : https://docs.informatica.com/data-catalog/shared-content-for-data-catalog/10-5/transformation-language-reference/functions/reg_extract.html .
Règles existantes	Choisissez des règles existantes dans le référentiel modèle. Vérifiez que vous disposez des autorisations appropriées pour afficher les règles avant d'en choisir une. Vous pouvez choisir des règles basées sur des mapplets qui ont été créées à l'aide d'Informatica Analyst ou d'Informatica Developer. Pour plus d'informations sur l'utilisation de règles basées sur des mapplets, reportez-vous à la section https://docs.informatica.com/data-engineering/shared-content-for-data-engineering/10-5/developer-mapping-guide/maplets/maplet-validation/maplet-as-a-rule-validation.html .

Résolution de conflit

Vous pouvez décider d'utiliser une règle de données ou une règle de nom de colonne pour configurer des critères d'acceptation automatique. Sélectionnez l'option **Résolution de conflit** pour accepter

automatiquement le domaine de données selon la règle de données ou de nom de colonne. Vous pouvez choisir l'une des options suivantes :

- En cas de conflit de règles, la règle de données remplace la règle de nom de colonne. Permet à Enterprise Data Catalog d'accepter automatiquement les domaines de données selon la règle de données. Si aucune règle de données n'est spécifiée, les résultats de la règle de nom de colonne sont utilisés pour accepter automatiquement les domaines de données.
- En cas de conflit de règles, la règle de nom de colonne remplace la règle de données. Permet à Enterprise Data Catalog d'accepter automatiquement les domaines de données selon la règle de nom de colonne. Si aucune règle de nom de colonne n'est spécifiée, les résultats de la règle de données sont utilisés pour accepter automatiquement les domaines de données.
- Faire correspondre la règle de données et de nom de colonne. Enterprise Data Catalog accepte automatiquement les domaines de données lorsque les résultats de la règle de données et de la règle de nom de colonne correspondent. Si aucune règle de données n'est spécifiée, les domaines de données ne sont pas acceptés automatiquement.
- Faire correspondre la règle de données et de nom de colonne. Enterprise Data Catalog accepte automatiquement les domaines de données lorsque les résultats de la règle de données ou de la règle de nom de colonne correspondent. Si aucune règle de données ou de nom de colonne n'est spécifiée, les domaines de données ne sont pas acceptés automatiquement.

Si vous ne sélectionnez pas l'option **Résolution de conflit**, Enterprise Data Catalog applique les résultats de la règle de conflit pour accepter automatiquement les domaines de données. Les domaines de données ne sont pas acceptés automatiquement si vous configurez **Critères de correspondance de domaine de données** sur **Ligne** dans la section **Ressource** de Catalog Administrator.

Les domaines de données sont acceptés automatiquement dans les scénarios suivants :

La règle de données remplace la règle de nom de colonne.

Correspondance de règle de nom de colonne	Correspondance de règle de données	État de domaine de données
Oui	Non	Non
Oui	Oui, atteint la valeur de conformité minimale configurée.	Inféré
Oui	Oui	Accepté automatiquement
Non	Non	Non
Non	Oui, atteint la valeur de conformité minimale configurée.	Inféré
Non	Oui	Accepté automatiquement
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Non	Non
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Oui, atteint la valeur de conformité minimale configurée.	Inféré

Correspondance de règle de nom de colonne	Correspondance de règle de données	État de domaine de données
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Oui	Accepté automatiquement
Oui	La découverte de domaine sur les métadonnées source n'est pas effectuée.	Accepté automatiquement
Non	La découverte de domaine sur les métadonnées source n'est pas effectuée.	Non

La règle de nom de colonne remplace la règle de données.

Correspondance de règle de nom de colonne	Correspondance de règle de données	État de domaine de données
Oui	Non	Accepté automatiquement
Oui	Oui, atteint la valeur de conformité minimale configurée.	Accepté automatiquement
Oui	Oui	Accepté automatiquement
Non	Non	Non
Non	Oui, dépasse la valeur de seuil d'acceptation automatique configurée.	Non
Non	Oui	Non
Non	La découverte de domaine sur les données source n'est pas effectuée.	Non
Oui	La découverte de domaine sur les données source n'est pas effectuée.	Accepté automatiquement
La découverte de domaine sur les données source n'est pas effectuée.	Oui	Accepté automatiquement
La découverte de domaine sur les données source n'est pas effectuée.	Oui, dépasse la valeur de seuil d'acceptation automatique configurée.	Inféré
La découverte de domaine sur les données source n'est pas effectuée.	Non	Non

Faire correspondre la règle de données et de nom de colonne.

Correspondance de règle de nom de colonne	Correspondance de règle de données	État de domaine de données
Oui	Oui, dépasse la valeur de seuil d'acceptation automatique configurée.	Accepté automatiquement
Oui	Oui, atteint la valeur de conformité minimale configurée.	Inféré
Oui	Non	Non
Non	Non	Non
Non	Oui, atteint la valeur de conformité minimale configurée.	Non
Non	Oui	Non
Non	La découverte de domaine sur les métadonnées source n'est pas effectuée.	Non
Oui	La découverte de domaine sur les métadonnées source n'est pas effectuée.	Accepté automatiquement
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Oui	Accepté automatiquement
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Non	Non
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Oui, atteint la valeur de conformité minimale configurée.	Inféré

Faire correspondre la règle de données et de nom de colonne.

Correspondance de règle de nom de colonne	Correspondance de règle de données	État de domaine de données
Oui	Non	Accepté automatiquement
Oui	Oui, atteint la valeur de conformité minimale configurée.	Accepté automatiquement
Oui	Oui	Accepté automatiquement
Non	Oui, atteint la valeur de conformité minimale configurée.	Inféré
Non	Oui	Accepté automatiquement

Correspondance de règle de nom de colonne	Correspondance de règle de données	État de domaine de données
Non	La découverte de domaine sur les métadonnées source n'est pas effectuée.	Non
Oui	La découverte de domaine sur les métadonnées source n'est pas effectuée.	Accepté automatiquement
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Oui	Accepté automatiquement
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Non	Non
La découverte de domaine sur les métadonnées source n'est pas effectuée.	Oui, atteint la valeur de conformité minimale configurée.	Inféré

Domaines de données de proximité

Enterprise Data Catalog utilise les domaines de données de proximité pour affiner les résultats inférés afin d'identifier les colonnes ou champs très similaires pour un domaine de données. Les domaines de données de proximité sont un des facteurs qui permettent de déterminer le type de données dans une colonne. Lorsqu'un ou plusieurs domaines de données sont inférés avec une conformité ou une probabilité égale, vous pouvez utiliser les domaines de données de proximité pour identifier le domaine de données presque identique. Lorsque vous créez ou modifiez un domaine de données, ajoutez un ou plusieurs domaines de données en tant que domaines de données de proximité.

Enterprise Data Catalog affiche les résultats sous la forme d'un score de correspondance pour le domaine de données. Le score de correspondance est le rapport entre les domaines de données proximaux découverts dans la source de données et le nombre de domaines de données proximaux configurés pour un domaine de données inféré.

Lorsque vous activez la découverte de données, ajoutez des domaines de données et exécutez la ressource, le scanneur de profilage analyse la source de données pour le domaine de données et les domaines de données de proximité dans la ressource. Il affiche ensuite un score de correspondance dans Enterprise Data Catalog.

Exemple

Vous exécutez un profil sur un fichier CSV qui comprend les colonnes suivantes :

- ID. Contient des nombres de sept chiffres.
- Fname. Contient les prénoms.
- Lname. Contient les noms.
- Dname. Contient les noms de département.
- Addr1. Contient les adresses.
- Phone. Contient les numéros de téléphone.

Vous souhaitez déterminer si la colonne ID fait référence au domaine *EMPID*, *NODEID* ou *CUSTID* qui contiennent tous sept chiffres.

Lorsque vous analysez le fichier, vous déterminez que la colonne ID comprend des ID d'employés, car la table comprend des colonnes de prénoms, de noms et de départements. Les noms de département s'affichent habituellement dans les tables relatives aux employés. La colonne Dname est une preuve que la colonne ID est EMPID, et non NODEID ni CUSTID.

Dans Enterprise Data Catalog, vous pouvez ajouter des domaines de données de proximité afin de déterminer le type de données dans la colonne ID.

Pour identifier les données dans la colonne ID, effectuez les tâches suivantes dans Catalog Administrator :

1. Créez les domaines de données *DeptName* et *EMPID*.
2. Ajoutez les domaines de données *FirstName*, *LastName* et *DeptName* en tant que domaines de données de proximité au domaine de données *EMPID*.
3. Modifiez la ressource avec le fichier CSV, puis ajoutez le domaine de données *EMPID*.
4. Exécutez la ressource.

Lors de l'exécution de la ressource, la colonne ID est identifiée en tant que *EMPID*, *NODEID* et *CUSTID* en raison des scores de conformité, mais les domaines de données de proximité déterminent et infèrent que la colonne ID est *EMPID*. Le catalogue affiche la colonne ID en tant que *EMPID*.

Critères de conformité

Lorsque vous choisissez une règle de données pour un domaine de données, vous pouvez configurer les critères de conformité du domaine de données. Enterprise Data Catalog identifie les domaines de données en fonction des critères de conformité.

Vous pouvez configurer les critères de conformité suivants pour un domaine de données :

- Conformité minimale. Entrez un pourcentage minimal de conformité pour le domaine de données. Par défaut, cette option est définie sur la valeur 40.
- Accepter automatiquement si supérieur à. Entrez un pourcentage. Enterprise Data Catalog accepte automatiquement le domaine de données si la correspondance de domaine de données dépasse le pourcentage configuré. Par défaut, cette option est définie sur la valeur 80.00 pour cent.
- Nombre de lignes. Entrez le nombre de lignes minimales pour la conformité des données. Par défaut, cette option est définie sur 1.

Les valeurs par défaut sont appelées valeurs de conformité prédéfinies. Vous pouvez configurer une valeur personnalisée pour remplacer les valeurs prédéfinies. Lorsque vous choisissez plusieurs domaines de données ou groupes de domaines de données, Enterprise Data Catalog calcule la valeur de conformité en fonction des valeurs prédéfinies ou personnalisées.

Lorsque vous choisissez une règle de nom de colonne pour un domaine de données, vous pouvez définir la valeur de l'option **Conformité minimale** ou **Accepter automatiquement si supérieur à** sur 0 ou 100. La règle de nom de colonne n'utilise pas la valeur de l'option **Nombre de lignes** pour calculer la correspondance de domaine de données.

Lorsque vous créez ou modifiez une ressource, vous pouvez choisir les domaines de données, ses valeurs de conformité configurées ou par défaut, ou choisir des valeurs de conformité personnalisées comme critères de conformité. Vous pouvez également choisir d'exclure les valeurs NULL pendant le calcul des domaines de données.

Domaines de données composites

Un domaine de données composite est une collection de domaines de données ou d'autres domaines de données composites liés à l'aide de règles. Un domaine de données composite vous permet de rechercher les informations requises sur une entité dans plusieurs schémas définis pour la base de données. Une entité fait référence à un sujet particulier pour lequel vous avez besoin de toutes les informations associées. Par exemple, pour une entité telle que Client, vous pouvez rechercher toutes les informations associées telles que le nom du client, l'âge, l'emplacement et le numéro de contact. Les informations peuvent être réparties entre plusieurs tables et colonnes dans la base de données. Les domaines de données composites vous permettent de définir une requête de recherche qui inclut les informations sur l'entité que vous voulez rechercher à partir de plusieurs schémas définis pour la base de données.

Remarque: Vous devez configurer une relation clé primaire-clé étrangère entre des tables si vous voulez rechercher des informations sur l'entité dans plusieurs schémas.

Un domaine de données composite vous permet de définir des modèles de recherche basés sur une combinaison de domaines de données ou de domaines de données composites existants. Les domaines de données de l'exemple peuvent être le nom du client, son âge, son emplacement et d'autres informations associées à l'entité *client*. Dans une définition de domaine de données composite, vous pouvez définir des règles permettant de lier des domaines de données existants ou d'autres domaines de données composites. Enterprise Data Catalog lie plusieurs règles à l'aide des opérateurs OR et lie les domaines de données ou les domaines de données composites dans chaque règle à l'aide des opérateurs AND. Assurez-vous de créer des domaines de données avant de créer des domaines de données composites.

Après avoir créé des domaines de données composites et activé la découverte de domaines de données composites pour les ressources, vous pouvez rechercher l'entité à l'aide de domaines de données composites dans Enterprise Data Catalog. Enterprise Data Catalog recherche les ressources associées à l'entité dans le catalogue.

Exemple d'entreprise

Sophie, gestionnaire de données dans une chaîne de magasins, est invitée à trouver des informations spécifiques sur les clients à partir des informations de clients stockées dans la chaîne de magasins. Les informations de clients se trouvent dans plusieurs tables de la base de données de la chaîne de magasins.

Les tables de la base de données incluent les informations suivantes :

- ID du client
- Prénom
- Nom
- E-mail
- Numéro de téléphone mobile
- Rue
- Ville
- Informations sur les comptes de médias sociaux
- Code postal
- Pays

À partir des informations, Sophie doit trouver l'une des informations suivantes susceptible de lui permettre d'identifier l'emplacement du client :

- Rue et ville
- Code postal

- Pays

Le tableau suivant répertorie les étapes que Sophie doit effectuer pour collecter les informations requises à propos du client :

Scénario	Résolution d'écran
Condition préalable à la création de domaines de données composites	Sophie doit définir des domaines de données basés sur les ressources existantes du catalogue. Sophie crée les domaines de données suivants en fonction des informations sur le client qu'elle veut collecter : <ul style="list-style-type: none"> - Rue - Ville - Code_postal - Pays
Créer des domaines de données composites	Sophie crée un domaine de données composite appelé <code>customer_location</code> et définit les règles en liant des domaines de données à l'aide des opérateurs AND et OR comme indiqué dans la liste suivante : <ul style="list-style-type: none"> - Rue AND ville OR - Code_postal OR - Pays
Identifier les ressources et activer la découverte de domaines de données composites	Sophie identifie les ressources requises en effectuant une recherche dans Enterprise Data Catalog basée sur les domaines de données composites créés. Sophie active ensuite la découverte de domaines de données composites pour les ressources identifiées.
Rechercher les informations requises sur les clients	Sophie peut effectuer une recherche en utilisant le domaine de données composite <code>customer_location</code> dans Enterprise Data Catalog. La recherche récupère une liste de ressources et de tables qui incluent les informations sur les clients exigées par la chaîne de magasins.

Par ailleurs, Sophie peut créer les trois domaines de données composites suivants (`Street_county`, `Zip_code` et `State`) et inclure les domaines de données composites dans une définition de domaine de données composites, appelée `customer_location` :

- Rue_ville Rue AND ville
- Code postal.
- Pays.

La définition du domaine de données composite `customer_location` est la suivante :

- Rue_ville
- OR
- Code_postal
- OR
- Pays

Flux de travail de la découverte de domaines de données composites

Le flux de travail de la découverte de domaines de données composites inclut les étapes suivantes :

1. Synchronisation des domaines de données composites entre le référentiel modèle et le catalogue.
2. Identification de la liste des domaines de données composites configurés pour une ressource.
3. Identification de la liste des domaines de données requis pour chaque domaine de données composites.
4. Récupération des résultats de la découverte de domaines de données pour les domaines de données de la ressource actuelle.
5. Préparation d'expressions pour chaque domaine de données composite et des résultats de la découverte de domaines de données pour chaque domaine composite.
6. Utilisation de la logique machine interne pour évaluer les expressions et les résultats.
7. Publication des résultats de la recherche et de la découverte dans le catalogue.

Remarque: Enterprise Data Catalog effectue les étapes répertoriées dans le flux de travail pendant la découverte de domaines de données composites.

Découverte de données

Lorsque vous activez la découverte de données pour une ressource et que vous analysez la ressource, Enterprise Data Catalog identifie les métadonnées liées au profilage, telles que les valeurs Null, les valeurs distinctes, les types de données déduites, les clés uniques et les domaines de données dans la ressource.

Lorsque vous créez une ressource, choisissez l'option **Paramètres de chargement des métadonnées > Découverte de données > Activer la découverte de données** pour découvrir les métadonnées liées au profilage et les clés uniques. En fonction de vos besoins, vous pouvez configurer les paramètres de connexion de domaine, les paramètres de profil de base, les paramètres d'inférence de clé unique et les paramètres de similarité de colonne.

Important: Le profilage peut échouer avec une erreur de licence si vous exécutez la découverte de domaine de données à l'aide d'un service d'intégration de données dans une grille.

Types de ressources pris en charge pour la découverte de données

Vous pouvez activer la découverte de données pour les types de ressources suivants :

- Nuage
 - Amazon Redshift
 - Amazon S3
 - Microsoft Azure Data Lake Store
 - Microsoft Azure SQL Data Warehouse
 - Microsoft Azure SQL Server
 - Microsoft Azure Blob Storage
 - Salesforce
 - Google BigQuery

- Snowflake
- Data Engineering
 - HDFS
 - Hive
- Gestion de fichiers
 - Système de fichiers. Les protocoles pris en charge incluent les protocoles SFTP, SMB/CIFS et le protocole de fichier local.
 - OneDrive
 - SharePoint
- Gestion de la base de données
 - IBM DB2
 - IBM DB2 for z/OS
 - IBM Netezza
 - JDBC
 - Microsoft SQL Server
 - Oracle
 - Sybase
 - Teradata
- Application
 - SAP R/3
 - Ressources SAP S/4HANA
- Aucun SQL
 - Apache Cassandra

Vous pouvez également exécuter la découverte de données sur les types de fichiers structurés, les types de fichiers non structurés et les formats non structurés étendus.

Types de fichiers structurés

La découverte de données dans Enterprise Data Catalog prend en charge les types de fichiers structurés suivants :

- Avro. Le type d'extension pris en charge est .avro.
Ce type de fichier est disponible pour une ressource HDFS et une ressource du système de fichiers. Pour une ressource du système de fichiers, vous pouvez choisir uniquement le protocole de fichier local.
- Délimité et texte
- JSON
- Parquet.
Ce type de fichier est disponible pour les ressources HDFS, Amazon S3, Azure Data Lake Store Gen2 et les ressources du système de fichiers. Pour une ressource du système de fichiers, vous pouvez choisir uniquement le protocole de fichier local.
- XML

Types de fichiers non structurés et formats non structurés étendus

Lorsque vous choisissez HDFS, Amazon S3 ou Système de fichiers en tant que ressource, vous pouvez choisir des formats non structurés étendus ou des types de fichiers non structurés. Les formats non structurés étendus incluent les formats MP3, MP4, BMP et JPG. Les formats non structurés étendus ne font pas partie des types de fichiers structurés ou non structurés.

La découverte de domaine de données dans Enterprise Data Catalog prend en charge les types de fichiers non structurés suivants :

- Fichiers Apple. Les types d'extension pris en charge incluent les fichiers .key, .pages, .numbers, .ibooks et .ipa.
- Fichiers compressés. Les types d'extension pris en charge incluent les fichiers .gz, .tgz et .emz.
- E-mail. Les types d'extension pris en charge incluent les fichiers .eml, .emlx et .mime.
- Microsoft Excel
- Microsoft PowerPoint
- Microsoft Word
- Les fichiers Open Office. Les types d'extension pris en charge incluent les fichiers .odt, .ott, .odm, .ods, .ots, .odp, .odg, .otp, .odg, .otg, et .odf.
- PDF
- Fichiers Webpage. Les types d'extension pris en charge incluent les fichiers .chm, .oth et .xhtml.
- CLOB Pris en charge sur les sources de données DB2 et DB2 z/OS.

Paramètres de connexion au domaine

Dans la section **Paramètres de connexion au domaine**, vous pouvez configurer les propriétés du service d'intégration de données. Après avoir configuré les propriétés, le service d'intégration de données exécute le profil, effectue la découverte de domaines de données et déduit la similarité de colonne pour la ressource. Vous pouvez choisir un service d'intégration de données différent pour déduire la similarité de colonne.

Sélectionnez l'une des options suivantes pour configurer les propriétés du service d'intégration de données :

Personnalisé

Configurez les paramètres suivants du service d'intégration de données :

Nom du domaine

Nom du domaine Informatica.

Service d'intégration de données

Nom du service d'intégration de données pour exécuter les profils.

Nom d'utilisateur

Nom d'utilisateur employé par le service d'intégration de données pour accéder au service de référentiel modèle.

Mot de passe

Mot de passe de l'utilisateur du référentiel modèle.

Domaine de sécurité

Nom du domaine de sécurité auquel l'utilisateur du domaine est rattaché.

Hôte

Nom d'hôte du nœud principal de passerelle.

Port

Numéro de port du nœud principal de passerelle.

Profil de système d'exploitation

Choisissez un profil de système d'exploitation si vous ne disposez pas d'un profil de système d'exploitation par défaut. Vous devez attribuer un profil de système d'exploitation aux utilisateurs Enterprise Data Catalog si le service d'intégration de données utilise des profils de système d'exploitation. Si vous n'attribuez pas de profil de système d'exploitation à un utilisateur et que celui-ci tente d'exécuter un profil dans Catalog Administrator, l'exécution du profil échoue. Le service d'intégration de données se sert des informations d'identification de l'utilisateur du profil de système d'exploitation pour exécuter la découverte de données. La découverte de données englobe les profils des colonnes et les profils de découverte des domaines de données.

Si vous disposez de plusieurs profils de système d'exploitation mais que vous n'avez pas de profil de système d'exploitation par défaut, choisissez un profil de système d'exploitation dans la section

Paramètres de connexion de domaine.

Dans l'onglet **Informatica Administrator** > **Sécurité** > **Profils de système d'exploitation**, vous pouvez créer ou attribuer un profil de système d'exploitation. Pour configurer un profil de système d'exploitation par défaut pour un utilisateur, cliquez sur l'option **Attribuer ou modifier le profil de système d'exploitation par défaut** dans la section **Sécurité** > **Utilisateurs** > **Autorisations** > **Profils de système d'exploitation**.

Remarque: Vous ne pouvez choisir un profil de système d'exploitation que si vous choisissez l'option **Personnalisé** pour la propriété **Spécifiez les paramètres de configuration du service d'intégration de données**. Si vous choisissez l'option **Global** pour la propriété **Spécifiez les paramètres de configuration du service d'intégration de données**, vous devez disposer d'un profil de système d'exploitation par défaut pour exécuter le profil.

Pour plus d'informations sur la configuration du service d'intégration de données pour qu'il utilise les profils de système d'exploitation, consultez le *Guide du service d'application Informatica*. Pour plus d'informations sur la création et l'attribution de profils de système d'exploitation, consultez le *Guide de sécurité Informatica*.

Global

Choisissez cette option pour activer une configuration réutilisable.

Une configuration réutilisable comporte des paramètres de service d'intégration de données que vous pouvez utiliser pour une ressource afin d'extraire des métadonnées de profil. Vous pouvez configurer une ou plusieurs configurations réutilisables. Accédez à **Gérer** > **Configuration réutilisable** pour afficher, créer ou supprimer une configuration réutilisable.

Paramètres de profil de base

Dans la section **Paramètres de profil de base**, vous pouvez configurer les options suivantes pour une ressource :

Option d'exécution de profil

Choisissez l'une des options d'exécution de profil suivantes pour que le scanneur de profilage exécute la tâche de profil sur la ressource :

Profil de colonne

Identifie le nombre de valeurs NULL, de valeurs distinctes et de valeurs non distinctes, et déduit les modèles de données et les types de données des colonnes de la ressource.

Découverte de domaines de données

Découvre tous les domaines de données associés à une colonne en fonction de la valeur ou du nom de la colonne.

Profil de colonne et découverte de domaines de données

Identifie le nombre de valeurs NULL, de valeurs distinctes et de valeurs non distinctes, et déduit les modèles de données, les types de données et les domaines de données de la ressource.

Lorsque vous exécutez plusieurs fois une analyse sur une ressource, les derniers résultats de l'analyse incluent toutes les analyses. Par exemple, vous choisissez le profil de colonne lorsque vous analysez une ressource. Ensuite, avant de réexécuter l'analyse, vous choisissez d'effectuer la découverte de domaines de données. Les résultats de la deuxième analyse incluent à la fois les résultats du profil de colonne et les résultats de la découverte de domaines de données.

Les résultats de la découverte de domaines de données affichent tous les domaines de données déduits de toutes les exécutions. Par exemple, si le domaine de données D1 est déduit lors de la première analyse de ressource et que le domaine de données D4 est déduit lors de l'analyse suivante, les résultats de la deuxième analyse afficheront D1 et D4.

Lorsque vous exécutez une analyse sur une ressource pour la deuxième fois ou plus, vous pouvez éventuellement exécuter uniquement la découverte de données sur la source. Pour exécuter uniquement la découverte de données sur la ressource, désactivez l'option **Paramètres de chargement des métadonnées > Métadonnées source**.

Type de découverte de domaines de données

Choisissez l'une des options suivantes pour que le scanneur de profilage déduise des domaines de données selon le nom de colonne, les données de colonne ou les deux :

Exécuter la découverte sur les données source

Exécute la découverte de domaines de données sur les données source.

Exécuter la découverte sur les métadonnées source

Exécute la découverte de domaines de données sur les noms de colonnes.

Exécuter la découverte sur les métadonnées et les données source

Exécute la découverte de domaines de données sur les métadonnées et les données source.

Exécuter la découverte sur les données source où les métadonnées correspondent

Exécute la découverte de domaines de données sur les métadonnées source pour identifier les noms de colonnes qui correspondent aux domaines de données. Le scanneur exécute ensuite la découverte de domaines de données sur les données source des colonnes identifiées.

Remarque: Vous pouvez choisir uniquement l'option **Exécuter la découverte sur les données sources** lorsque vous exécutez la découverte de domaine de données sur des sources de données non structurées.

Priorité

Choisissez l'une des valeurs suivantes pour que le scanneur de profilage hiérarchise l'exécution de ressource :

- Élevé
- Faible

Le scanneur de profilage exécute les ressources ayant la priorité Élevé, puis celles ayant la priorité Faible.

Par exemple, vous disposez de trois ressources : R1, R2 et R3. La priorité définie pour R1 et R3 est Élevé tandis que celle de R2 est Faible. Lors de l'exécution des ressources, le scanner exécute d'abord R1 et R3, puis R2.

Option d'échantillonnage

Choisissez l'une des options d'échantillonnage suivantes pour déterminer le nombre de lignes sur lesquelles exécuter la tâche de profil :

Toutes les lignes

Exécute le profil sur toutes les lignes de la source de données.

Lignes aléatoires automatiques

Exécute le profil sur un échantillon aléatoire de lignes. Enterprise Data Catalog calcule le nombre de lignes aléatoires en fonction du nombre de lignes source.

N lignes aléatoires

Exécute le profil sur le nombre configuré de lignes aléatoires.

Dans le champ **Lignes d'échantillonnage aléatoire**, entrez le nombre de lignes sur lesquelles vous souhaitez exécuter le profil.

premières lignes N

Exécute le profil sur les N premières lignes de la ressource.

Dans le champ **Nombre des N premières lignes d'échantillonnage**, entrez le nombre de lignes sur lesquelles exécuter le profil.

Limite de N lignes

Exécute le profil basé sur le nombre de lignes dans l'objet de données.

Dans le champ **Nombre de lignes à limiter**, entrez le nombre de lignes sur lesquelles exécuter le profil.

Pourcentage aléatoire

Exécute le profil sur un pourcentage de lignes dans l'objet de données.

Dans le champ **Pourcentage aléatoire**, entrez le nombre de lignes sur lesquelles exécuter le profil.

Exclure les vues

Choisissez l'option Exclure les vues si vous ne souhaitez pas que le scanneur de profilage analyse les vues dans les sources de données relationnelles.

Profilage incrémentiel

Choisissez cette option pour exécuter le profil uniquement pour les modifications apportées à la source de données. Si vous ne sélectionnez pas cette option, le profil s'exécute sur l'intégralité de la source de données.

Enterprise Data Catalog prend en charge le profilage incrémentiel pour les ressources suivantes :

- Oracle. Découvre les modifications apportées aux données et métadonnées pendant le profilage incrémentiel.
- Microsoft SQL Server. Découvre les modifications apportées aux données et métadonnées pendant le profilage incrémentiel.
- Système de fichiers avec le protocole **Fichier local**. Découvre les modifications apportées aux données et métadonnées pendant le profilage incrémentiel.
- HDFS Découvre les modifications apportées aux métadonnées pendant le profilage incrémentiel.
- Amazon S3. Découvre les modifications apportées aux données pendant le profilage incrémentiel.
- Snowflake. Découvre les modifications apportées aux données pendant le profilage incrémentiel.

Lorsque vous activez le profilage incrémentiel pour une ressource contenant une table et que vous exécutez plusieurs fois le profil sur la ressource, le scanneur de profilage valide le profil et l'exécute chaque fois sur la même table.

Filtre de profil de données

Vous pouvez inclure ou exclure des tables et des vues depuis l'exécution du profil. Utilisez des points-virgules (;) pour séparer les noms des tables et les noms des vues.

Pour plus d'informations sur le champ Filtre de profil de données, voir la rubrique ["Filtre de profil des données et métadonnées source" à la page 71](#).

Cumulatif

Enterprise Data Catalog ne conserve pas les résultats de l'analyse précédente. Seuls les résultats de l'analyse la plus récente sont affichés. Pour conserver les résultats de profil de l'exécution précédente dans les derniers résultats d'analyse, choisissez l'option **Cumulatif**. Si vous ne choisissez pas cette option, les résultats de profil de colonne et de similarité de colonne obtenus lors de l'exécution précédente sont supprimés et seuls les résultats les plus récents s'affichent dans Enterprise Data Catalog.

Les scénarios d'utilisation suivants expliquent l'incidence de l'option **Cumulatif** associée au champ **Filtre de profil de données** et à l'option **Profilage incrémentiel** sur les résultats du profilage :

- Option **Cumulatif** associée au champ **Filtre de profil de données**
 - Vous exécutez une ressource après avoir entré les noms de tables et les noms de vues dans le champ **Filtre de profil de données**, puis vous choisissez l'option **Cumulatif**.
Dans ce scénario, le scanneur conserve les résultats précédents, ajoute les résultats les plus récents et affiche les résultats du profil consolidés dans Enterprise Data Catalog.
 - Vous exécutez la ressource après avoir entré les noms de tables et les noms de vues dans le champ **Filtre de profil de données**, mais vous ne choisissez pas l'option **Cumulatif**.
Dans ce scénario, les résultats de profil précédents, à l'exclusion des résultats de la découverte de domaines de données, sont supprimés et les résultats de profil les plus récents s'affichent dans Enterprise Data Catalog.
- Option **Cumulatif** associée à l'option **Profilage incrémentiel**
 - Vous exécutez une ressource après avoir choisi l'option **Profilage incrémentiel**.
Dans ce scénario, le scanneur conserve les résultats de profil précédents, que vous choisissiez l'option **Cumulatif** ou pas. Enterprise Data Catalog affiche les résultats du profil consolidés.
- Vous ne choisissez pas les options **Cumulatif** et **Profilage incrémentiel**
 - Vous exécutez une ressource sans choisir les options **Cumulatif** et **Profilage incrémentiel**.
Dans ce scénario, les résultats précédents, à l'exclusion des résultats de la découverte de domaines de données, sont supprimés lors de l'exécution de profil suivante. Enterprise Data Catalog affiche les résultats du profil les plus récents.

Nom de la connexion source

Choisissez la connexion source pour exécuter la découverte de données. Vous pouvez créer les connexions dans Informatica Administrator.

Remarque: Ce paramètre est facultatif pour une ressource du système de fichiers.

Exécuté le

Choisissez l'un des environnements d'exécution suivants pour exécuter le profil :

Blaze

Exécute le profil dans l'environnement Hadoop sur le moteur Blaze.

Spark

Exécute le profil dans l'environnement Hadoop sur le moteur Spark.

Native

Exécute le profil sur la même machine sur laquelle le service d'intégration de données s'exécute.

Databricks

Exécute le profil dans l'environnement Hadoop sur le moteur Spark dans le cluster Databricks.

L'environnement d'exécution Databricks prend en charge les ressources JDBC et Azure Data Lake Store.

Remarque: Choisissez Blaze ou Natif comme environnement d'exécution pour exécuter le travail de profil pour toutes les ressources, à l'exception des ressources Hive. Lorsque vous choisissez le moteur Blaze ou Spark, sélectionnez une connexion Hadoop pour exécuter les profils.

Sélectionner un domaine de données

Choisissez l'une des options de domaine de données suivantes :

Tous les domaines de données

Découvre tous les domaines de données dans la ressource.

Groupes de domaines de données spécifiques

Découvre les domaines de données dans les groupes de domaines de données sélectionnés.

Dans le champ Groupes de domaines de données, choisissez un ou plusieurs groupes de domaines de données.

Domaines de données spécifiques

Découvre les domaines de données sélectionnés.

Dans le champ Domaines de données, choisissez un ou plusieurs domaines de données.

Dans l'espace de travail **Bibliothèque**, vous pouvez afficher tous les domaines de données et groupes de domaines de données disponibles dans Enterprise Data Catalog. Pour créer un domaine de données ou un groupe de domaines de données, accédez à la page **Nouveau > Domaine de données**, ou **Nouveau > Groupe de domaines de données**. Dans l'espace de travail **Bibliothèque**, vous pouvez afficher ou supprimer des domaines de données ou des groupes de domaines de données.

Utiliser la conformité de

Choisissez l'une des valeurs de conformité suivantes pour le domaine de données :

Domaine de données

Utilise les valeurs de conformité prédéfinies que vous avez configurées pour les domaines de données.

Lorsque vous créez un domaine de données, vous pouvez configurer le pourcentage minimal de lignes source et le nombre minimal de lignes source comme critères de conformité pour le domaine de données correspondant. Ces valeurs sont des valeurs de conformité prédéfinies.

Personnalisation

Utilise la valeur de conformité que vous entrez dans le champ **Valeur de conformité personnalisée** pour les domaines de données. La valeur personnalisée remplace les valeurs de conformité prédéfinies.

Critères de correspondance de domaine de données

Choisissez l'un des critères de conformité suivants pour le domaine de données correspondant :

Pourcentage

Nombre de lignes correspondantes divisé par le nombre total de lignes.

Lignes

Nombre total des lignes.

Enterprise Data Catalog utilise les propriétés de conformité des données que vous avez configurées pour les domaines de données. Pour afficher les domaines de données, accédez à **Bibliothèque > Ressources > Domaines de données**. Ouvrez chaque domaine de données pour afficher ses propriétés configurées.

Exclure les valeurs NULL de la découverte de domaines de données

Choisissez cette option pour exclure les valeurs NULL de la source de données lorsque vous exécutez la découverte de domaines de données. Lorsque vous utilisez cette option, l'inférence du domaine de données gagne en précision et en fiabilité. Par exemple, vous disposez d'une table comportant 100 lignes, dont 30 contiennent des valeurs NULL. Le nombre de lignes de conformité est 40. Si vous ne choisissez pas cette option, la découverte de domaines de données s'exécute sur l'ensemble des 100 lignes pour découvrir les domaines de données, ce qui peut entraîner une inférence inexacte. Si vous choisissez cette option, la découverte de domaines de données s'exécute sur 70 lignes seulement et les résultats sont plus précis.

Lorsque vous sélectionnez le pourcentage minimal de lignes avec l'option Exclure les valeurs NULL, le pourcentage de conformité représente le rapport du nombre de lignes correspondantes dans une colonne divisé par le nombre de lignes qui ne contiennent pas de valeurs NULL. Par exemple, soit T le nombre total de lignes d'une colonne, M le nombre de lignes correspondantes, N le nombre de lignes avec des valeurs NULL, alors le pourcentage de conformité est $M/(T-N) \%$.

Paramètres de préparation des données du profil de similarité et de fréquence de valeur

Configurez les propriétés de similarité de colonne pour identifier des colonnes et une fréquence de valeur similaires dans la ressource.

Exécuter le profil de similarité

Choisissez l'une des options suivantes :

- Oui. Le scanner de profilage analyse la source de données et prépare les données pour découvrir des colonnes similaires selon les noms de colonnes, les modèles de colonnes et les valeurs uniques.
- Non

Enregistrer les données source

Choisissez l'une des options suivantes :

- Oui. Le scanner de profilage prépare des données pour découvrir des colonnes similaires selon les noms de colonnes, les modèles de colonnes et les valeurs uniques. Il calcule également les fréquences de valeur. Le scanner conserve ensuite les informations calculées dans PostgreSQL. Les informations calculées sont conservées dans PostgreSQL jusqu'à ce que vous choisissiez de supprimer ou de purger la ressource.
- Non. Le scanner de profilage prépare des données pour découvrir des colonnes similaires selon les noms de colonnes, les modèles de colonnes et les valeurs uniques. Le scanner conserve ensuite les informations calculées dans PostgreSQL. Les informations calculées sont conservées dans PostgreSQL jusqu'à ce que vous choisissiez de supprimer ou de purger la ressource.

Option d'échantillonnage

Choisissez l'une des options d'échantillonnage suivantes pour déterminer le nombre de lignes sur lesquelles Enterprise Data Catalog peut exécuter le profil :

- Réutiliser les paramètres de profil de base. Utilisez l'option d'échantillonnage dans la section **Paramètres de profil de base**.

- Toutes les lignes. Exécute le profil sur toutes les lignes de la source de données.
- Lignes aléatoires automatiques. Exécute le profil sur un échantillon aléatoire de lignes. Enterprise Data Catalog calcule le nombre de lignes aléatoires en fonction du nombre de lignes source.
- N lignes aléatoires. Exécute le profil sur le nombre configuré de lignes aléatoires.
Dans le champ **Lignes d'échantillonnage aléatoire**, entrez le nombre de lignes sur lesquelles vous souhaitez exécuter le profil.
- N premières lignes. Exécute le profil sur les N premières lignes de la ressource.
Dans le champ **Nombre de premières lignes d'échantillonnage N**, entrez le nombre de lignes sur lesquelles exécuter le profil.

Paramètres de connexion au domaine

Choisissez l'une des options de paramètres de connexion de domaine suivantes :

- Utiliser les paramètres de configuration du profil. Enterprise Data Catalog utilise le service d'intégration de données spécifié dans la section paramètres de connexion de domaine pour identifier des colonnes similaires dans les sources de données.
- Spécifier les paramètres de connexion au domaine. Pour utiliser un service d'intégration de données différent pour identifier des colonnes similaires dans les sources de données, entrez les paramètres de connexion de domaine pour le service d'intégration de données.

Autorisations et privilèges

Vous pouvez afficher la section Fréquence de valeur dans Enterprise Data Catalog si vous disposez des autorisations et privilèges suivants :

- Dans l'outil Administrator tool, l'administrateur doit attribuer le privilège **Privilèges de données : Afficher les données** à l'utilisateur.
- Dans Catalog Administrator, accédez à la page **Gérer > Sécurité > Ressources**, et attribuez l'autorisation **Lecture des métadonnées et des données** ou **Toutes les autorisations** à la ressource.
- Dans Catalog Administrator, attribuez l'autorisation d'**accès en lecture** à la ressource DataDomain sur la page **Gérer > Sécurité > Ressources**.

Si la ressource de données contient des données sensibles, vous pouvez les afficher dans la ressource de données après que l'administrateur a attribué le privilège **Privilèges de données : Afficher les données sensibles**. Pour plus d'informations sur les privilèges et les autorisations, consultez le guide *Informatica Administrator Reference for Catalog*.

Paramètres d'inférence de clé unique

Une clé unique est une colonne ou une combinaison de colonnes qui identifie de façon unique une ligne dans une source de données. Le service de profilage identifie les colonnes dans l'objet de données afin de générer des clés uniques. Enterprise Data Catalog affiche des inférences de clé unique pour les actifs tabulaires.

La clé unique ne peut pas contenir de doublons. Si une colonne comprend des doublons, elle n'est pas identifiée en tant que clé unique. L'inférence de clé unique est prise en charge dans l'environnement d'exécution natif.

Dans la section **Paramètres d'inférence de clé unique**, vous pouvez configurer les options suivantes pour une ressource afin qu'elle génère la clé unique :

Exécuter l'inférence de clé unique

Le scanner de profil analyse et infère les clés uniques de la source de données.

Seuil % Null dans l'inférence de clé unique

Définit le seuil des valeurs Null dans l'inférence de clé unique. Vous pouvez entrer une valeur comprise entre 0 et 1.

Ignorer l'inférence de clé unique lorsqu'une clé unique acceptée ou documentée existe

Ignore la table avec les clés uniques documentées ou acceptées.

Options d'échantillonnage de la clé unique

Vous pouvez choisir les options d'échantillonnage suivantes :

- Toutes les lignes. Exécute l'inférence de clé unique sur toutes les lignes de l'objet de données.
- <number> premières lignes. Exécute l'inférence de clé unique sur les lignes sélectionnées dans l'objet de données.

Les ressources suivantes prennent en charge l'inférence de clé unique :

Type de ressource relationnelle

- Hive
- Oracle
- Microsoft SQL Server
- Teradata
- IBM Netezza
- Amazon Redshift

Ressource du système de fichiers

- Amazon S3
- HDFS
- ADLS Gen1

Types de fichiers

Fichiers CSV

Lorsque vous configurez une ressource non prise en charge pour inférer les clés uniques, le message d'erreur suivant s'affiche :

```
L'inférence de clé unique n'est pas prise en charge pour le type de ressource : Type de
ressource
```

Similarité de colonne

En tant qu'analyste de données ou architecte de données, vous pouvez numériser vos données d'entreprise pour trouver des colonnes similaires. Lorsque vous exécutez le scanneur de propagation de domaine de données ou lorsque le scanneur est exécuté en fonction d'une planification, il utilise les colonnes similaires déduites pour propager les domaines de données intelligents à d'autres colonnes similaires. Ce processus vous permet de rechercher et de découvrir plus rapidement des ressources intéressantes dans le catalogue.

Dans Catalog Administrator, lorsque vous exécutez une ressource après avoir choisi l'option **Exécuter le profil de similarité**, puis que vous créez et exécutez la ressource **SimilarityDiscovery**, le scanner découvre des colonnes similaires selon les facteurs suivants : nom de colonne, modèle de données de colonnes et valeurs uniques.

Par défaut, Enterprise Data Catalog utilise tous les facteurs pour propager les domaines de données intelligents à d'autres colonnes similaires. Vous pouvez choisir un ou plusieurs facteurs pour propager les domaines de données intelligents à d'autres colonnes similaires. Par exemple, identifiez une colonne ID d'employé comme domaine de données potentiel. Créez un domaine de données intelligent pour la colonne. Pour propager cela à d'autres colonnes similaires, vous ne pouvez choisir que les facteurs de correspondance de noms et de correspondance de modèles.

Vous pouvez identifier la fréquence des valeurs après avoir activé la similarité de colonne pour une ressource et choisi l'option Enregistrer les données source. En fonction de vos besoins professionnels, vous pouvez utiliser la fréquence de la valeur pour analyser les données d'une ressource. Vous pouvez calculer la fréquence de la valeur dans la colonne vue, la colonne de table, le champ CSV, le champ fichier XML et le champ fichier JSON pour les sources relationnelles, les sources semi-structurées et les sources sur lesquelles vous pouvez exécuter le profil de colonne.

Mode de fonctionnement de la similarité de colonne

Dans une organisation, il est possible qu'un nom de colonne tel qu'ID de client existe dans plusieurs sources de données. Pour identifier les sources de données qui contiennent de telles colonnes, vous pouvez utiliser la similarité de colonne dans Enterprise Data Catalog. Elle utilise un clustering non supervisé qui est une technique d'apprentissage automatique pour identifier les colonnes similaires. Enterprise Data Catalog effectue le clustering non supervisé dans plusieurs sources de données en fonction de plusieurs facteurs, tels que la correspondance de valeurs distinctes, la correspondance de modèles et la correspondance de noms. Il attribue ensuite un score de similitude global ainsi que la probabilité de correspondance pour chaque facteur.

Pour déterminer la similarité de colonne, le clustering non supervisé utilise les facteurs suivants :

Correspondance de valeurs distinctes

Détermine les colonnes similaires dans les sources de données en fonction des valeurs distinctes. Le chevauchement des valeurs distinctes est calculé en pourcentage des valeurs distinctes qui se chevauchent dans deux colonnes.

Le catalogue affiche cette mesure en tant que **Valeurs distinctes** dans la section **Colonnes similaires** > **% de confiance**.

Correspondance de modèles

Détermine le modèle de données de colonnes dans les sources de données et calcule le pourcentage de chevauchement des modèles dans les paires de colonnes pour tous les types de données, sauf les types de données numériques. La correspondance de modèles utilise le profilage pour identifier les modèles de données dominants pour chaque colonne et champ. La correspondance de modèles recherche ensuite un chevauchement de ces modèles dans les paires de colonnes. Enterprise Data Catalog n'infère pas les modèles X, XX et XXX pour tous les types de données.

Le catalogue affiche le pourcentage en tant que **Modèle** dans la section **Colonnes similaires** > **% de confiance**.

Par exemple, la correspondance de modèles identifie les paires de colonnes susceptibles de contenir le modèle de carte de crédit.

Correspondance de noms

Détermine les colonnes similaires dans les sources de données en fonction des noms de colonnes. La correspondance de noms utilise la correspondance approximative de chaînes pour identifier les colonnes qui portent des noms similaires.

Le catalogue affiche le pourcentage en tant que **Nom** dans la section **Colonnes similaires** > **% de confiance**.

Par exemple, la correspondance de noms peut identifier toutes les colonnes qui portent le nom de colonne, customerId.

Processus de similarité de colonne

Le processus de similarité de colonne inclut la préparation des données, la transition des données, la découverte et l'ingestion des données dans le catalogue.

Le processus de similarité de colonne inclut les tâches suivantes :

Préparation des données

Après avoir configuré les paramètres dans la section **Paramètres de préparation des données du profil de similarité et de fréquence de la valeur** et exécuté la ressource, le scanner de profilage prépare les données dans la ressource pour inférer des colonnes similaires, puis conserve les informations dans PostgreSQL. De la même manière, vous pouvez préparer et conserver des données pour plusieurs ressources.

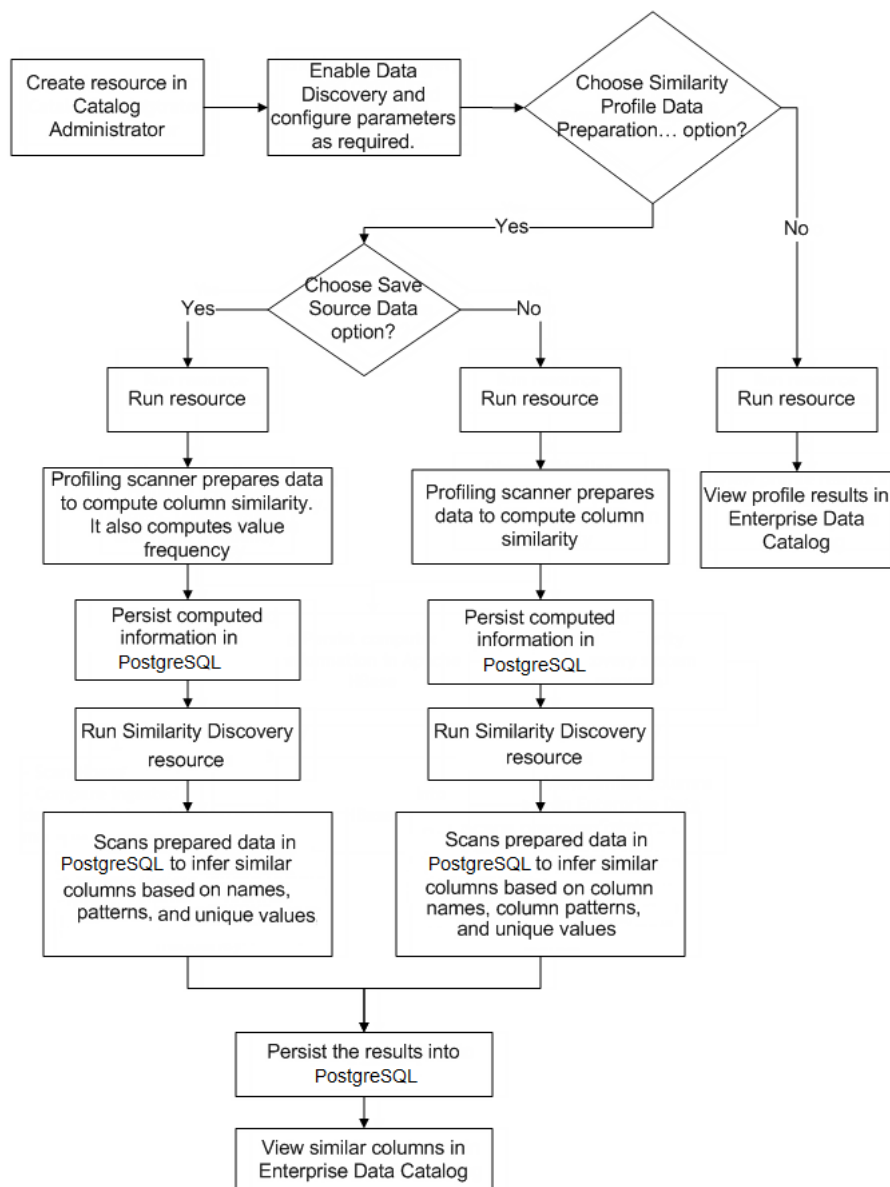
Découverte de colonnes similaires

Pour identifier les colonnes similaires, créez et exécutez la ressource Informatica Similarity Discovery. La ressource Informatica Similarity Discovery s'exécute sur plusieurs ressources pour découvrir des colonnes similaires dans le catalogue. Lorsque vous exécutez la ressource Informatica Similarity Discovery, elle analyse PostgreSQL, compare les données préparées pour inférer des colonnes similaires et conserve les résultats dans PostgreSQL.

Les administrateurs Informatica peuvent soit installer le serveur de base de données PostgreSQL fourni avec le programme d'installation d'Enterprise Data Catalog, soit configurer une base de données PostgreSQL externe après l'installation. Les administrateurs Informatica peuvent configurer une base de données PostgreSQL externe si vous exécutez la ressource Informatica Similarity Discovery. Pour plus d'informations sur la manière dont les administrateurs Informatica peuvent configurer une base de données PostgreSQL externe, reportez-vous à l'article

https://knowledge.informatica.com/s/article/You-can-t-create-logical-partitions-or-configure-high-availability-and-failover-options-when-you-use-the-PostgreSQL-database-bundled-with-the-Enterprise-Data-Catalog-installer?language=en_US de la base de connaissances (KB).

L'image suivante illustre le processus de similarité de colonne :



Dans Catalog Administrator, vous pouvez activer la découverte de données et configurer les propriétés de la ressource afin de découvrir la similarité de colonne pour plusieurs ressources. Après avoir exécuté les ressources et la ressource Informatica Similarity Discovery dans Catalog Administrator, vous pouvez afficher des colonnes similaires selon les noms de colonnes, les modèles de données de colonnes et les valeurs uniques pour les actifs de données dans Enterprise Data Catalog.

Choix des facteurs de propagation des domaines de données intelligents

Vous pouvez choisir un ou plusieurs facteurs pour propager les domaines de données intelligents à d'autres colonnes similaires.

1. Dans Catalog Administrator, créez ou ouvrez une ressource.
2. Exécutez la ressource après avoir choisi l'option **Exécuter le profil de similarité**.

3. Cliquez sur **Bibliothèque > Ressource**.
4. Créez et exécutez la ressource **Informatica Similarity Discovery**.
Le catalogue affiche les actifs liés à la ressource.
5. Créez et enregistrez un fichier texte sur votre machine locale. Nommez le fichier si nécessaire.
6. Modifiez le fichier. Entrez un domaine de données intelligent et les abréviations de facteurs choisis au format suivant :
`<data domain name>=<factor1>,<factor2>`
Les abréviations des facteurs sont N pour la correspondance de noms, P pour la correspondance de modèles et U pour la correspondance de valeurs distinctes.
7. Continuez d'entrer d'autres domaines de données et leurs facteurs choisis dans le fichier. Assurez-vous que chaque domaine de données et ses facteurs se trouvent sur une nouvelle ligne. Les noms de domaines de données sont insensibles à la casse. N'entrez aucun espace dans le fichier.
L'extrait suivant montre un exemple de fichier texte :

```
SDD_Oracle_DD1=P
SDD_PatternCheck=P
DD_NameCheck=N,VF
DD_Check_Pattern=P,U
```
8. Enregistrez le fichier.
9. Dans Catalog Administrator, modifiez la ressource système **DataDomainPropagation**.
10. Dans l'onglet **Paramètres de chargement de métadonnées**, choisissez le fichier texte pour l'option **Types de similarité activés**.
11. Vous pouvez configurer les champs facultatifs suivants selon vos besoins :
 - Choisissez d'activer ou de désactiver le seuil du score de confiance.
 - Choisissez Oui pour activer le seuil de score de confiance et accepter les domaines de données intelligents si la correspondance du domaine de données dépasse la valeur de seuil d'acceptation automatique configurée. Spécifiez le seuil d'inférence et le seuil d'acceptation automatique pour la ressource système **DataDomainPropagation**.
 - Choisissez Non si vous souhaitez organiser séparément les domaines de données intelligents inférés.
 - **Seuil d'inférence**. Entrez un pourcentage minimal de conformité requis pour prendre en compte les domaines de données intelligents pour l'inférence.
 - **Seuil d'acceptation automatique**. Entrez un pourcentage pour accepter les domaines de données intelligents si la correspondance de domaine de données dépasse le pourcentage configuré.
12. Enregistrez et exécutez la ressource système.
Le scanner **DataDomainPropagation** propage les domaines de données intelligents dans le fichier texte selon les facteurs choisis. Pour propager le reste des domaines de données, le scanner utilise les facteurs suivants : correspondance de noms, correspondance de modèles et correspondance de valeurs distinctes.

Exemple d'entreprise

Alex est analyste de données dans une institution financière qui a des succursales et des franchises à travers l'Amérique du Nord. L'institution a récemment acquis une autre institution financière de taille égale. Certains clients ont des comptes dans les deux institutions financières. Alex et son équipe sont invités à intégrer toutes les informations relatives aux clients dans une base de données unique. Il souhaite également rechercher les clients en fonction des régions telles que le Nord-Est, le Sud, le Midwest et l'Ouest.

Voici les défis auxquels Alex et son équipe font face :

- Parcourir plusieurs sources pour identifier des données de clients similaires.
- Identifier le lignage et l'analyse d'impact des données avant de supprimer des données dupliquées.
- Identifier les éléments de données qui peuvent être joints.
- Baliser les colonnes similaires avec des attributs supplémentaires afin qu'Alex et son équipe puissent rechercher des données nécessaires plus rapidement.

Le tableau suivant répertorie les scénarios qu'Alex et son équipe doivent gérer et la manière dont Alex utilise la ressource Informatica Similarity Discovery et les domaines de données pour extraire les informations requises :

Scénario	Résolution d'écran
Différents systèmes de base de données sont utilisés par l'institution financière et l'institution acquise.	<p>Identifiez les sources de données qui doivent être analysées pour trouver les clients requis correspondant aux critères d'admissibilité. Ajoutez ces sources de données en tant que ressources dans Catalog Administrator afin d'extraire les métadonnées de ces ressources.</p> <p>Alex identifie les bases de données de l'entreprise qui incluent les informations relatives aux clients.</p>
Il existe un manque de cohérence et de contexte dans les noms de colonnes, ce qui rend difficile la recherche et l'analyse de colonnes source avec des données similaires.	<p>Dans l'administrateur de catalogue, activez les options de découverte de données et d'exécution du profil de similitude pour les ressources sélectionnées. Le scanner de profilage et la ressource Informatica Similarity Discovery identifient les colonnes similaires dans les ressources. Les colonnes similaires sont identifiées en fonction des noms de colonnes, des modèles de données de colonnes et des valeurs uniques. Enterprise Data Catalog affiche les colonnes similaires pour les actifs de données.</p> <p>Alex utilise des colonnes similaires pour identifier les colonnes qui contiennent des données similaires dans toutes les sources de données, pour identifier les données qui peuvent être jointes et pour dupliquer les données qui peuvent être supprimées.</p> <p>Par exemple, Alex découvre d'après un rapport bancaire existant que les deux organisations stockent le numéro de sécurité sociale dans tous les dossiers qui contiennent des informations relatives aux clients. Sur la base de ces informations, il déduit que si les tables contiennent une colonne avec des informations SSN, les informations relatives aux clients peuvent être présentes dans les tables. Pour vérifier, Alex recherche une colonne SSN dans le catalogue, Enterprise Data Catalog répertorie la colonne recherchée ainsi que d'autres colonnes à partir de toutes les sources de données qui sont similaires à la colonne recherchée. D'après les résultats de la recherche, Alex et son équipe identifient les données qui peuvent être jointes et les données dupliquées qui peuvent être supprimées.</p>

Scénario	Résolution d'écran
Identifier le lignage de chaque ressource de données, les autres ressources qui sont liées à une ressource particulière et l'impact que l'adhésion ou la suppression d'une ressource de données spécifique peut causer sur les autres ressources de données connexes.	<p>Dans Enterprise Data Catalog, affichez le lignage, le résumé de l'impact et la vue des relations pour les ressources identifiées.</p> <p>Alex et son équipe peuvent afficher le lignage, le résumé de l'impact et les détails de la ressource afin d'identifier l'impact avant la mise à jour ou la suppression d'une ressource spécifique.</p>
Classer les clients en fonction des régions et effectuer des recherches plus rapidement.	<p>Dans Catalog Administrator, créez des domaines de données et des groupes de domaines de données pour rechercher une liste de colonnes avec des informations relatives aux clients spécifiques à une région.</p> <p>Pour définir des domaines de données et des groupes de domaines de données afin de classer les clients en fonction des régions, Alex effectue les étapes suivantes :</p> <ol style="list-style-type: none"> 1. Il crée un domaine de données appelé <code>customer_details</code>. 2. Il attribue le domaine de données à l'une des colonnes qui contient le SSN dans Enterprise Data Catalog. 3. Il crée le domaine de données <code>ZIP_code_<area></code>. <p>Dans Informatica Analyst ou Informatica Developer, Alex spécifie une règle pour remplacer le paramètre <code><area></code> dans le domaine de données avec les succursales des institutions financières. Enterprise Data Catalog utilise la règle pour faire correspondre un modèle de colonne avec le code postal pour une succursale spécifique.</p> <p>Remarque: Une règle est une logique d'entreprise qui définit les conditions appliquées aux données lorsque vous exécutez un profil. Vous pouvez ajouter une règle au profil afin de nettoyer, modifier ou valider les données du profil.</p> <ol style="list-style-type: none"> 4. Il crée quatre groupes de domaines de données basés sur les régions appelées <code>Northeast</code>, <code>South</code>, <code>Midwest</code>, et <code>West</code>, et inclut les domaines de données dans leur groupe de domaines de données respectif. <p>Par exemple, le domaine de données qui correspond au code postal <code>ZIP_code_LosAngeles</code> est inclus dans le groupe de domaines de données <code>West</code>.</p> <ol style="list-style-type: none"> 5. Alex effectue une recherche dans Enterprise Data Catalog pour <code>customer_details</code>. Enterprise Data Catalog répertorie toutes les colonnes qui incluent les détails du SSN des clients et affiche également les domaines de données (<code>ZIP_code_<area></code>) et les groupes de domaines de données associés à la colonne. <p>Alex peut également effectuer une recherche basée sur les groupes de domaines de données définis pour trouver une liste de colonnes contenant des informations relatives aux clients spécifiques à une région.</p>

Propagation des termes d'entreprise

Enterprise Data Catalog identifie les termes d'entreprise potentiels qui peuvent être associés à des colonnes et des champs du catalogue. Enterprise Data Catalog exécute régulièrement la ressource système **DataDomainPropagation** pour accepter automatiquement ou recommander les termes d'entreprise pour des

colonnes similaires. Ce processus limite l'intervention manuelle afin d'identifier et d'associer des termes d'entreprise à des tables, colonnes, vues ou champs.

Dans Enterprise Data Catalog, un terme d'entreprise ayant le score de confiance supérieur à 80 % est automatiquement attribué comme fonction à une ressource. Le score de confiance est calculé par un algorithme interne basé sur les termes d'entreprise acceptés sur les domaines de données, la similarité de colonnes et la correspondance de noms entre une colonne et un terme d'entreprise. S'il existe plusieurs termes d'entreprise avec un seuil supérieur à 80 pour cent, le terme dont le niveau de confiance est le plus élevé est automatiquement attribué comme titre d'entreprise. Enterprise Data Catalog affiche au maximum trois termes d'entreprise recommandés si le niveau de confiance est compris entre 60 et 80 pour cent pour une colonne.

Le processus de propagation de termes d'entreprise implique les étapes suivantes :

1. **Vous configurez et exécutez une ressource**

Vous créez une ressource et configurez les paramètres dans l'onglet **Paramètres de chargement des métadonnées > Paramètres de préparation des données du profil de similarité et de fréquence de la valeur**, puis exécutez la ressource. Parallèlement, avec l'option **Exécuter le profil de similarité**, configurez les paramètres dans la section **Paramètres de chargement des métadonnées > Association des termes d'entreprise**, puis exécutez la ressource.

Remarque: Vous pouvez choisir de configurer uniquement **Association des termes d'entreprise** sans configurer l'option **Exécuter le profil de similarité** d'une ressource afin de propager les termes d'entreprise.

Pour plus d'informations sur la configuration de l'**Association des termes d'entreprise**, consultez la rubrique ["Activer l'association des termes d'entreprise " à la page 81](#).

2. **Vous exécutez une ressource Business Glossary ou Axon**

Vous créez et exécutez une ressource Business Glossary ou Axon pour récupérer et afficher les termes d'entreprise du catalogue.

3. **Vous pouvez créer et exécuter la ressource Informatica Similarity Discovery**

Vous pouvez créer et exécuter la ressource **Informatica Similarity Discovery**. La ressource système **SimilarityDiscovery** découvre les colonnes similaires selon le nom de colonne, le modèle de données de colonnes et les valeurs uniques dans la ressource.

Remarque: Il n'est pas nécessaire de configurer l'option **Exécuter le profil de similarité** si vous choisissez de propager les termes d'entreprise uniquement avec l'option **Association des termes d'entreprise**.

4. **Enterprise Data Catalog exécute la ressource système DataDomainPropagation**

Enterprise Data Catalog exécute la ressource système **DataDomainPropagation** toutes les 120 minutes, conformément à une planification prédéfinie. La ressource système **DataDomainPropagation** effectue les tâches suivantes :

- Le scanneur utilise les colonnes similaires inférées pour propager les domaines de données intelligents à d'autres colonnes similaires.
- Acceptation automatique ou recommandation des termes d'entreprise pour d'autres colonnes similaires en fonction du score de confiance.

Le scanner accepte automatiquement les termes d'entreprise ou les affiche sous forme de recommandations pour les colonnes lorsque l'une des conditions suivantes est remplie :

- Un domaine de données contient un terme d'entreprise accepté
- Une colonne contient un terme d'entreprise accepté
- Les noms de la colonne et du terme d'entreprise sont similaires

Business Glossary Synchronisation

Lorsque vous associez un terme d'entreprise à un actif technique dans Enterprise Data Catalog, les informations de relations sont mises à jour dans le glossaire d'entreprise de l'outil Analyst tool. Lorsque vous mettez à jour les liens d'actifs d'un terme d'entreprise dans un glossaire d'entreprise de l'outil Analyst tool, Enterprise Data Catalog peut extraire les informations de liens mises à jour si le glossaire est attribué à la ressource BG_DEFAULT_RESOURCE.

La synchronisation entre Enterprise Data Catalog et le glossaire d'entreprise de l'outil Analyst tool se produit comme suit :

Du catalogue vers le glossaire d'entreprise de l'outil Analyst tool.

Lorsque vous associez un terme d'entreprise à une ressource technique dans le catalogue, la relation est affichée dans le catalogue et dans le glossaire d'entreprise de l'outil Analyst tool. Lorsque vous ouvrez le terme dans le glossaire d'entreprise de l'outil Analyst tool, l'outil Analyst tool importe le lien de ressource et l'affiche dans la section Ressources associées.

Par exemple, vous associez le terme d'entreprise « Salaire annuel » à la colonne SALARY d'une table Oracle. Enterprise Data Catalog affiche le terme Salaire annuel en regard du nom de la colonne dans les résultats de la recherche et dans la vue Détails de l'actif. Lorsque vous ouvrez le terme d'entreprise dans le glossaire d'entreprise de l'outil Analyst tool, la colonne SALARY est répertoriée comme une ressource de données connexe.

Du glossaire d'entreprise de l'outil Analyst tool vers le catalogue.

Lorsque vous mettez à jour un lien d'actif d'un terme dans le glossaire d'entreprise de l'outil Analyst tool, Enterprise Data Catalog peut extraire les informations du lien mises à jour. Enterprise Data Catalog peut extraire les informations de liens si le glossaire est attribué à la ressource BG_DEFAULT_RESOURCE.

Lorsqu'un administrateur d'Informatica crée et active un service Analyst avec les propriétés du service de catalogue, le service Analyst crée une ressource de glossaire d'entreprise par défaut nommée « BG_DEFAULT_RESOURCE ». L'administrateur du catalogue attribue les glossaires d'entreprise à BG_DEFAULT_RESOURCE. La tâche de chargement des métadonnées qui est associée à BG_DEFAULT_RESOURCE extrait les informations du glossaire d'entreprise de l'outil Analyst tool et les importe dans le catalogue. La tâche de chargement des métadonnées démarre lorsqu'un utilisateur de l'outil Analyst tool crée ou met à jour un lien de ressource dans un glossaire d'entreprise qui est attribué à la ressource.

Par exemple, un administrateur du catalogue attribue le glossaire d'entreprise Finance_Glossary à BG_DEFAULT_RESOURCE. Un utilisateur de l'outil Analyst tool lie le terme d'entreprise « Coût prévu » dans le glossaire Finance_Glossary à une ressource de données de fichier plat nommé « COSTS ». Cette action lance la tâche de chargement des métadonnées pour BG_DEFAULT_RESOURCE. Lorsque la tâche de chargement des métadonnées est terminée, vous voyez le terme « Coût planifié » dans le catalogue. La vue Détails de la ressource indique la ressource de données « COSTS » dans la section Autres ressources.

Si vous créez une ressource de glossaire d'entreprise pour extraire les actifs d'un glossaire qui n'est pas attribué à BG_DEFAULT_RESOURCE, Enterprise Data Catalog ne synchronise pas automatiquement les actifs. Pour cela, vous devez charger manuellement la ressource de glossaire d'entreprise par le biais de l'administrateur de catalogue. Vous pouvez également configurer une planification dans l'administrateur de catalogue pour charger la ressource à intervalles réguliers.

Pour plus d'informations sur le glossaire d'entreprise de l'outil Analyst tool, consultez le *Guide d'Informatica Business Glossary*. Pour plus d'informations sur le chargement de ressources via Catalog Administrator, reportez-vous au *Guide d'Informatica Catalog Administrator*.

Aperçu, préparation et provisionnement des données

Un catalogue de données efficace permet aux utilisateurs de la veille stratégique et aux experts en science des données de localiser, de prévisualiser, de préparer et de provisionner des données pour l'analyse ad hoc.

Le provisionnement de données est le processus de déplacement des données d'une source du catalogue vers une cible pour un traitement approfondi. Vous pouvez prévisualiser des données pour les types d'actifs table et fichier, puis créer des tâches de provisionnement dans Enterprise Data Catalog. Prévisualisez un exemple des données source avant de créer une tâche de provisionnement.

La préparation des données est le processus de combinaison, de nettoyage, de transformation et de structuration des données d'un ou plusieurs actifs de données afin que les données soient prêtes pour l'analyse.

Remarque: Vous pouvez préparer des données pour des actifs tels que des tables et des vues dans des ressources relationnelles.

Prérequis

Prérequis pour l'aperçu et le provisionnement de données

Avant d'activer la capacité d'aperçu et de provisionnement de données, vérifiez que les prérequis suivants sont respectés :

1. Veillez à avoir accès au compte Informatica Intelligent Cloud Services en tant qu'administrateur.
2. Créez une organisation Informatica Intelligent Cloud Services.
3. Créez les connexions nécessaires à la base de données dans Informatica Intelligent Cloud Services pour les sources et les cibles.
4. Vérifiez que vous avez créé le mappage `EdcProvisionMapping` à l'aide du modèle d'intégration **Copier les données dans une cible existante** dans Informatica Intelligent Cloud Services.
5. Configurez les propriétés personnalisées **LdmCustomOptions.enableDataProvision** et **LdmCustomOptions.provision.ics.master.app.url** dans le service de catalogue pour le provisionnement de données dans la console d'Informatica Administrator.
6. Ajoutez les organisations Informatica Intelligent Cloud Services à l'outil Catalog Administrator.

Prérequis pour la préparation des données

Avant d'activer la préparation des données, vérifiez les prérequis suivants :

1. Veillez à avoir accès au compte Informatica Intelligent Cloud Services en tant qu'administrateur.
2. Créez une organisation Informatica Intelligent Cloud Services.
3. Créez les connexions nécessaires à la base de données dans Informatica Intelligent Cloud Services pour les sources.
4. Configurez les propriétés personnalisées **LdmCustomOptions.enableDataProvision** et **LdmCustomOptions.dataprep.app.url** dans le catalogue de service pour la préparation des données dans l'outil Informatica Administrator.

Remarque: Vous devez activer la propriété personnalisée **LdmCustomOptions.enableDataProvision** pour afficher et configurer la préparation des données dans l'onglet **Accès aux données** de l'outil Catalog Administrator.

5. Ajoutez les organisations Informatica Intelligent Cloud Services dans l'outil Catalog Administrator.

Aperçu des données

Vous pouvez prévisualiser un exemple de données source dans Enterprise Data Catalog. Vous souhaitez peut-être prévisualiser des données avant d'effectuer une tâche de provisionnement. Prévisualisez les données pour évaluer les données source avant de transférer les données vers la cible. Par exemple, supposez une source Oracle qui contient des informations sur les clients, et vous voulez migrer les données vers une cible Oracle. Avant de déplacer des données de la source vers la cible, vous souhaitez peut-être vérifier les données dans la table et savoir si celle-ci contient des informations pertinentes. Pour vérifier les données, vous pouvez prévisualiser la table et vous assurer qu'elle contient des colonnes pertinentes, telles qu'Adresse, Numéro de téléphone et d'autres informations sur les clients.

Vous pouvez effectuer l'aperçu des données pour les sources suivantes :

- Amazon Redshift
- Amazon S3
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure SQL Server
- Hive
- JDBC
- Microsoft SQL Server
- Oracle
- Salesforce
- Teradata

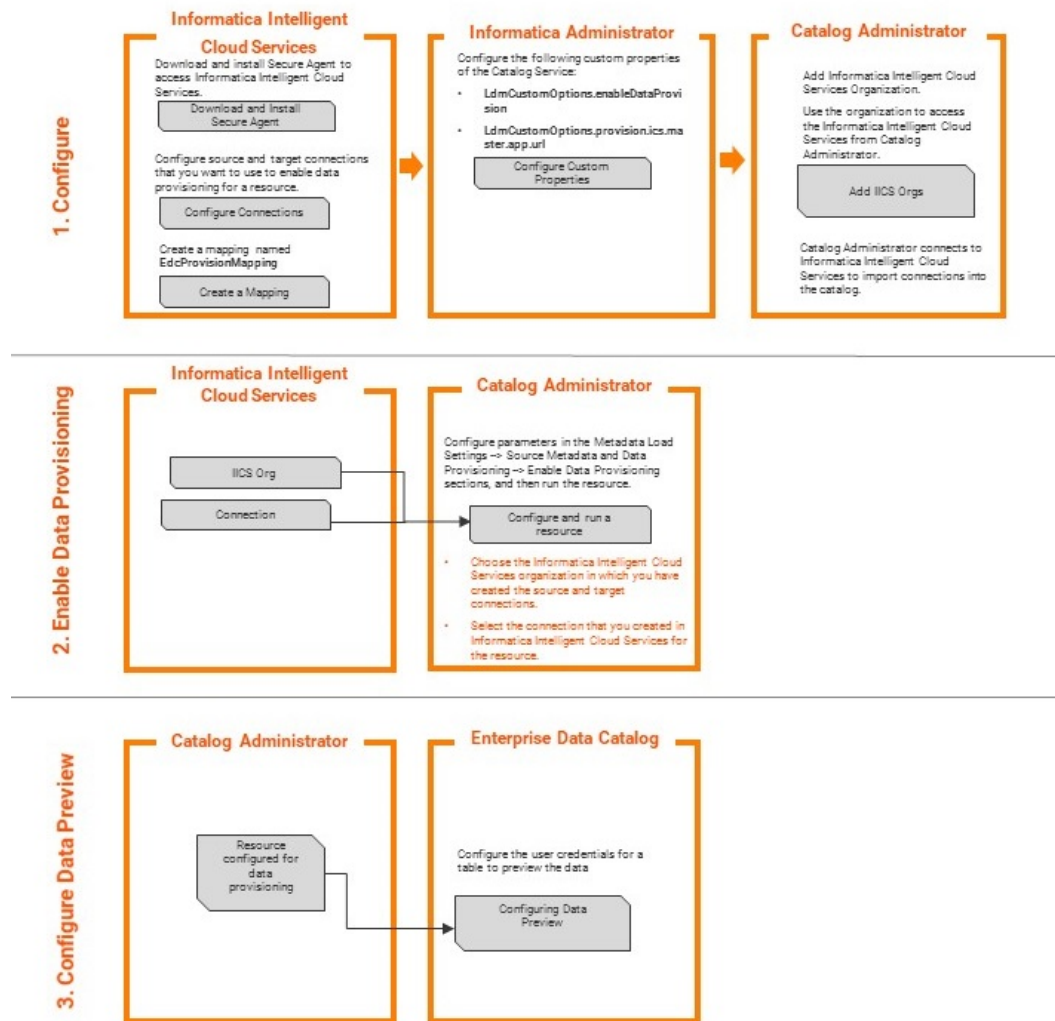
Vous pouvez uniquement prévisualiser les actifs de table et de fichier du catalogue. Avant de prévisualiser les données dans le catalogue, configurez le provisionnement de données dans Catalog Administrator.

Vous pouvez également prévisualiser les données des objets de données à l'aide des API REST. Pour plus d'informations, reportez-vous à la section [Preview Data for a Data Object](#) du *Guide de référence de l'API REST d'Enterprise Data Catalog*.

Processus d'aperçu des données

Le processus d'aperçu des données implique plusieurs étapes de configuration dans Informatica Intelligent Cloud Services, Informatica Administrator et Catalog Administrator. Ces étapes sont identiques aux étapes de configuration du provisionnement de données.

L'image suivante illustre le processus d'aperçu des données :



1. Configurer. Configurez Informatica Intelligent Cloud Services et des propriétés personnalisées pour le provisionnement de données dans Informatica Administrator, puis ajoutez des organisations Informatica Intelligent Cloud Services dans Catalog Administrator.
2. Activez le provisionnement de données. Configurez une ressource pour le provisionnement de données.
3. Configurez l'aperçu des données. Configurez l'aperçu des données pour les types d'actif table et fichier dans Enterprise Data Catalog.

Le processus d'aperçu des données implique les étapes suivantes :

1. **Configurez les propriétés personnalisées du provisionnement de données dans la console d'Informatica Administrator**

Pour plus d'informations sur l'ajout de propriétés personnalisées, reportez-vous à la section Activer l'aperçu, le provisionnement et la préparation des données du *Guide d'Informatica Administrator pour Enterprise Data Catalog*.

2. **Vous ajoutez des organisations Informatica Intelligent Cloud Services à Catalog Administrator**

Vous ajoutez les organisations pour obtenir l'accès à Informatica Intelligent Cloud Services. Catalog Administrator se connecte à Informatica Intelligent Cloud Services afin d'importer des connexions à

la base de données pour la source et la cible. Pour plus d'informations sur l'ajout d'organisations Informatica Intelligent Cloud Services, reportez-vous à la section Ajout d'une organisation Informatica Intelligent Cloud Services.

3. **Activez le provisionnement de données pour une ressource**

Créez une ressource et configurez les paramètres dans la section **Accès aux données > Activer le provisionnement de données**, puis exécutez la ressource. Lorsque vous activez le provisionnement de données pour la ressource, Enterprise Data Catalog récupère les informations de la connexion source des données à prévisualiser.

Remarque: Pour activer le provisionnement de données d'une ressource, vous devez sélectionner le même nom de schéma ou nom de répertoire que vous avez configuré pour une connexion dans Informatica Intelligent Cloud Services.

Pour plus d'informations sur l'activation du provisionnement de données pour une ressource, reportez-vous à la section Activer le provisionnement de données.

4. **Vous configurez les propriétés de connexion de la source de données dans Enterprise Data Catalog**

Vous spécifiez les informations d'identification d'utilisateur de la base de données dans les propriétés de connexion de la source de données, puis prévisualisez les données.

Pour plus d'informations sur la configuration des propriétés de connexion des données source, reportez-vous à la section Aperçu des données du *Guide de l'utilisateur d'Informatica Enterprise Data Catalog*.

Provisionnement de données

Le provisionnement de données est un processus de transfert de données de la source vers la cible pour une analyse et un traitement approfondis. Vous pouvez provisionner des tables et des fichiers vers la cible. Enterprise Data Catalog utilise un mappage et des connexions à la base de données depuis Informatica Intelligent Cloud Services pour effectuer l'approvisionnement de données. Un mappage est un ensemble d'entrées et de sorties qui représentent le flux de données entre les sources et les cibles. Il contient des composants tels que des objets sources, des objets cibles et des transformations. Activez le provisionnement de données dans Catalog Administrator. Après avoir activé le provisionnement de données d'une ressource, vous pouvez créer des tâches de provisionnement dans Enterprise Data Catalog. Vous pouvez créer des tâches de provisionnement pour les types d'actifs des tables et des fichiers.

Vous pouvez effectuer le provisionnement de données pour les sources suivantes :

- Amazon Redshift
- Amazon S3
- Azure Data Lake Store
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure SQL Server
- Hive
- JDBC
- Microsoft SQL Server
- Microsoft Azure Blob Storage
- Oracle
- Salesforce
- Teradata

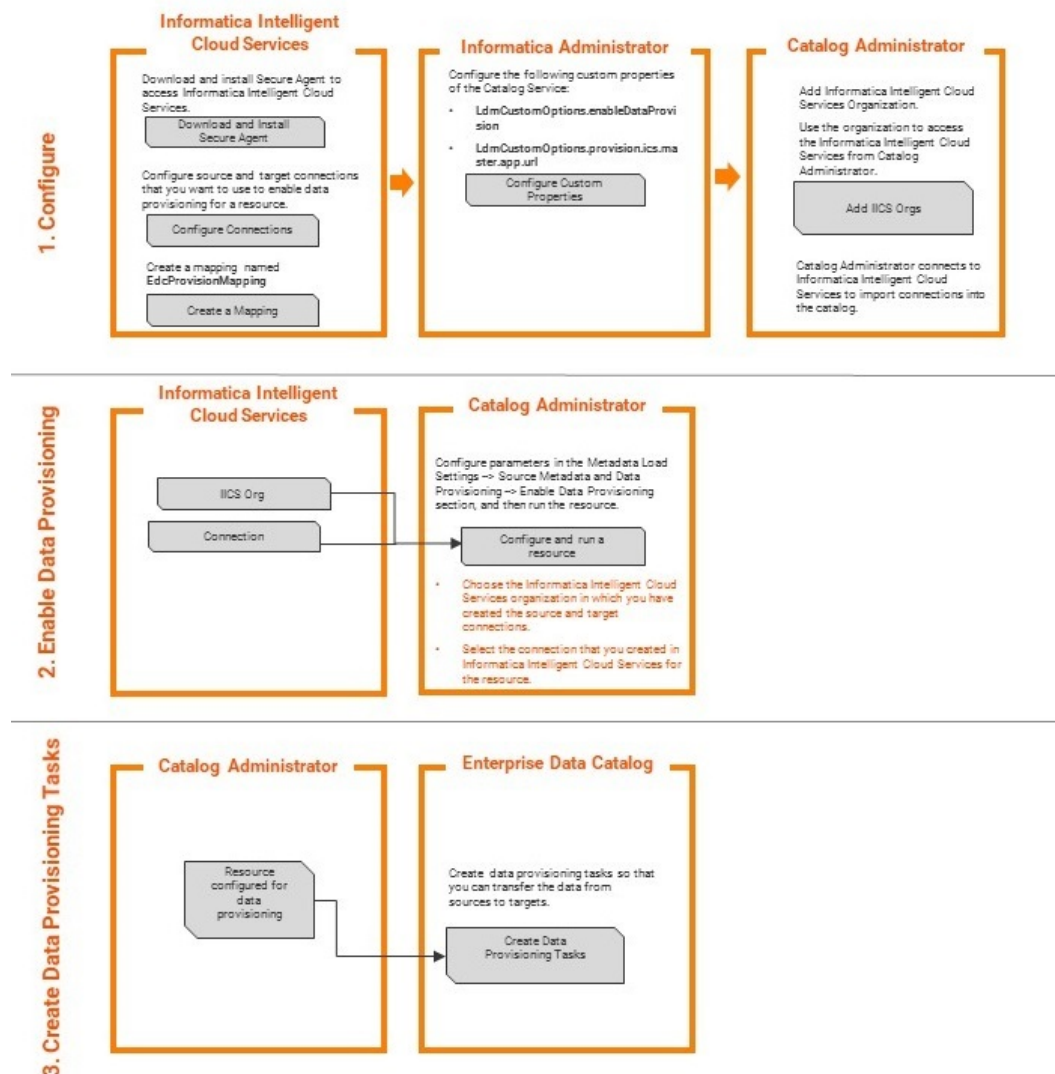
Vous pouvez effectuer le provisionnement de données pour les cibles suivantes :

- Amazon Redshift
- Amazon S3
- Azure Data Lake Store
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure SQL Server
- Google BigQuery
- Google Cloud Storage
- HDFS
- Hive
- JDBC
- Microsoft Azure Blob Storage
- Microsoft SQL Server
- Oracle
- QlikView
- Teradata
- Tableau Online
- Tableau Server

Processus de provisionnement de données

Le processus de provisionnement de données implique plusieurs étapes de configuration dans Informatica Intelligent Cloud Services, Informatica Administrator et Catalog Administrator.

L'image suivante illustre le processus de provisionnement de données :



1. Configurer. Configurez Informatica Intelligent Cloud Services et des propriétés personnalisées pour le provisionnement de données dans Informatica Administrator, puis ajoutez des organisations Informatica Intelligent Cloud Services dans Catalog Administrator.
2. Activez le provisionnement de données. Configurez une ressource pour le provisionnement de données.
3. Créez les tâches de provisionnement de données. Créez les tâches de provisionnement de données pour les actifs de la ressource configurée pour le provisionnement de données dans Enterprise Data Catalog.

Le processus de provisionnement de données implique les étapes suivantes :

1. **Vous configurez les propriétés personnalisées du provisionnement de données dans la console d'Informatica Administrator**

Pour plus d'informations sur l'ajout de propriétés personnalisées, reportez-vous à la section Activer l'aperçu, le provisionnement et la préparation des données du *Guide d'Informatica Administrator pour Enterprise Data Catalog*.

2. Vous ajoutez des organisations Informatica Intelligent Cloud Services à Catalog Administrator

Vous ajoutez les organisations pour obtenir l'accès à Informatica Intelligent Cloud Services. Catalog Administrator se connecte à Informatica Intelligent Cloud Services afin d'importer des connexions à la base de données pour la source et la cible. Pour plus d'informations sur l'ajout d'organisations Informatica Intelligent Cloud Services, reportez-vous à la section Ajout d'une organisation Informatica Intelligent Cloud Services.

3. Configurez et exécutez une ressource dans Catalog Administrator

Créez une ressource et configurez des paramètres dans les sections **Paramètres de chargement des métadonnées** > **Métadonnées source** et **Accès aux données** > **Activer le provisionnement de données**, puis exécutez la ressource.

Remarque: Pour activer le provisionnement de données d'une ressource, vous devez sélectionner le même nom de schéma que vous avez configuré pour une connexion dans Informatica Intelligent Cloud Services.

Pour plus d'informations sur l'activation du provisionnement de données pour une ressource, reportez-vous à la section Activer le provisionnement de données.

4. Vous créez des tâches de provisionnement de données dans Enterprise Data Catalog

Vous configurez les informations de connexions source et cible de la base de données afin de provisionner des données.

Pour plus d'informations sur la création de tâches de provisionnement de données, reportez-vous à la section Création d'une tâche de provisionnement du *Guide de l'utilisateur d'Informatica Enterprise Data Catalog*.

Vous pouvez également effectuer cette tâche à l'aide des API REST. Pour plus d'informations, reportez-vous aux sections API REST de provisionnement de données et Répertoire toutes les tâches de provisionnement du *Guide de référence des API REST d'Enterprise Data Catalog*.

Préparation des données

La préparation des données est le processus de combinaison, de nettoyage, de transformation et de structuration des données d'un ou plusieurs actifs de données afin que les données soient prêtes pour l'analyse. Enterprise Data Catalog utilise des connexions à la base de données à partir d'Informatica Intelligent Cloud Services pour préparer les données. Lorsque vous activez la préparation des données pour une ressource, vous pouvez préparer les données pour les types d'actifs tels que les tables et les vues.

Vous pouvez configurer la préparation des données pour toutes les ressources relationnelles dans Enterprise Data Catalog.

Processus de préparation des données

Le processus de préparation des données implique plusieurs étapes de configuration dans Informatica Intelligent Cloud Services, Informatica Administrator et Catalog Administrator.

Le processus de préparation des données implique les étapes suivantes :

1. Configurez les propriétés personnalisées de la préparation des données dans l'outil Informatica Administrator

Pour plus d'informations sur l'ajout de propriétés personnalisées, reportez-vous à la section Activer l'aperçu, le provisionnement et la préparation des données du *Guide d'Informatica Administrator pour Enterprise Data Catalog*.

2. **Ajoutez des organisations Informatica Intelligent Cloud Services dans l'outil Catalog Administrator.**

Ajoutez les organisations Informatica Intelligent Cloud Services dans l'outil Catalog Administrator. L'outil Catalog Administrator se connecte à Informatica Intelligent Cloud Services afin d'importer des connexions à la base de données pour la source et la cible. Pour plus d'informations sur l'ajout d'organisations Informatica Intelligent Cloud Services, reportez-vous à la section Ajout d'une organisation Informatica Intelligent Cloud Services.

3. **Configurer et exécuter une ressource dans l'outil Catalog Administrator**

Créez une ressource et configurez des paramètres dans les sections **Paramètres de chargement des métadonnées > Métadonnées source** et **Accès aux données > Activer la préparation des données**, puis exécutez la ressource. Pour plus d'informations sur les propriétés de préparation des données, reportez-vous à la section ["Activer la préparation des données" à la page 84](#).

4. **Préparer les données dans Enterprise Data Catalog**

Ouvrez l'actif requis dans l'outil Catalogue et préparez les données.

Pour plus d'informations sur la préparation des données, reportez-vous à la section Préparation des données du *Guide de l'utilisateur d'Informatica Enterprise Data Catalog*.

CHAPITRE 3

Utilisation de Catalog Administrator

Ce chapitre comprend les rubriques suivantes :

- [Présentation de Catalog Administrator, 58](#)
- [Espace de travail Présentation, 60](#)
- [Espace de travail Ressource, 60](#)
- [Espace de travail Surveillance, 61](#)
- [Espace de travail Bibliothèque, 61](#)
- [Espace de travail Domaines de données, 62](#)

Présentation de Catalog Administrator

Catalog Administrator est l'outil d'administration que vous pouvez utiliser pour effectuer des tâches d'administration, telles que la gestion des ressources, des planifications et des attributs.

Utilisez Catalog Administrator pour effectuer les types de tâches suivants :

Gérer les ressources

Créez, configurez, modifiez et supprimez des ressources. Une ressource est un objet qui représente une source de données externe ou un référentiel de métadonnées à partir duquel les analyseurs extraient les métadonnées. Enterprise Data Catalog effectue toutes les opérations de base, telles que l'extraction de métadonnées, le stockage de métadonnées dans le cluster Informatica et la gestion des métadonnées au niveau des ressources.

Gérer les planifications.

Créez des planifications que vous pouvez attacher aux ressources. Vous pouvez créer des planifications globales et récurrentes que vous pouvez attribuer à plusieurs ressources.

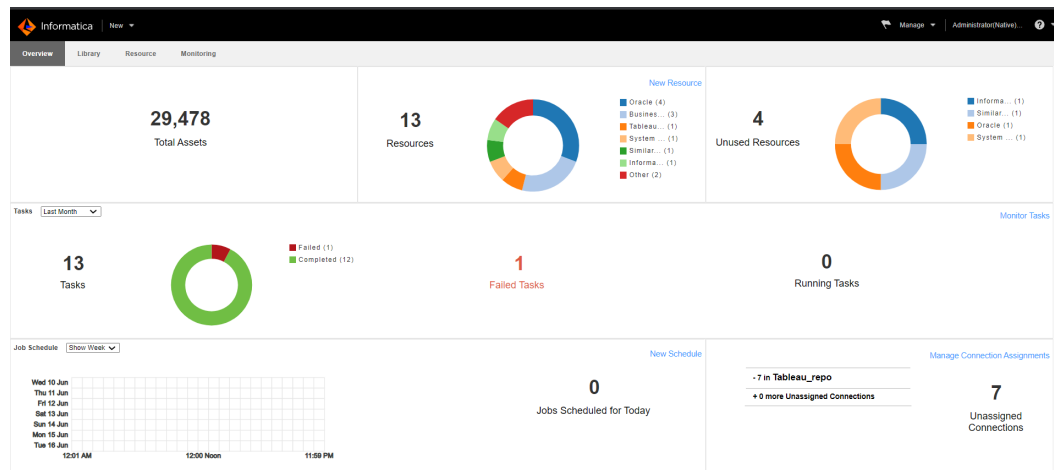
Gérer les attributs

Attribuez des attributs système prédéfinis à des types d'objets de métadonnées spécifiques, tels que Table, Colonne, Rapport et Ressource. Vous pouvez créer des attributs personnalisés et les attribuer aux types d'objets de métadonnées en fonction des besoins professionnels. Les attributs affectés aux ressources permettent aux utilisateurs d'Enterprise Data Catalog de trouver rapidement les ressources de données et les informations connexes. Vous pouvez configurer les attributs système et personnalisés afin qu'Enterprise Data Catalog affiche les attributs en tant que filtres de recherche.

Gérer les connexions PowerCenter, SAP BusinessObjects et Big Data Object

Vous pouvez afficher les informations relatives aux connexions qui sont automatiquement attribuées à chaque ressource. Vous pouvez également afficher des connexions d'objets PowerCenter, SAP BusinessObjects, Cloudera et Hive attribuées et non-attribuées qui sont définies par l'utilisateur et les schémas de chaque connexion. Vous pouvez attribuer des schémas spécifiques aux ressources appropriées. Annulez l'attribution des connexions et des schémas selon les besoins.

L'image suivante présente l'interface de Catalog Administrator :



Catalog Administrator possède les onglets suivants :

Présentation

Consultez les statistiques de surveillance des ressources et des tâches. Vous pouvez afficher la distribution des tâches par statut de tâche et temps d'exécution. Vous pouvez également afficher la distribution des ressources et le chargement de tâche prédictive pour la semaine en cours.

Ressource

Créez des ressources. Vous pouvez également ouvrir les objets récemment configurés.

Surveillance

Afficher les statistiques de surveillance par type de tâche et statut de tâches. Appliquez des filtres pour diminuer le nombre de tâches et de ressources qui répondent à des conditions spécifiques.

Bibliothèque

Afficher la liste des ressources et des planifications. Ouvrez une ressource ou une planification pour une analyse plus poussée.

Catalog Administrator possède les éléments d'en-tête suivants :

Nouveau

Créez des ressources et des planifications.

Gérer

Gérez les attributs système et personnalisés, les attributions de connexion et les paramètres de configuration réutilisables.

Administrator

Modifiez le mot de passe de Catalog Administrator et déconnectez-vous de Catalog Administrator.

Aide

Accédez aux rubriques d'aide spécifiques pour l'espace de travail ou la page en cours, lancer l'aide en ligne et afficher la version d'Informatica.

Espace de travail Présentation

L'espace de travail Présentation affiche des graphiques visuels qui représentent les statistiques de surveillance des tâches, des ressources et de la charge système. Cliquez sur les sections appropriées des graphiques interactifs pour afficher les informations.

Vous pouvez afficher les statistiques de surveillance visuelles suivantes dans l'espace de travail Présentation :

- Nombre total de ressources dans le catalogue.
- Nombre total de ressources et nombre de ressources par type de ressource. Ces nombres peuvent vous aider à surveiller le chargement des ressources dans Catalog Administrator.
- Nombre total de ressources configurées que vous n'avez pas encore utilisées pour extraire des métadonnées. Vous pouvez également afficher le nombre de ces ressources par type de ressource.
- Nombre total de tâches et leurs statuts. Les statuts de la tâche incluent Exécution, Échec, Mis en attente et Suspendu. Vous pouvez vous concentrer sur toute divergence de tâche inattendue en fonction des informations sur le statut de la tâche. Vous pouvez afficher toutes les tâches des dernières 24 heures, de la semaine dernière ou du mois passé.
- Nombre de tâches ayant échoué et tâches en cours d'exécution qui requièrent votre attention. Cliquez sur le lien de la tâche pour afficher les informations dans l'espace de travail Surveillance.
- Nombre de tâches en cours d'exécution en fonction du temps d'exécution de la tâche. Par exemple, vous pouvez afficher les informations sur les tâches en cours d'exécution depuis plus d'un jour ou de 4 à 12 heures.
- Représentation graphique de la charge de travail prédictive en termes de nombre d'emplois à différents créneaux horaires pendant la semaine ou le jour en cours.
- Nombre total de tâches prévues pour la journée.
- Nombre total de connexions non attribuées.

Espace de travail Ressource

Vous pouvez créer des ressources à partir de l'espace de travail Ressource. Vous pouvez également lancer les ressources de données récemment ouvertes.

Pour créer une ressource, cliquez sur le lien sous le panneau **Nouvelles ressources**. Pour ouvrir une ressource de données récemment ouverte, cliquez sur la ressource sous le panneau **Récemment ouverts**.

Espace de travail Surveillance

Vous pouvez surveiller le statut des tâches d'Enterprise Data Catalog dans l'espace de travail Surveillance. L'espace de travail affiche la représentation visuelle des tâches et leur distribution. Vous pouvez afficher le statut des tâches, comme Exécution, Échec et Suspendu.

Vous pouvez afficher les informations suivantes sur les tâches dans le volet inférieur de l'espace de travail Surveillance :

- Type. Indique le type de tâche, comme le chargement des métadonnées, l'exécuteur de profil et l'extraction de résultats de profil.
- Nom de ressource. Affiche le nom de la ressource.
- Planification. Affiche le nom de la planification associée à la ressource.
- Déclenché par. Indique si la tâche a été déclenchée manuellement ou si elle fait partie d'une planification configurée.
- Statut. Affiche le statut de la tâche, tel que Exécution, Échec, Terminé et Suspendu.
- Heure de début de la tâche.
- Heure de fin de la tâche.
- Temps d'exécution de la tâche.
- Planification suivante. Affiche la date et l'heure auxquelles le scanneur exécute la tâche suivante.
- URL du journal. Affiche un lien sur lequel vous pouvez cliquer pour ouvrir le fichier journal des tâches terminées.

Vous pouvez actualiser l'espace de travail Surveillance pour afficher les dernières tâches et les derniers statuts de tâche. Vous pouvez également appliquer des filtres sur la liste des tâches en fonction des conditions, telles que le type de ressource, le temps de création de tâche et l'historique des échecs de tâche.

Espace de travail Bibliothèque

Utilisez l'espace de travail Bibliothèque pour parcourir, afficher, rechercher et appliquer des filtres à une collection de ressources, de domaines de données composites, de domaines de données, de groupes de domaines de données, de ressources de référence et de planifications sur lesquels vous disposez d'un droit d'accès.

Dans l'espace de travail Bibliothèque, spécifiez les critères de recherche pour trouver une ressource ou un groupe de ressources spécifique. Vous pouvez trier les ressources par nom ou par type. Vous pouvez également regrouper les ressources en fonction des besoins spécifiques.

Vous pouvez ouvrir une ressource ou une planification à partir de la liste des ressources. Lorsque vous cliquez sur une ressource, elle s'ouvre dans l'espace de travail Ressources. Si vous cliquez sur une planification, elle s'ouvre dans l'espace de travail Planifications. Vous pouvez filtrer la liste des ressources selon plusieurs conditions, telles que le nom de la ressource, la personne qui l'a créée par et le type de ressource. Vous pouvez filtrer une planification en fonction des conditions, telles que le nom de la planification et l'heure à laquelle vous l'avez créée.

Espace de travail Domaines de données

L'espace de travail Domaines de données s'ouvre lorsque vous affichez ou créez un domaine de données ou un groupe de domaines de données. Vous pouvez également lancer les domaines de données et les groupes de domaines de données récemment ouverts à l'aide de l'espace de travail.

Pour ouvrir l'espace de travail Domaine de données, effectuez l'une des étapes suivantes :

- Cliquez sur **Nouveau > Domaine de données**
- Cliquez sur **Nouveau > Groupe de domaines de données**.
- Dans l'espace de travail **Bibliothèque**, sélectionnez **Domaines de données** ou **Groupes de domaines de données**, puis cliquez sur un domaine de données ou un groupe de domaines de données existant pour ouvrir les détails associés dans l'espace de travail Domaines de données.

Pour ouvrir un domaine de données récemment ouvert, placez le pointeur de la souris sur l'onglet de l'espace de travail **Domaines de données**, puis cliquez sur l'option **Récemment ouverts** qui s'affiche.

CHAPITRE 4

Gestion des ressources

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion des ressources, 63](#)
- [Ressources et scanners, 64](#)
- [Ressources et planifications, 64](#)
- [Ressources et attributs, 64](#)
- [Ressources de référence, 65](#)
- [Affichage des ressources de référence, 68](#)
- [Création d'une ressource, 68](#)
- [Type de ressource, 69](#)
- [Filtre de profil des données et métadonnées source, 71](#)
- [Liaison insensible à la casse , 72](#)
- [Activation de la découverte de données, 73](#)
- [Découverte de domaines de données composites, 80](#)
- [Activer l'association des termes d'entreprise , 81](#)
- [Activer le provisionnement de données , 84](#)
- [Activer la préparation des données, 84](#)
- [Modification d'une ressource, 85](#)
- [Clonage d'une ressource, 85](#)
- [Exécution d'une analyse sur une ressource, 86](#)
- [Ressources système, 86](#)
- [Affichage d'une ressource, 87](#)

Présentation de la gestion des ressources

Une ressource est un objet de référentiel qui représente une source de données externe ou un référentiel de métadonnées à partir duquel les scanners extraient des métadonnées. Les opérations de métadonnées de base, telles que l'extraction, le stockage et la gestion des métadonnées, sont effectuées au niveau des ressources.

Vous pouvez créer, modifier et supprimer une ressource. Une ressource possède un type de ressource qui détermine le type de source de données à partir duquel les scanners extraient les métadonnées. Vous pouvez également choisir les types spécifiques de métadonnées que vous voulez que les scanners puissent

extraire des sources de données. Par exemple, vous pouvez choisir d'extraire les métadonnées sources de base ou les métadonnées de base et les métadonnées de profil. Vous pouvez afficher les classifications attribuées aux ressources. Les classifications vous permettent de rechercher et de filtrer rapidement des ressources spécifiques dans Enterprise Data Catalog.

Une ressource a les caractéristiques suivantes :

- Une ressource possède un type de ressource qui identifie le type de système source.
- Une ressource possède une identité globale et unique.
- Une ressource peut avoir une planification unique ou récurrente. Vous pouvez joindre une planification à une ressource ou joindre une planification réutilisable à plusieurs ressources.

Ressources et scanners

Les scanners attachés aux ressources identifient le type de ressource, y compris le nom du type de ressource, le nom d'affichage, la description et les versions prises en charge. Vous pouvez configurer plusieurs scanners pour différents types de métadonnées. Vous pouvez configurer les propriétés du scanner pour une ressource lorsque vous créez la ressource. Chaque scanner possède un ID unique qui correspond au type de ressource.

Un type de ressource peut avoir plusieurs scanners. Par exemple, un scanner peut extraire les métadonnées d'une source de données Oracle directement à partir de la source. Un autre scanner peut dériver les métadonnées relatives à la source de données Oracle à partir d'autres référentiels de métadonnées, tels qu'un référentiel PowerCenter ou un référentiel modèle.

Ressources et planifications

Les planifications de ressources déterminent la fréquence à laquelle les scanners extraient les métadonnées des sources de métadonnées. Vous pouvez créer des planifications pour des ressources spécifiques ou créer des planifications réutilisables à attribuer à plusieurs ressources.

Vous pouvez créer des planifications qui ont une date de fin ou des planifications qui s'exécutent indéfiniment. En plus d'une date de début, vous pouvez également configurer l'heure de début en fonction du moment où vous voulez que les scanners commencent à extraire des métadonnées.

Ressources et attributs

Les ressources disposent d'attributs qui représentent des propriétés spécifiques des types d'objet, tels que Colonne Alias, Catégorie, Jeu de données et Table, en fonction des types de données des attributs. Vous pouvez utiliser les attributs de ressource dans Enterprise Data Catalog pour rechercher des informations pertinentes parmi un grand nombre de ressources de données et de relations de données.

La ressource peut avoir des attributs personnalisés associés. Les attributs système incluent des attributs prédéfinis qui représentent des propriétés spécifiques des types d'objet, telles que la procédure stockée, la table, le déclencheur, la requête XML, le schéma et la ressource. Par exemple, Commentaire est un attribut système attribué à une table Hive ou à une vue Hive. Vous pouvez créer des attributs personnalisés en

fonction de vos besoins professionnels et les attribuer à différentes ressources. Vous pouvez choisir de créer des attributs personnalisés en fonction des types de scanners utilisés.

Ressources de référence

Une ressource de référence indique que la ressource que vous avez configurée inclut des références d'autres ressources ou sources de données. Les actifs des ressources de référence sont appelés des actifs de référence.

Les ressources configurées dans Enterprise Data Catalog peuvent inclure des informations de lignage et des données de connexion d'autres ressources et sources de données. Les informations sur la source de données peuvent ne pas être répertoriées dans le catalogue, mais vous pouvez afficher la source de données en tant que ressource de référence lorsque vous affichez le lignage pour les actifs de la ressource.

Par exemple, une ressource Informatica PowerCenter que vous souhaitez configurer peut comprendre des mappages qui pointent vers les tables ou les colonnes d'une base de données Oracle. La ressource PowerCenter est appelée la ressource principale. La base de données Oracle est une ressource de référence pour la ressource PowerCenter. Les tables et les colonnes de la base de données Oracle sont des actifs de référence de la ressource de référence.

Une fois les ressources et actifs de référence extraits, vous pouvez afficher le lignage d'un actif dans le catalogue avec les actifs de référence. Les actifs de référence incluent des sources de données de référence et des ensembles de données de référence que vous pouvez afficher dans Enterprise Data Catalog. Vous pouvez configurer des ressources dans Enterprise Data Catalog pour inclure l'extraction des ressources et actifs de référence. Pour plus d'informations sur la configuration des ressources pour l'extraction des métadonnées des ressources et actifs de référence, consultez le *Guide de configuration du scanner d'Enterprise Data Catalog*.

Pour afficher le lignage avec les actifs inclus dans les sources de données de ressource de référence, vous devez configurer les ressources pour les sources de données de ressource de référence dans Enterprise Data Catalog. Vous devez ensuite effectuer des attributions de connexion entre les ressources de référence et les ressources que vous avez configurées. Par exemple, pour afficher les détails de colonne d'une table Oracle dans une base de données Oracle référencée par une ressource PowerCenter, vous devez ajouter la ressource Oracle dans Enterprise Data Catalog et effectuer des attributions de connexion entre la ressource de référence et le schéma dans la ressource Oracle.

Vous pouvez également attribuer des connexions d'une ressource de référence vers d'autres ressources de référence. Par exemple, si une base de données Microsoft SQL Server est référencée par une ressource PowerCenter et une ressource Tableau, vous pouvez attribuer des connexions de la ressource de référence de la base de données vers des ressources PowerCenter et Tableau.

Avantages de l'extraction des actifs de référence

Une fois que vous avez extrait les actifs de référence, vous pouvez les utiliser comme des actifs standard dans le catalogue.

L'extraction d'actifs de référence dans le catalogue fournit les avantages suivants :

- Recherche de ressources et d'actifs de référence dans Enterprise Data Catalog.
- Filtrage des résultats de la recherche en fonction des ressources et des actifs de référence.
- Affichage du lignage d'une ressource principale qui inclut des ressources et des actifs de référence.
- Affichage du lignage, de l'impact et du résumé de contrôle des actifs de référence.

- Filtrage du lignage en fonction des actifs de référence.
- Annotation et enrichissement des actifs de référence à l'aide d'attributs et de termes d'entreprise personnalisés.
- Exportation ou importation en bloc des attributs personnalisés, des associations de termes d'entreprise et d'autres enrichissements des actifs de référence.
- Affichage du lignage personnalisé entre les ressources de référence.

Pour plus d'informations sur la recherche, les filtres, le lignage, les attributs et les termes d'entreprise personnalisés, et l'importation et l'exportation en bloc, consultez le *Guide de l'utilisateur d'Enterprise Data Catalog*.

Attribuer des connexions aux ressources de référence

L'attribution de connexions aux ressources de référence implique d'attribuer des connexions à partir des ressources de référence aux ressources qui représentent les sources de données ou d'autres ressources de référence.

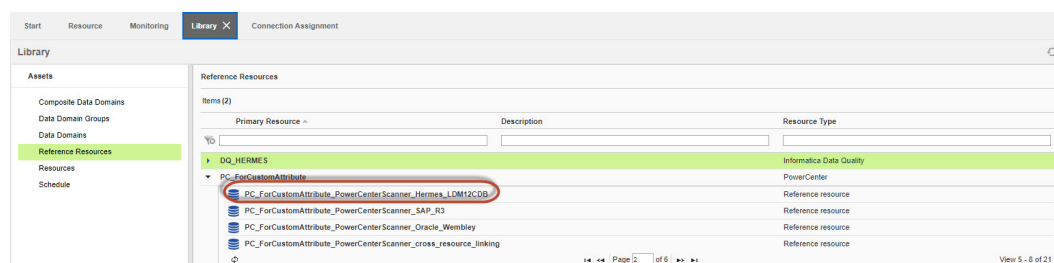
Vous devez attribuer des connexions pour les ressources de référence afin d'afficher les actifs des sources de données référencées dans le lignage. Avant d'attribuer des connexions, vous devez configurer des ressources dans Enterprise Data pour les sources de données de ressources de référence. Pour créer une ressource, reportez-vous à la section *Création d'une ressource*. Vous pouvez ensuite effectuer les étapes répertoriées dans la section *Gestion des connexions* pour attribuer des connexions entre les ressources de référence et les ressources configurées. Lorsque vous attribuez des connexions aux ressources de référence, Enterprise Data Catalog convertit les ressources et les actifs de référence en ressources et actifs standard dans le catalogue.

Une fois que les connexions ont été attribuées aux ressources de référence, elles ne figurent plus dans le catalogue en tant que ressources et actifs de référence. Si vous avez enrichi des actifs de référence avec plusieurs valeurs pour les attributs et les termes d'entreprise personnalisés, ces valeurs sont fusionnées et associées à la ressource après l'attribution de la connexion. Si vous avez enrichi des actifs de référence avec une seule valeur pour les attributs et les termes d'entreprise personnalisés, la valeur la plus récente est associée à la ressource.

Exemple de cas d'utilisation

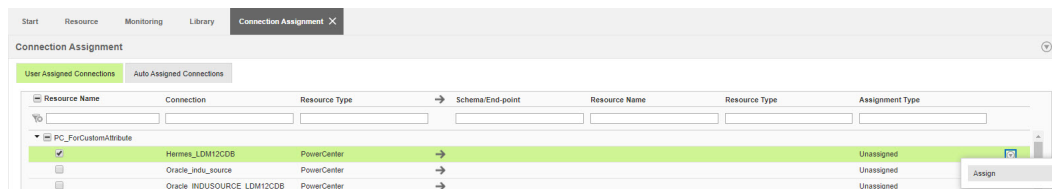
L'exemple de cas d'utilisation décrit le flux de travail impliqué dans l'attribution de connexions entre un exemple de ressource de référence et un schéma dans une ressource dans Enterprise Data Catalog.

Une ressource PowerCenter nommée `PC_ForCustomAttribute` inclut la ressource de référence `PC_ForCustomAttribute_PowerCenterScanner_Hermes_LDM12CDB` sur la page **Ressources de référence** comme indiqué dans l'image suivante :

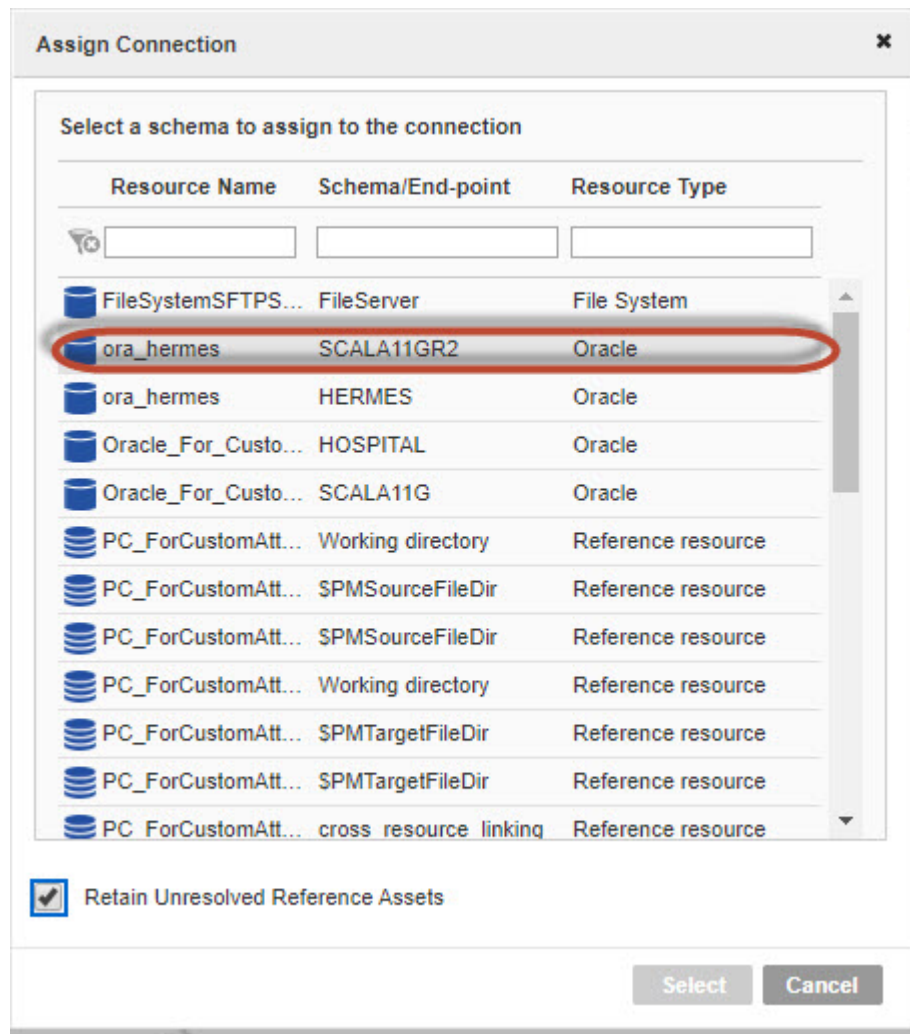


Le préfixe `<resource name>_<Provider ID>_<Connection name>` est ajouté aux ressources de référence.

À partir de l'espace de travail **Attribution de connexion**, vous pouvez attribuer la connexion dans la ressource de référence à une autre ressource comme indiqué dans l'image suivante :



Vous pouvez attribuer le schéma **SCALA11GR2** dans la ressource Oracle **ora_hermes** à la connexion **Hermes_LDM12CDB** comme indiqué dans l'image suivante :



Une fois l'attribution de connexions effectuée, la ressource de référence **Hermes_LDM12CDB** cesse d'exister en tant que ressource de référence et la ressource **ora_hermes** et les actifs inclus dans la ressource sont répertoriés lors de l'affichage du lignage pour les actifs de la ressource **PowerCenter PC_ForCustomAttribute**.

Une fois l'attribution de connexions effectuée, la ressource de référence **PC_ForCustomAttribute_PowerCenterScanner_Hermes_LDM12CDB** ne figure plus dans le catalogue et le lignage des actifs de la ressource **PC_ForCustomAttribute** inclut les actifs de la ressource **ora_hermes**.

Affichage des ressources de référence

Vous pouvez afficher les ressources de référence sur la page **Bibliothèque**.

Pour afficher les ressources de référence dans Catalog Administrator, procédez comme suit :

1. Dans Catalog Administrator, cliquez sur **Bibliothèque**. La page **Bibliothèque** s'affiche.
2. Dans le volet **Bibliothèque** sur la gauche, cliquez sur **Ressources de référence**. La page **Ressources de référence** s'affiche.

Vous pouvez afficher la ressource principale, sa description et son type dans la page **Ressources de référence**. Vous pouvez cliquer sur la liste déroulante à côté de la ressource principale pour développer la liste des ressources de référence.

Création d'une ressource

Lorsque vous créez une ressource, vous pouvez spécifier le type de ressource, le type de métadonnées qu'Enterprise Data Catalog extrait et une planification facultative de la ressource. Vous pouvez également attribuer des attributs personnalisés à la ressource lorsque vous la créez.

Remarque: Vous pouvez valider la connexion lorsque vous configurez une ressource. Cliquez sur **Tester la connexion** pour valider les paramètres de connexion.

1. Cliquez sur **Nouveau > Ressource**.
L'assistant **Nouvelle ressource** s'affiche.
2. Entrez un nom et une description facultative pour la ressource.
3. Si vous souhaitez affecter des valeurs d'attribut personnalisées à la ressource, vous pouvez éventuellement entrer des valeurs pour les attributs tels que le département, le propriétaire des données, le gestionnaire de données et les experts techniques dans la section **Propriétés supplémentaires**.
 - Si vous souhaitez appliquer des valeurs d'attribut personnalisées à toutes les éléments chargés via la ressource, sélectionnez **Propager** sous chaque type d'attribut personnalisé.
4. Si vous souhaitez que les utilisateurs reçoivent des notifications d'échec d'analyse pour la ressource, sélectionnez éventuellement les utilisateurs en tant qu'abonnés aux notifications d'échec dans la section **Propriétés supplémentaires**.
5. Cliquez sur **Sélectionner** pour ouvrir la boîte de dialogue **Sélectionner le type de ressource**.
6. Sélectionnez un type de ressource, puis cliquez sur **Sélectionner**.
D'autres champs apparaissent dans l'assistant en fonction du type de ressource que vous avez sélectionné.
7. En fonction du type de ressource, configurez les propriétés de la connexion.
8. Cliquez sur **Suivant** pour passer à la page **Paramètres de chargement des métadonnées**.
Vous pouvez choisir le type de métadonnées que vous voulez voir extraites par Enterprise Data Catalog des systèmes source.
9. Configurez les métadonnées source requises et les paramètres de métadonnées de profilage.
10. Cliquez sur **Suivant** pour passer à la page **Attributs personnalisés**, et configurez les paramètres d'attribut. Vous pouvez sélectionner les attributs personnalisés à associer à la ressource.
11. Cliquez sur **Suivant** pour passer à la page **Accès aux données**.

12. Pour créer des tâches de provisionnement ou de préparation dans Enterprise Data Catalog, activez éventuellement l'une des deux tâches.
13. Cliquez sur **Suivant** pour accéder à la page **Autorisations** et limitez les autorisations pour la ressource. Vous pouvez également attribuer ou modifier les autorisations de la ressource à partir des paramètres de sécurité.
14. Cliquez sur **Suivant** pour accéder à la page **Planification**.
15. Vous pouvez éventuellement sélectionner les planifications pour le chargement des métadonnées source et les métadonnées de profilage. Vous pouvez créer une planification globale si nécessaire.
Remarque: Vous pouvez afficher le fuseau horaire du serveur pour les planifications personnalisées et réutilisables. Sélectionnez ou créez des horaires en fonction du fuseau horaire du serveur.
16. Cliquez sur **Enregistrer** ou sur **Enregistrer et exécuter**.

Type de ressource

Un type de ressource représente le type de système source à partir duquel les scanners extraient les métadonnées. Un type de ressource identifie les types de métadonnées requis et facultatifs que vous devez configurer pour chaque ressource. Business Objects, Oracle, PowerCenter et Teradata sont des exemples de type de ressource.

Chaque type de ressource a des propriétés différentes que vous devez configurer lorsque vous créez une ressource. Les propriétés incluent les propriétés de la connexion et les propriétés supplémentaires, telles que le propriétaire des ressources et la classification des ressources.

Vous pouvez créer les types de ressources suivants :

- Amazon Redshift
- Amazon S3
- Apache Atlas
- Microsoft Azure SQL Server
- Microsoft Azure SQL Data Warehouse
- Azure Data Lake Store
- Microsoft Azure Blob Storage
- Business Glossary Classification
- Cloudera Navigator
- Custom Lineage
- Scripts de base de données
- Erwin
- Système de fichiers
- Google BigQuery
- Hadoop File System (HDFS)
- Gestionnaire de fichiers à parties multiples Hadoop
- Hive
- IBM Cognos

- IBM DB2
- IBM DB2 for z/OS
- IBM Netezza
- Informatica Axon
- Informatica Cloud
- Informatica Platform
- JDBC
- MDM
- MicroStrategy
- OneDrive
- Oracle
- PowerCenter
- ProfileWarehouse
- QlikView
- Salesforce
- SAP Business Objects
- SAP Business Warehouse (SAP BW)
- SAP BW/4HANA
- Ressources SAP S/4HANA
- SAP R/3
- SharePoint
- SQL Server Integration Services
- Sybase
- Tableau Server
- Teradata
- Workday

Pour exécuter un travail d'analyse sur des types de ressources tels qu'Oracle, DB2 et SQL Server, utilisez les types de ressources standard disponibles dans Enterprise Data Catalog. Informatica vous recommande de ne pas configurer le type de ressource JDBC pour récupérer des données des sources de métadonnées standard prises en charge.

Cependant, vous pouvez utiliser le type de ressource JDBC lorsqu'il n'existe aucun connecteur standard pour récupérer les métadonnées et effectuer la découverte des données à partir des sources suivantes :

- MySQL
- i5/OS
- Amazon Athena
- Vertica
- Postgres
- Tables Databricks Delta

Pour plus d'informations sur la configuration des ressources JDBC, contactez le support client international Informatica.

Pour plus d'informations sur la configuration d'une ressource, consultez le *Guide de configuration du scanneur d'Enterprise Data Catalog*.

Filtre de profil des données et métadonnées source

Lorsqu'une ressource contient des tables et des vues multiples, vous pouvez choisir d'extraire les métadonnées source de tables et vues spécifiques, et d'extraire les données et métadonnées de profil de tables et vues spécifiques.

Vous pouvez configurer le champ de filtre si l'une des conditions suivantes est remplie :

- Vous souhaitez extraire les métadonnées uniquement pour quelques tables et vues d'une liste contenant toutes les tables et vues d'une ressource.
- Vous souhaitez inclure quelques tables et vues de l'exécution de ressource suivante.
- Vous souhaitez exclure quelques tables et vues de l'exécution de ressource.
- Vous souhaitez exclure quelques tables et vues de l'exécution de ressource suivante.

Dans une ressource, vous pouvez utiliser les filtres suivants pour inclure et exclure des tables et des vues :

- Filtre de métadonnées source. Dans l'onglet **Paramètres de chargement des métadonnées** de l'assistant Ressource, indiquez les tables et les vues à inclure ou exclure dans une exécution de ressource. Le scanneur extrait les métadonnées source en fonction des tables et des vues indiquées dans le champ Filtre des métadonnées source.
- Filtre de profil de données. Dans l'onglet **Paramètres de chargement des métadonnées** de l'assistant Ressource, indiquez les tables et les vues à inclure ou exclure dans une exécution de profil. Le scanneur de profil extrait les métadonnées de profil en fonction des tables et des vues indiquées dans le champ Filtre de profil de données.

Pour inclure des tables et des vues, vous pouvez indiquer leurs noms complets ou utiliser le caractère générique % dans le nom ou le modèle. Utilisez des points-virgules (;) pour séparer les noms des tables et les noms des vues. Si le modèle contient un caractère spécial, tel qu'un blanc, placez-le entre des guillemets simples. Par exemple, HR_id ; 'HR Emp%' ; HR_job%.

Remarque: Le filtre de profil de données prend en charge les expressions régulières telles que %, ^, \$, ., | et (a | b) tandis que le filtre de métadonnées source prend en charge l'expression %. La contrainte de filtre s'applique à Microsoft Azure SQL Server, Microsoft Azure SQL Data Warehouse, Hive ainsi qu'à toutes les sources de bases de données relationnelles.

Vous pouvez choisir un ou plusieurs des modèles suivants pour identifier les tables et les vues :

- <name>%. Inclut les tables et les vues qui commencent par <name>.
Par exemple, A% inclut toutes les tables et les vues qui commencent par le caractère A.
- %<name>. Inclut les tables et les vues qui se terminent par <name>.
Par exemple, %id inclut toutes les tables et les vues qui se terminent par le texte 'id'.
- %<name>%. Inclut les tables et les vues qui contiennent <name>.
Par exemple, %DB% inclut toutes les tables et les vues qui contiennent le mot clé « DB ».
- <name>. Inclut les tables et les vues qui correspondent à <name>.
Par exemple, le filtre HR_id inclut n'importe quelle table ou vue de la ressource qui correspond à HR_id.

Pour exclure des tables et des vues de l'analyse, faites précéder le nom ou le modèle du terme NOT. Par exemple, NOT %ID ; NOT 'ID%'. Utilisez des points-virgules (;) pour séparer les noms des tables et les noms des vues.

Dépannage

Que se passe-t-il si j'inclus dans le filtre de métadonnées source une table liée à un objet externe ?

Lorsqu'une table ou une vue filtrée dépend d'un objet externe tel qu'une base de données, le scanneur de ressource extrait les métadonnées de l'objet externe même si celui-ci ne remplit pas les conditions de filtre, car le filtre ne s'applique pas à l'objet externe dépendant.

Comment puis-je identifier les tables et les vues qui ont été exclues de l'analyse de profil ?

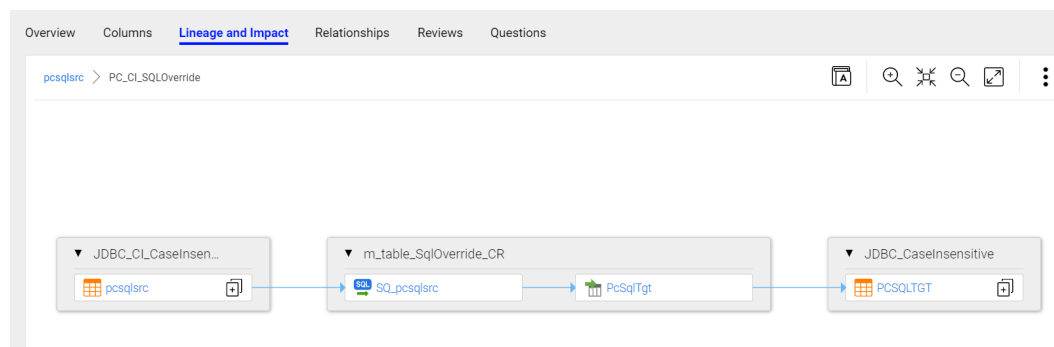
Par défaut, les fichiers journaux ne répertorient pas toutes les tables et les vues qui ont été exclues d'une analyse. Pour afficher la liste des tables et des vues exclues, activez le niveau de journal *Débogage* puis exécutez l'analyse. Les fichiers journaux répertorient toutes les tables et les vues exclues de l'analyse.

Liaison insensible à la casse

Vous pouvez configurer une ressource comme non sensible à la casse. Si vous définissez une ressource comme non sensible à la casse, Enterprise Data Catalog peut créer des liens entre la source et la cible; indépendamment des noms en majuscules ou en minuscules des ressources. Ces noms n'impactent pas les diagrammes de lignage dans le catalogue. La liaison de l'insensibilité à la casse permet d'afficher des informations de lignage des actifs de données pour les ressources qui contiennent une source et une cible sensibles à la casse.

Supposez, par exemple que Catalog Administrator contient des ressources Oracle et de veille stratégique entre lesquelles vous voulez effectuer l'attribution de la connexion. Les points de terminaison source et cible des ressources contiennent des tables et des colonnes sensibles et insensibles à la casse. Dans ce scénario, traditionnellement, le catalogue ne créait pas de liens entre la source et cible. Pour créer des liens entre la source et la cible, vous configurez les ressources Oracle et de veille stratégique comme insensibles à la casse, puis effectuez l'attribution de la connexion pour les ressources. Enterprise Data Catalog crée ensuite des liens, indépendamment de la sensibilité à la casse dans la source et la cible, et affiche des informations de lignage.

L'image suivante affiche l'exemple de vue de lignage et d'impact d'une table insensible à la casse :



Activation de la découverte de données

Vous pouvez sélectionner l'option **Activer la découverte de données** pour une ressource afin qu'elle trouve le contenu, la qualité et la structure de la source de données. Vous pouvez exécuter un profil de colonne, effectuer la découverte de domaine de données et préparer des données pour inférer des colonnes similaires dans plusieurs sources de données. Vous pouvez générer une clé unique à partir de la colonne ou d'une combinaison de colonnes.

Pour effectuer la découverte de données sur une ressource, sélectionnez l'option **Activer la découverte de données** de la section **Découverte de données** dans l'onglet **Paramètres de chargement des métadonnées**. Une fois que la découverte de données est activée pour une ressource, vous pouvez configurer les propriétés du service d'intégration de données, les paramètres liés au profil, les paramètres d'inférence de clé unique et activer la similarité de colonne. Pour plus d'informations sur les options de la section **Découverte de données**, voir la rubrique ["Découverte de données" à la page 31](#).

Choisissez les options de profilage suivantes pour les types de fichiers non structurés et les formats non structurés étendus :

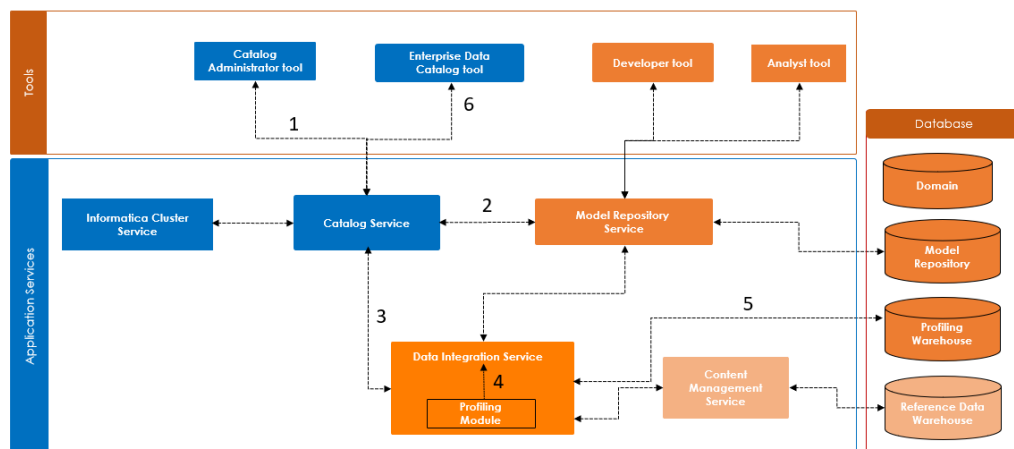
- Option d'exécution du profil **Découverte de domaines de données** ou **Profil de colonne et découverte de domaines de données**.
- Options d'échantillonnage **Toutes les lignes**.
- **Lignes** comme critère de correspondance de domaine de données. Le nombre de lignes est le nombre d'occurrences d'un domaine de données dans une source de données.

Vous ne pouvez pas exécuter la découverte sur les métadonnées source pour les types de fichiers non structurés et les formats non structurés étendus.

Architecture de profilage

L'architecture de profilage comprend des outils, des services et des bases de données. Le composant Outils comprend les applications clientes. Le composant Services comprend des services d'applications nécessaires pour gérer les outils, effectuer les tâches d'intégration de données et gérer les métadonnées des objets de profils. Le composant de bases de données comprend le référentiel modèle, l'entrepôt de profilage et l'entrepôt de données de référence.

L'image suivante montre les composants de l'architecture du profilage :



Le tableau suivant décrit les composants de l'architecture :

Composant	Description
Outil Catalog Administrator	Outil pour configurer les paramètres de métadonnées de profilage afin que vous disposiez d'informations relatives à la qualité des données pour une analyse plus approfondie.
Outil Enterprise Data Catalog	Outil pour découvrir, explorer et relier différents types de métadonnées provenant de sources disparates dans l'entreprise. Vous pouvez afficher les résultats du profilage dans l'outil Enterprise Data Catalog.
Outil Developer tool	Application cliente que vous utilisez pour concevoir et implémenter des solutions d'intégration des données, de qualité des données, de profilage des données et de services des données. Vous pouvez utiliser l'outil Developer tool pour créer des règles de profilage.
Outil Analyst tool	Outil client Web disponible pour plusieurs produits Informatica et qui est utilisé par les utilisateurs professionnels pour collaborer à des projets d'une organisation. Vous pouvez utiliser l'outil Analyst tool pour créer des règles de profilage.
Service de cluster Informatica	Service d'application qui exécute et gère tous les services Hadoop, le serveur Apache Ambari et les agents Apache Ambari sur un cluster Hadoop intégré.
Service de catalogue	Service d'application qui exécute et gère les connexions entre les composants de service et les utilisateurs ayant accès aux outils Catalog Administrator et Enterprise Data Catalog.
Service de référentiel modèle	Service d'application qui gère le référentiel modèle.
Service d'intégration de données	Service d'application qui effectue des tâches d'intégration de données pour l'outil Analyst tool, l'outil Developer tool et les clients externes.
Service de gestion de contenu	Service d'application qui gère les données de référence pour les domaines de données qui utilisent des tables de référence. Il utilise le service d'intégration de données pour exécuter des mappages en vue de transférer des données entre les tables de référence et les sources de données externes.
Référentiel modèle	Base de données relationnelle qui stocke les métadonnées de projets créés par l'outil Analyst tool ou l'outil Developer tool.
Entrepôt de profilage	Base de données qui stocke les informations de profilage, telles que les résultats de profils.
Entrepôt de données de référence	Base de données qui stocke des valeurs de données pour les objets de table de référence que vous définissez dans le référentiel modèle.

Comme illustré dans le diagramme d'architecture de profilage, les processus suivants se produisent lorsque vous activez le profilage pour une ressource et analysez cette dernière :

1. Activez la découverte de données pour la ressource dans l'outil Catalog Administrator afin de découvrir les métadonnées liées au profilage et les clés uniques.
2. Le service de catalogue obtient la définition de profil à partir du référentiel modèle.
3. Le service de catalogue appelle le module de profilage dans le service d'intégration de données.
4. Le module de profilage traite le travail de profil et le soumet au service d'intégration de données.
5. Celui-ci génère les résultats de profilage et les écrit dans l'entrepôt de profilage.

6. Découvrez, explorez et affichez les résultats du profilage dans l'outil Enterprise Data Catalog.

Paramètres de connexion au domaine

Configurez les propriétés du service d'intégration de données. Après avoir configuré les propriétés, le service d'intégration de données exécute le profil, effectue la découverte de domaines de données et infère la similarité de colonne. Vous pouvez choisir un service d'intégration de données différent pour inférer la similarité de colonne.

Le tableau suivant décrit les propriétés que vous pouvez configurer dans l'onglet **Paramètres de chargement des métadonnées > Paramètres de connexion au domaine** :

Propriété	Description
Spécifiez les paramètres de configuration du service d'intégration de données.	Choisissez l'une des options suivantes : <ul style="list-style-type: none">- Personnalisé. Entrez les paramètres de connexion au domaine pour le service d'intégration de données- Global. Choisissez une configuration réutilisable.
Nom du domaine	Nom du domaine du service d'intégration de données.
Service d'intégration de données	Nom du service d'intégration de données.
Nom d'utilisateur	Nom d'utilisateur permettant de vous connecter au service d'intégration de données.
Mot de passe	Mot de passe permettant de vous connecter au service d'intégration de données.
Domaine de sécurité	Nom du domaine de sécurité.
Hôte	Nom d'hôte du service d'intégration de données.
Port	Numéro de port du service d'intégration de données.
Profil de système d'exploitation	<p>Choisissez un profil de système d'exploitation si vous ne disposez pas d'un profil de système d'exploitation par défaut. Le service d'intégration de données se sert du profil de système d'exploitation pour exécuter le profil.</p> <p>Remarque: Vous ne pouvez choisir un profil de système d'exploitation que si vous choisissez l'option Personnalisé pour la propriété Spécifiez les paramètres de configuration du service d'intégration de données. Si vous choisissez l'option Global pour la propriété Spécifiez les paramètres de configuration du service d'intégration de données, vous avez besoin d'un profil de système d'exploitation par défaut pour exécuter le profil.</p> <p>Pour plus d'informations sur l'utilisation des profils de système d'exploitation dans Enterprise Data Catalog, voir la rubrique "Paramètres de connexion au domaine" à la page 33.</p>

Paramètres de profil de base

Configurez les paramètres de profil pour exécuter un profil de colonne et effectuez une découverte de domaines de données pour une ressource.

Le tableau suivant décrit les propriétés que vous pouvez configurer dans la section **Paramètres de chargement des métadonnées > Paramètres de profil de base** :

Propriété	Description
Option d'exécution du profil	Choisissez l'un des types de profils suivants : <ul style="list-style-type: none"> - Profil de colonne. - Découverte de domaines de données. - Profil de colonne et découverte de domaines de données.
Priorité	Choisissez l'une des valeurs de priorité suivantes : <ul style="list-style-type: none"> - Élevée - Faible
Option d'échantillonnage	Choisissez l'une des options d'échantillonnage suivantes : <ul style="list-style-type: none"> - Toutes les lignes. - Lignes aléatoires automatiques. - N lignes aléatoires. - N premières lignes. - Limite de N lignes - Pourcentage aléatoire <p>Remarque: Pour les ressources Hive, choisissez uniquement l'option Toutes les lignes ou l'option d'échantillonnage N premières lignes. Pour les ressources XML, JSON, Avro et Parquet, choisissez uniquement l'option d'échantillonnage Toutes les lignes.</p> <p>Remarque: Pour une ressource Cassandra, sélectionnez l'option d'échantillonnage Toutes les lignes, N premières lignes ou Limite de N lignes.</p>
Exclure les vues	Choisissez cette option si vous ne souhaitez pas que le scanner de profilage analyse les vues dans les sources de données relationnelles. <p>Remarque: Pour une ressource Cassandra, activez l'option Exclure les vues, car la découverte de données n'est pas prise en charge sur les vues.</p>
Profilage incrémentiel	Choisissez cette option pour exécuter le profil uniquement sur les modifications apportées à la source de données. Si vous ne sélectionnez pas cette option, le profil s'exécute sur l'intégralité de la source de données. <p>Pour plus d'informations sur les ressources qui prennent en charge le profilage incrémentiel, voir la rubrique "Paramètres de profil de base" à la page 34.</p>
Filtre de profil de données	Vous pouvez inclure ou exclure des tables et des vues depuis l'exécution du profil. Utilisez des points-virgules (;) pour séparer les noms des tables et les noms des vues. <p>Pour plus d'informations sur le champ de filtre, voir "Filtre de profil des données et métadonnées source" à la page 71.</p>
Cumulatif	Choisissez cette option pour conserver les résultats de similarité de colonne et de profil de colonne issus de l'analyse précédente dans les résultats d'analyse de la ressource suivante. Si vous ne choisissez pas cette option, les résultats de profil précédents sont supprimés. <p>Pour plus d'informations sur l'utilisation de cette option avec le champ Filtre de profil de données et l'option Profilage incrémentiel, voir la rubrique "Paramètres de profil de base" à la page 34.</p>

Propriété	Description
Nom de la connexion source	Choisissez une connexion source pour exécuter la découverte de données. Remarque: Ce paramètre est facultatif pour une ressource du Ressource du système de fichiers .
Exécuter sur	Choisissez l'un des environnements d'exécution suivants : <ul style="list-style-type: none"> - Blaze. Exécute le profil dans l'environnement Hadoop sur le moteur Blaze. Cliquez sur Sélectionner... et choisissez un nom de connexion Hadoop dans la boîte de dialogue Sélectionner le nom de la connexion Hadoop. - Spark. Exécute le profil dans l'environnement Hadoop sur le moteur Spark. Cliquez sur Sélectionner... et choisissez un nom de connexion Hadoop dans la boîte de dialogue Sélectionner le nom de la connexion Hadoop. - Databricks. Exécute le profil dans l'environnement Databricks sur le moteur Databricks Spark dans le cluster Databricks. Cliquez sur Sélectionner... et choisissez un nom de connexion Hadoop dans la boîte de dialogue Sélectionner le nom de la connexion Hadoop. Remarque: Vous pouvez exécuter des ressources JDBC et Azure Data Lake Store à l'aide de l'environnement d'exécution Databricks. <ul style="list-style-type: none"> - Native. Exécute le profil sur la même machine sur laquelle le service d'intégration de données s'exécute.

Vous pouvez choisir un environnement d'exécution natif, Hadoop ou Databricks pour un profil de colonne. Vous pouvez choisir un moteur Blaze ou Spark dans l'environnement d'exécution Hadoop. Vous pouvez choisir l'option Databricks Spark dans l'environnement d'exécution Databricks. Enterprise Data Catalog définit l'environnement d'exécution dans la définition de profil une fois que vous avez choisi un environnement d'exécution.

Remarque: Choisissez Blaze ou Natif comme environnement d'exécution pour exécuter le travail de profil pour toutes les ressources, à l'exception des ressources Hive. Lorsque vous choisissez le moteur Blaze ou Spark, sélectionnez une connexion Hadoop pour exécuter les profils.

Vous pouvez exécuter le profil dans les environnements d'exécution suivants :

Environnement natif

Lorsque vous exécutez un profil dans l'environnement d'exécution natif, le service d'intégration de données exécute les mappages et écrit les résultats du profil dans l'entrepôt de profilage. Par défaut, tous les profils s'exécutent dans un environnement d'exécution natif. Vous pouvez utiliser des sources de données natives pour créer et exécuter des profils dans l'environnement natif.

Environnement Hadoop

Vous pouvez exécuter des profils sur le moteur Spark ou Blaze dans l'environnement Hadoop ou sur le moteur Databricks Spark dans l'environnement Databricks. Le service d'intégration de données envoie le traitement au cluster. Une fois le traitement terminé, le service d'intégration de données écrit les résultats du profil dans l'entrepôt de profilage.

Environnement Databricks

Vous pouvez choisir l'option Databricks Spark pour exécuter les profils dans l'environnement d'exécution Databricks. Après avoir choisi l'option Databricks Spark, vous pouvez sélectionner une connexion Databricks. Le service d'intégration de données envoie la logique de profil au moteur Databricks Spark sur le cluster Databricks pour l'exécution des profils. Lorsque vous exécutez un profil dans l'environnement Databricks, l'outil Analyst tool soumet les travaux de profil au module de service de profilage. Celui-ci fractionne alors les travaux de profil en un ensemble de mappages. Le service d'intégration de données envoie les mappages au moteur Databricks Spark via la connexion Hadoop. Le moteur Databricks Spark traite les mappages, et le service d'intégration de données écrit les résultats de profil dans l'entrepôt de profilage.

Configurez les propriétés suivantes lorsque vous choisissez l'option **Découverte de domaines de données** ou **Profil de colonne et découverte de domaines de données** :

Propriété	Description
Type de découverte de domaines de données	Choisissez l'un des types de découverte de domaines de données suivants : <ul style="list-style-type: none"> - Exécuter la découverte sur les données sources. - Exécuter la découverte sur les métadonnées sources. - Exécuter la découverte sur les métadonnées et les données sources. - Exécuter la découverte sur les données sources où les métadonnées correspondent.
Sélectionner un domaine de données	Choisissez l'une des options de domaine de données suivantes : <ul style="list-style-type: none"> - Tous les domaines de données. - Groupes de domaines de données spécifiques. <p>Dans le champ Groupe de domaines de données, choisissez un ou plusieurs groupe de domaines de données dans la boîte de dialogue Groupe de domaines de données.</p> <ul style="list-style-type: none"> - Domaines de données spécifiques. <p>Dans le champ Domaines de données, choisissez un ou plusieurs domaines de données dans la boîte de dialogue Domaines de données.</p>
Utiliser la conformité de	Choisissez l'une des valeurs suivantes : <ul style="list-style-type: none"> - Domaine de données - Personnalisation
Critères de correspondance de domaine de données	Choisissez l'une des valeurs suivantes : <ul style="list-style-type: none"> - Pourcentage - Lignes
Exclure les valeurs NULL de la découverte de domaines de données	Sélectionnez cette option pour exclure les valeurs NULL de la source de données lorsque vous exécutez la découverte de domaines de données.

Paramètres de profil de similarité et de fréquence de valeur

Configurez les propriétés de similarité de colonne pour identifier des colonnes et une fréquence de valeur similaires dans la ressource.

Le tableau suivant décrit les propriétés que vous pouvez configurer dans l'onglet **Paramètres de chargement des métadonnées > Paramètres de préparation des données du profil de similarité et de fréquence de la valeur** :

Propriété	Description
Exécuter le profil de similarité	<p>Choisissez l'une des options suivantes :</p> <ul style="list-style-type: none"> - Oui. Le scanneur de profilage analyse la source de données et prépare les données pour effectuer les tâches suivantes : <ul style="list-style-type: none"> - Découverte de colonnes similaires. L'algorithme découvre des colonnes similaires dans la ressource selon les noms de colonnes, les modèles de colonnes et les valeurs uniques. - Identification des termes d'entreprise. L'algorithme identifie et recommande les termes d'entreprise pour une colonne selon les domaines de données acceptés et les colonnes similaires. - Non.
Enregistrer les données source	<p>Choisissez l'une des options suivantes :</p> <ul style="list-style-type: none"> - Oui. Conserve les informations calculées sur les colonnes similaires, les modèles de colonnes et les valeurs uniques similaires dans la ressource de la base de données PostgreSQL. - Non.
Options d'échantillonnage	<p>Choisissez l'une des options d'échantillonnage suivantes :</p> <ul style="list-style-type: none"> - Réutiliser les paramètres de profil de base. - Toutes les lignes. - Lignes aléatoires automatiques. - N lignes aléatoires. - N premières lignes.
Paramètres de connexion au domaine	<p>Choisissez l'une des options suivantes :</p> <ul style="list-style-type: none"> - Utiliser les paramètres de configuration du profil. - Spécifier les paramètres de connexion au domaine. <p>Pour plus d'informations sur les propriétés des paramètres de connexion de domaine, consultez la section <i>Paramètres de connexion au domaine</i>.</p>

Paramètres d'inférence de clé unique

Vous pouvez configurer les paramètres suivants pour générer un candidat de clé unique à partir des colonnes sélectionnées.

Propriété	Description
Exécuter l'inférence de clé unique	<p>Choisissez l'une des options suivantes :</p> <ul style="list-style-type: none"> - Oui. Le scanneur de profil analyse et infère les clés uniques de la source de données. - Non
Seuil % Null dans l'inférence de clé unique	Seuil des valeurs Null dans l'inférence de clé unique. Vous pouvez entrer une valeur comprise entre 0 et 1.

Propriété	Description
Ignorer l'inférence de clé unique lorsqu'une clé unique acceptée ou documentée existe	Choisissez l'une des options suivantes : - Oui. Le service de profilage ignore l'inférence de clé unique pour les colonnes avec des clés documentées ou acceptées. - Non
Option d'échantillonnage de la clé unique	Choisissez l'une des options suivantes : - Toutes les lignes. Sélectionne toutes les lignes dans l'objet de données pour l'inférence de clé unique. - N premières lignes. Sélectionne uniquement les N premières lignes dans l'objet de données pour l'inférence de clé unique.

Pour obtenir une liste des types de ressource et de fichier pris en charge, reportez-vous à la section ["Paramètres d'inférence de clé unique" à la page 40](#).

Découverte de domaines de données composites

Comme unique étape prérequis, assurez-vous d'effectuer le profilage avant d'exécuter la découverte de domaines de données composites. Enterprise Data Catalog découvre également les domaines de données composites si des domaines de données déduits existent dans votre environnement. Enterprise Data Catalog découvre les domaines de données qui ne sont pas activés pour la découverte pendant la découverte de domaines de données composites.

Enterprise Data Catalog découvre les domaines de données composites pour les ressources suivantes :

- Amazon Redshift
- Amazon S3
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure SQL Server
- Système de fichiers
- HDFS
- Hive
- IBM DB2
- IBM DB2 for zOS
- IBM Netezza
- JDBC
- Microsoft SQL Server
- Oracle
- Salesforce
- SAP R/3
- Sybase
- Teradata

Les options suivantes de la section **Découverte de domaine composite** vous aident à activer et à configurer la découverte de domaine de données composites pour la ressource :

Propriété	Description
Activer la découverte de domaine composite	Permet la découverte de domaines de données composites pour la ressource.
Sélectionner un domaine de données composites	Sélectionnez les domaines de données composites que la ressource doit utiliser pour l'inférence. Sélectionnez l'une des options suivantes pour spécifier les domaines de données composites : <ul style="list-style-type: none">- Tous les domaines de données composites. Sélectionne tous les domaines de données composites.- Domaines de données composites spécifiques. Sélectionne des domaines de données composites spécifiques. Lorsque vous sélectionnez cette propriété, la zone de texte Domaines de données composites apparaît. Cliquez sur Sélectionner... pour sélectionner une liste de domaines de données composites à partir de la boîte de dialogue Sélectionner un domaine de données composites.

Remarque: Si vous activez la découverte de domaines de données composites et la découverte de domaines de données pour la ressource, Enterprise Data Catalog découvre tous les domaines de données inclus dans les domaines de données composites pour la ressource.

Activer l'association des termes d'entreprise

Configurez les paramètres d'**Association des termes d'entreprise** d'une ressource afin d'associer automatiquement ou de recommander les termes d'entreprise en tant que fonction pour les tables, les colonnes, les vues, les fichiers et les ressources de champ.

Si vous choisissez l'option **Paramètres de chargement des métadonnées > Association des termes d'entreprise > Activer l'association des termes d'entreprise** afin de configurer la propagation des termes d'entreprise, vous ne devez pas configurer l'option **Exécuter le profil de similarité** d'une ressource afin d'accepter automatiquement ou de recommander le terme d'entreprise pour des colonnes ou des ressources de champ. Configurez les options **Exécuter le profil de similarité** et **Activer l'association des termes d'entreprise** d'une ressource pour améliorer le score de confiance de la propagation des termes d'entreprise.

Enterprise Data Catalog prend en charge l'option **Activer l'association des termes d'entreprise** pour les ressources suivantes :

- Amazon S3
- Amazon Redshift
- Azure Data Lake Store
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure SQL Server
- Système de fichiers
- Google BigQuery
- HDFS
- Hive

- IBM DB2 for z/OS
- IBM DB2
- IBM Netezza
- JDBC
- Microsoft Azure Blob Storage
- Microsoft SQL Server
- OneDrive
- Oracle
- Salesforce
- SAP
- SharePoint
- Snowflake
- Sybase
- Teradata
- Workday
- MongoDB

Remarque: Après avoir configuré les paramètres d'association des termes d'entreprise, exécutez la ressource système DataDomainPropagation pour afficher les recommandations de termes d'entreprise immédiatement dans l'application Enterprise Data Catalog.

Configurez les propriétés suivantes dans la section **Association des termes d'entreprise** de la page **Paramètres de chargement des métadonnées** pour associer automatiquement les termes d'entreprise :

Propriété	Description
Activer l'association des termes d'entreprise	Active l'association des termes d'entreprise pour les colonnes et champs dans la ressource.
Sélectionner des glossaires	Sélectionnez les glossaires d'entreprise ou les glossaires Axon à partir desquels vous voulez qu'Enterprise Data Catalog associe automatiquement la fonction aux ressources.

Propriété	Description
Utiliser des synonymes et des mots vides	<p>Sélectionnez cette option pour utiliser des synonymes et des mots vides lors du processus de propagation des termes d'entreprise. « the », « a », « an » et « in » sont des exemples de mots vides.</p> <p>Par exemple, supposez une ressource qui contient une colonne nommée Rapport de profits et pertes. Le mot « and » est un mot vide, et vous voulez qu'Enterprise Data Catalog attribue ou recommande le terme d'entreprise nommé ProfitLossReport pour la colonne. Après avoir sélectionné l'option Utiliser des synonymes et des mots vides, Enterprise Data Catalog évite le mot vide « and » dans le nom de la colonne, puis attribue ou recommande le terme d'entreprise ProfitLossReport pour la colonne.</p> <p>Chargez la liste de définitions des synonymes dans Catalog Administrator pour qu'Enterprise Data Catalog utilise des synonymes afin de propager les termes d'entreprise.</p> <p>Par exemple, Enterprise Data Catalog contient un terme d'entreprise nommé Numéro de sécurité sociale et une colonne nommée NSS. Vous voulez que le catalogue associe automatiquement ou recommande le terme d'entreprise Numéro de sécurité sociale pour la colonne NSS. Après avoir défini la liste de définitions de synonymes dans Catalog Administrator, Enterprise Data Catalog recommande ou attribue Numéro de sécurité sociale comme fonction à la colonne.</p> <p>Pour plus d'informations sur les définitions de synonymes, consultez la section Chapitre 13, "Gérer les définitions de synonymes" à la page 127.</p>
Éviter les préfixes	<p>Sélectionnez cette option pour éviter des préfixes dans le nom de la ressource lors du processus de propagation des termes d'entreprise.</p> <p>Par exemple, supposez une ressource qui contient une colonne nommée Reorderdetails. Vous voulez qu'Enterprise Data Catalog recommande ou attribue automatiquement le terme d'entreprise orderdetails. Après avoir sélectionné l'option Éviter les préfixes, Enterprise Data Catalog ignore le préfixe Re dans le nom de la colonne, puis attribue automatiquement le terme d'entreprise orderdetails à la colonne Reorderdetails.</p>
Options personnalisées	<p>Paramètres que vous pouvez définir pour configurer la propagation des termes d'entreprise. Utilisez les options personnalisées de pondération de fonctionnalité pour personnaliser la priorité des domaines de données, la similarité de colonnes et les conditions de correspondance de noms lors du processus de propagation des termes d'entreprise. Utilisez les options personnalisées de seuil pour personnaliser la limite de seuil pour les associations et les recommandations automatiques de termes d'entreprise. Selon la limite de seuil que vous définissez, Enterprise Data Catalog attribue des termes d'entreprise ou recommande des termes d'entreprise aux actifs.</p> <p>Pour configurer la pondération de la fonctionnalité, ajoutez les options personnalisées suivantes à la ressource de propagation des termes d'entreprise :</p> <ul style="list-style-type: none"> - <code>-Dglossary.propagation.domain.curation.feature.weight=<value></code> <ul style="list-style-type: none"> - La valeur par défaut est 3. - <code>-Dglossary.propagation.column.similarity.feature.weight=<value></code> <ul style="list-style-type: none"> - La valeur par défaut est 1. <p>Pour configurer les seuils, ajoutez les options personnalisées suivantes à la ressource de propagation des termes d'entreprise :</p> <ul style="list-style-type: none"> - <code>-Dglossary.propagation.autoaccept.score.threshold=<value></code> <ul style="list-style-type: none"> - La valeur par défaut est 80. - <code>-Dglossary.propagation.inference.score.threshold=<value></code> <ul style="list-style-type: none"> - La valeur par défaut est 60.

Pour plus d'informations sur la propagation des termes d'entreprise, consultez la rubrique ["Propagation des termes d'entreprise" à la page 47](#).

Activer le provisionnement de données

En une seule étape préalable, vérifiez que vous avez importé l'organisation Informatica Intelligent Cloud Services dans Catalog Administrator. Configurez les paramètres de **Provisionnement de données** d'une ressource pour créer des tâches de provisionnement de données dans Enterprise Data Catalog. Enterprise Data Catalog prend en charge l'option **Activer le provisionnement de données** pour les ressources Oracle et Microsoft SQL Server.

Configurez les propriétés suivantes sur la page **Accès aux données** pour activer le provisionnement de données d'une ressource :

Avant d'activer le provisionnement de données, vérifiez que les prérequis figurant dans la section [Data Preview and Provisioning Prerequisites](#) sont appliqués.

Propriété	Description
Activer le provisionnement de données	Active le provisionnement de données pour les tables relationnelles dans la ressource.
Nom de l'organisation	Choisissez l'organisation Informatica Intelligent Cloud Services dans laquelle vous avez créé les connexions source et cible. Pour plus d'informations sur les organisations Informatica Intelligent Cloud Services, reportez-vous au chapitre Managing Cloud Organization du guide <i>Catalog Administrator</i> .
Connexion	Sélectionnez la connexion que vous avez créée dans Informatica Intelligent Cloud Services pour la ressource. Vérifiez que la connexion que vous sélectionnez comprend le même schéma configuré pour une ressource.

Pour plus d'informations sur le provisionnement de données, consultez la rubrique ["Provisionnement de données" à la page 53](#).

Activer la préparation des données

Avant d'activer la préparation des données, vérifiez que les prérequis figurant dans la section ["Prérequis pour la préparation des données" à la page 50](#) sont appliqués.

Configurez les propriétés suivantes sur la page **Accès aux données** pour activer la préparation des données d'une ressource :

Propriété	Description
Activer la préparation des données	Permet la préparation des données pour les ressources relationnelles.
Nom de l'organisation	Choisissez l'organisation Informatica Intelligent Cloud Services dans laquelle vous avez créé les connexions source. Pour plus d'informations sur les organisations Informatica Intelligent Cloud Services, reportez-vous au chapitre Managing Cloud Organization du guide <i>Catalog Administrator</i> .
Nom de la source du catalogue	Spécifiez le nom de la connexion à la source de catalogue associée au jeu de données.

Pour plus d'informations sur la préparation des données, reportez-vous à la section ["Préparation des données" à la page 56.](#)

Modification d'une ressource

Vous pouvez apporter des modifications à une ressource après l'avoir créée. Vous pouvez modifier les paramètres, tels que les propriétés de la connexion, les attributs personnalisés, les paramètres de métadonnées de source, les paramètres de métadonnées de profil et la planification associée. Une fois que la ressource est créée, vous ne pouvez pas modifier son nom ni son type.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Bibliothèque**
L'espace de travail **Bibliothèque** s'affiche.
2. Dans la liste des ressources, pointez vers une ressource, puis sélectionnez **Modifier** dans le menu système.
Les informations de la ressource apparaissent dans un nouvel onglet.
3. Modifiez la description, les propriétés de la connexion, les attributs personnalisés, les paramètres de métadonnées de source, les paramètres de métadonnées de profil et la planification selon les besoins.
4. Pour enregistrer les modifications sans réexécuter la ressource, cliquez sur **Enregistrer**.
5. Pour enregistrer les modifications et exécuter l'analyse, cliquez sur **Enregistrer et exécuter**.

Clonage d'une ressource

Vous pouvez créer une copie d'une ressource en clonant la ressource. Cette fonctionnalité est utile si vous souhaitez copier la configuration d'une ressource existante, puis modifier la configuration si nécessaire. Lorsqu'une ressource est clonée, certaines configurations de ressources telles que les mots de passe, les chargements de fichiers et les attributions de connexion ne sont pas copiées. Vous ne pouvez pas cloner une ressource système.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Bibliothèque**
L'espace de travail **Bibliothèque** s'affiche.
2. Dans la liste des ressources, pointez sur la ressource à cloner et sélectionnez **Cloner** dans le menu de contrôle.
Un message s'affiche vous demandant de confirmer si vous souhaitez continuer.
3. Sélectionnez **OK**.
Les détails de la ressource s'affichent.
Remarque: Vous pouvez également ouvrir une ressource et cliquer sur le bouton **Cloner** pour accéder à la page de détails de la ressource et créer une ressource à partir du clone.
4. Modifiez le nom de la ressource et apportez les modifications nécessaires à la configuration de la ressource copiée.
5. Pour enregistrer les modifications sans réexécuter la ressource, cliquez sur **Enregistrer**.
6. Pour enregistrer les modifications et exécuter l'analyse, cliquez sur **Enregistrer et exécuter**.

Exécution d'une analyse sur une ressource

Vous pouvez exécuter une analyse sur une ressource dans le cadre de la planification des ressources ou, manuellement, en tant que tâche unique en fonction de vos besoins.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Bibliothèque**
L'espace de travail **Bibliothèque** s'affiche.
2. Dans la liste des ressources, pointez vers une ressource et sélectionnez **Exécuter** dans le menu système.
Les détails de la ressource s'affichent sur une nouvelle page avec l'onglet **Surveillance** activé.

Ressources système

Enterprise Data Catalog crée les ressources système suivantes pour effectuer diverses tâches internes lorsque vous activez le service de catalogue :

DataDomain

Tâche système interne planifiée pour s'exécuter automatiquement toutes les 45 minutes. Le scanneur synchronise les domaines de données entre le référentiel modèle et le catalogue.

DataDomainPropagation

Tâche système interne planifiée pour s'exécuter automatiquement à intervalles réguliers. Utilisation des colonnes semblables déduites pour propager les domaines de données intelligents à d'autres colonnes semblables. Le scanneur identifie et propage également les termes d'entreprise à d'autres colonnes similaires.

DomainUsers

Tâche système interne qui synchronise les utilisateurs du domaine Informatica entre le domaine et le catalogue Informatica.

Vous pouvez configurer les paramètres suivants pour les ressources en fonction de vos besoins, comme vous configurez toutes les autres ressources que vous créez :

- Modifiez la mémoire requise pour exécuter la ressource.
- Attribuez des attributs personnalisés à la ressource.
- Modifiez la planification de la ressource.

Remarque: Vous pouvez modifier la description des scanners, mais vous ne pouvez pas modifier les informations de nom ou de type de ressource des ressources système.

Affichage d'une ressource

Vous pouvez afficher la liste des ressources dans l'onglet Bibliothèque. Vous pouvez lancer une vue en lecture seule d'une ressource spécifique. Vous pouvez également modifier une ressource à partir de la liste des ressources.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Bibliothèque**
L'espace de travail **Bibliothèque** s'affiche.
2. Dans la liste des ressources, pointez vers une ressource et sélectionnez **Ouvrir** dans le menu système.
Une vue en lecture seule des détails de la ressource s'affiche sur l'onglet **Général**. Vous pouvez voir plusieurs onglets relatifs aux informations sur la ressource.
3. Cliquez sur chaque onglet pour afficher plus d'informations sur la ressource.

CHAPITRE 5

Gestion de la sécurité des ressources

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion de la sécurité des ressources, 88](#)
- [Configuration des autorisations par défaut pour les ressources, 89](#)
- [Configuration des autorisations pour des utilisateurs et des groupes d'utilisateurs spécifiques, 90](#)
- [Application de limitations à une ressource, 92](#)
- [Attribution des autorisations de configuration aux ressources, 93](#)

Présentation de la gestion de la sécurité des ressources

Vous pouvez configurer des autorisations spécifiques sur les ressources pour les utilisateurs et les groupes d'utilisateurs configurés dans le domaine Informatica.

En tant qu'administrateur de catalogue, vous pouvez spécifier des autorisations d'accès sur des ressources pour des utilisateurs et des groupes d'utilisateurs spécifiques. Le type d'autorisation d'accès dépend des exigences spécifiques de sécurité et de privilèges de votre entreprise. Par exemple, dans une institution financière, en dehors du gestionnaire des données qui valide l'intégrité, la cohérence et la qualité des données, personne ne doit être en mesure d'afficher les informations des sources de données qui stockent les informations confidentielles des clients. L'identification des sources de données qui stockent les informations des clients par du personnel non-autorisé pourrait entraîner le piratage des sources de données et la fuite des informations confidentielles.

Vous pouvez spécifier les autorisations à l'aide de Catalog Administrator de la manière suivante :

- Spécifiez les autorisations par défaut sur toutes les ressources ou sur des ressources spécifiques pour les utilisateurs et les groupes d'utilisateurs.
- Sélectionnez une ressource spécifique et spécifiez les autorisations pour les utilisateurs et les groupes d'utilisateurs.
- Sélectionnez des utilisateurs ou des groupes d'utilisateurs spécifiques et configurez les autorisations sur les ressources.

En tant qu'administrateur de catalogue, vous pouvez attribuer aux utilisateurs et aux groupes d'utilisateurs les autorisations suivantes sur les ressources :

Lecture

Affichez les informations des ressources dans Enterprise Data Catalog.

Lecture et écriture

L'utilisateur ou les utilisateurs inclus dans le groupe d'utilisateurs disposent d'une autorisation d'accès en lecture et peuvent enrichir les ressources d'Enterprise Data Catalog. Vous pouvez enrichir les ressources en attribuant des attributs personnalisés, des termes d'entreprise ou des domaines de données à la ressource. L'enrichissement des ressources vous permet de rechercher la ressource à l'aide de l'attribut personnalisé, du terme d'entreprise ou du domaine de données attribué.

Remarque: Lorsque vous configurez les autorisations de données pour les sources relationnelles telles qu'Oracle, les utilisateurs ou les utilisateurs inclus dans les groupes d'utilisateurs disposent des autorisations sur tous les actifs, y compris les tables, les vues et les synonymes. Toutefois, lorsque vous configurez les autorisations de données pour les actifs suivants, les utilisateurs ou les utilisateurs inclus dans les groupes d'utilisateurs disposent des autorisations pour accéder uniquement à ces actifs :

- Tables
- Vues
- Synonymes

Lecture des métadonnées et des données

Permet à un ou à plusieurs utilisateurs inclus dans le groupe d'utilisateurs d'afficher la fréquence de valeur d'un actif dans Enterprise Data Catalog. Vous pouvez attribuer cette autorisation aux ressources qui prennent en charge le profilage. Pour obtenir la liste de ces ressources, reportez-vous à ["Découverte de données" à la page 31](#).

Toutes les autorisations

Permet à un ou à plusieurs utilisateurs inclus dans le groupe d'utilisateurs d'enrichir les actifs dans Enterprise Data Catalog et d'afficher la fréquence de valeur de l'actif. Vous pouvez enrichir les ressources en attribuant des attributs personnalisés, des termes d'entreprise ou des domaines de données à la ressource. L'enrichissement des ressources vous permet de rechercher la ressource à l'aide de l'attribut personnalisé, du terme d'entreprise ou du domaine de données attribué.

Configuration des autorisations par défaut pour les ressources

Vous pouvez spécifier les autorisations par défaut sur les ressources pour les utilisateurs et les groupes d'utilisateurs. Vous pouvez modifier les autorisations par défaut ultérieurement en fonction des besoins des utilisateurs et des groupes d'utilisateurs spécifiques. Enterprise Data Catalog applique les autorisations par défaut pour tous les utilisateurs ou groupes d'utilisateurs jusqu'à ce que vous spécifiez des autorisations pour un utilisateur ou un groupe d'utilisateurs. Les autorisations spécifiques que vous configurez pour les utilisateurs ou les groupes d'utilisateurs prennent effet uniquement sur les ressources pour lesquelles vous octroyez des autorisations en modifiant les autorisations par défaut.

1. Cliquez sur **Gérer > Sécurité**.

L'onglet **Sécurité** s'ouvre.

2. Cliquez sur **Définir les autorisations par défaut**.

La boîte de dialogue **Autorisations par défaut** s'affiche et répertorie les utilisateurs et les groupes configurés sur le domaine Informatica.

3. Sélectionnez les utilisateurs et les groupes d'utilisateurs pour lesquels vous voulez configurer les autorisations sur les ressources, puis cliquez sur **Suivant**.

La section **Spécifier les autorisations des utilisateurs sélectionnés et des groupes** apparaît.

4. Sélectionnez l'une des options suivantes pour spécifier les ressources sur lesquelles vous voulez attribuer des autorisations pour les utilisateurs ou groupes d'utilisateurs sélectionnés :
 - **Toutes les ressources**
Sélectionnez cette option si vous voulez spécifier les autorisations par défaut pour toutes les ressources dans Enterprise Data Catalog. Sélectionnez **Lecture** ou **Lecture et écriture** dans la liste déroulante pour attribuer l'autorisation sélectionnée à toutes les ressources.
 - **Personnalisation**
Sélectionnez cette option si vous voulez sélectionner les ressources pour lesquelles spécifier des autorisations. Sélectionnez les ressources requises et cliquez sur la liste déroulante adjacente à la ressource. Sélectionnez l'autorisation requise pour la ressource dans la liste déroulante. Si vous sélectionnez plusieurs ressources, cliquez sur la liste déroulante en regard d'***Autorisations héritées** et sélectionnez l'autorisation requise à appliquer pour toutes les ressources.

Remarque:

- Types de ressources. Entrez le nom de la ressource dans la zone de texte sous la section **Type de ressource** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les ressources avec des modèles correspondants.
- Autorisations. Sélectionnez l'autorisation dans la liste déroulante **Autorisations**. Enterprise Data Catalog répertorie les ressources qui ont l'autorisation correspondante.

Cliquez sur **Effacer le filtre** pour effacer les options de filtre que vous avez spécifiées.

5. Cliquez sur **OK**.

Configuration des autorisations pour des utilisateurs et des groupes d'utilisateurs spécifiques

Vous pouvez configurer des autorisations spécifiques pour les utilisateurs et les groupes d'utilisateurs afin qu'ils accèdent aux ressources.

Enterprise Data Catalog conserve les autorisations par défaut (autorisations héritées) configurées pour les utilisateurs ou les groupes d'utilisateurs jusqu'à ce que vous configuriez des autorisations spécifiques. Enterprise Data Catalog ajoute un astérisque (*) au nom de l'autorisation pour indiquer que l'utilisateur ou le groupe d'utilisateurs sélectionné possède les autorisations par défaut.

Remarque: Par défaut, les utilisateurs inclus dans un groupe d'utilisateurs héritent des autorisations attribuées au groupe. Si vous configurez des autorisations spécifiques pour un utilisateur dans un groupe d'utilisateurs, Enterprise Data Catalog applique à l'utilisateur les autorisations spécifiques configurées pour lui. Enterprise Data Catalog affiche le statut d'autorisation comme **Mixte** pour un utilisateur disposant d'autorisations différentes configurées au niveau de l'utilisateur et au niveau du groupe d'utilisateurs.

Pour sélectionner des utilisateurs et des groupes d'utilisateurs afin de leur attribuer des autorisations sur la ressource, effectuez les étapes suivantes :

1. Cliquez sur **Gérer > Sécurité** à partir de l'en-tête de Catalog Administrator.
L'onglet **Sécurité** s'ouvre avec l'option **Utilisateurs et groupes** sélectionnée dans la section **Afficher**.
2. Dans le panneau **Utilisateurs et groupes**, sélectionnez l'utilisateur ou le groupe d'utilisateurs requis.

Vous pouvez utiliser les filtres suivants pour répertorier les utilisateurs ou les groupes d'utilisateurs qui correspondent aux critères spécifiques :

- Nom. Entrez le nom de l'utilisateur ou du groupe d'utilisateurs dans la zone de texte sous **Nom** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les utilisateurs et les groupes d'utilisateurs correspondants.
- Type. Sélectionnez **Utilisateur** ou **Groupe** dans la liste déroulante sous **Type** pour spécifier si vous souhaitez afficher une liste d'utilisateurs ou de groupes.
- Domaine de sécurité. Entrez le domaine de sécurité configuré pour l'utilisateur ou le groupe d'utilisateurs dans la zone de texte sous **Domaine de sécurité**. Enterprise Data Catalog répertorie tous les utilisateurs et groupes d'utilisateurs configurés avec le domaine de sécurité spécifié.

Cliquez sur **Effacer le filtre** pour effacer les options de filtre que vous avez spécifiées.

3. Sélectionnez les ressources requises à partir du panneau **Ressources**, cliquez sur la colonne d'autorisation pour chaque ressource et sélectionnez l'autorisation requise à configurer sur la ressource.

Vous pouvez utiliser les filtres suivants pour répertorier les ressources en conséquence :

- Nom. Entrez le nom de la ressource dans la zone de texte sous la section **Nom** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les ressources avec des noms correspondants.
- Autorisations de données. Sélectionnez l'autorisation dans la liste déroulante **Autorisations de données**. Enterprise Data Catalog répertorie les ressources pour lesquelles l'autorisation de données sélectionnée est configurée.
- Type de ressource. Entrez le type de la ressource dans la zone de texte sous la section **Type de ressource** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les types de ressources correspondants.

Cliquez sur **Effacer le filtre** pour effacer les options de filtre que vous avez spécifiées.

Sélection de ressources pour attribuer des autorisations à des utilisateurs ou des groupes d'utilisateurs spécifiques

Comme alternative à la sélection des utilisateurs ou des groupes d'utilisateurs et l'attribution des autorisations pour les ressources, vous pouvez sélectionner plusieurs ressources et attribuer des autorisations pour des utilisateurs spécifiques ou des groupes d'utilisateurs.

1. Cliquez sur **Gérer > Sécurité**.

L'onglet **Sécurité** s'ouvre avec l'option **Utilisateurs et groupes** sélectionnée dans la section **Afficher**.

2. Sélectionnez **Ressources** à partir de la section **Afficher**.

Vous pouvez utiliser les filtres suivants pour répertorier les utilisateurs ou les groupes d'utilisateurs requis :

- Nom. Entrez le nom de la ressource dans la zone de texte sous **Nom** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les ressources correspondantes.
- Type de ressource. Entrez le type de la ressource dans la zone de texte sous **Type de ressource** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie la liste des types de ressources correspondants.
- Cliquez sur **Effacer le filtre** pour effacer les options de filtre que vous avez spécifiées.

La liste des ressources configurées dans Enterprise Data Catalog s'affiche.

3. Sélectionnez la ressource pour laquelle vous voulez attribuer des autorisations de données et de configuration pour les utilisateurs ou les groupes d'utilisateurs.

4. Effectuez cette étape si vous voulez attribuer des autorisations de données et de configuration à des groupes d'utilisateurs. Sélectionnez le groupe d'utilisateurs requis dans la page d'onglet **Groupes**, cliquez sur les colonnes d'autorisation de données et de configuration pour ce groupe d'utilisateurs, puis sélectionnez les autorisations de données et de configuration que vous souhaitez configurer pour le groupe d'utilisateurs.
5. Effectuez cette étape si vous voulez attribuer des autorisations de données et de configuration à des utilisateurs. Sélectionnez l'utilisateur requis dans la page d'onglet **Utilisateurs**, cliquez sur les colonnes d'autorisation de données et de configuration pour cet utilisateur, puis sélectionnez les autorisations de données et de configuration que vous souhaitez configurer pour l'utilisateur.

Vous pouvez utiliser les options de filtrage suivantes dans les pages d'onglet **Groupes** ou **Utilisateurs** pour filtrer la liste des utilisateurs et des groupes d'utilisateurs en fonction du nom, du domaine de sécurité, des autorisations de données et des autorisations de configuration configurés pour les utilisateurs ou les groupes d'utilisateurs :

- Nom. Entrez le nom de l'utilisateur ou du groupe d'utilisateurs dans la zone de texte sous **Nom** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les utilisateurs ou les groupes d'utilisateurs avec des noms correspondants.
- Autorisations de configuration. Sélectionnez l'autorisation dans la liste déroulante **Autorisations de configuration** pour filtrer les utilisateurs ou les groupes d'utilisateurs selon les autorisations de configuration attribuées. Enterprise Data Catalog répertorie les utilisateurs ou les groupes d'utilisateurs pour lesquels l'autorisation de configuration sélectionnée est configurée.
- Autorisations de données. Sélectionnez l'autorisation dans la liste déroulante **Autorisations de données** pour filtrer les utilisateurs ou les groupes d'utilisateurs selon les autorisations de configuration attribuées. Enterprise Data Catalog répertorie les utilisateurs ou les groupes d'utilisateurs pour lesquels l'autorisation de données sélectionnée est configurée.
- Domaine de sécurité. Entrez le domaine de sécurité dans la zone de texte sous la section **Domaine de sécurité** et appuyez sur la touche **Entrée**. Enterprise Data Catalog répertorie les utilisateurs ou les groupes d'utilisateurs configurés dans le domaine de sécurité spécifié.

Cliquez sur **Effacer le filtre** pour effacer les options de filtre que vous avez spécifiées.

Application de limitations à une ressource

En tant qu'administrateur de catalogue, vous pouvez contrôler les opérations effectuées sur une ressource en limitant celle-ci. Par défaut, toutes les ressources ne sont pas limitées.

1. Cliquez sur **Gérer > Sécurité**.
L'onglet **Sécurité** s'ouvre avec l'option **Utilisateurs et groupes** sélectionnée dans la section **Afficher**.
2. Sélectionnez **Ressources** à partir de la section **Afficher**.
La liste des ressources configurées dans Enterprise Data Catalog s'affiche.
3. Pour appliquer des limitations à une ressource, cliquez sur le menu **Actions**, puis sélectionnez **Limiter**.
Le statut des ressources est défini sur **Cette ressource est limitée**.

Remarque: Les utilisateurs ou les groupes d'utilisateurs disposant des autorisations de configuration de lecture, d'analyse et d'écriture sur une ressource peuvent annuler la limitation de celle-ci.

Attribution des autorisations de configuration aux ressources

En tant qu'administrateur de catalogue, vous pouvez attribuer des autorisations de configuration aux ressources pour contrôler les opérations effectuées sur celles-ci.

Vous pouvez attribuer les autorisations de configuration suivantes aux ressources :

Lire

L'utilisateur ou les utilisateurs du groupe d'utilisateurs peuvent afficher les détails de la ressource, les tâches associées, les connexions et les autorisations de données dans Catalog Administrator.

Lire et analyser

L'utilisateur ou les utilisateurs du groupe d'utilisateurs peuvent afficher les détails de la ressource, les tâches associées, les connexions et les autorisations de données, exécuter l'analyse des métadonnées et effectuer les opérations de restauration et de purge sur la ressource.

Lire, analyser et écrire

L'utilisateur ou les utilisateurs du groupe d'utilisateurs peuvent afficher les détails de la ressource, les tâches associées, les connexions et les autorisations de données, exécuter l'analyse des métadonnées, modifier la ressource, supprimer la ressource, attribuer ou supprimer des connexions, mettre à jour les autorisations de données et effectuer les opérations de restauration et de purge sur la ressource.

Aucun accès

L'utilisateur ou les utilisateurs du groupe d'utilisateurs ne peuvent pas accéder à la ressource.

Remarque: L'utilisateur ou les utilisateurs du groupe d'utilisateurs doivent disposer des autorisations de configuration Lire, Analyser et Écrire sur les ressources participant à la connexion pour afficher, attribuer ou supprimer des connexions.

CHAPITRE 6

Gestion des planifications

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion des planifications, 94](#)
- [Types de planifications, 94](#)
- [Création d'une planification, 95](#)
- [Affichage de la liste des planifications, 95](#)

Présentation de la gestion des planifications

Les planifications déterminent le moment où les scanners extraient les métadonnées des sources. Vous pouvez avoir des planifications récurrentes quotidiennes, hebdomadaires et mensuelles pour extraire des métadonnées à intervalles réguliers. Créez une planification réutilisable si vous voulez attribuer plusieurs ressources à la même planification. Si vous choisissez d'avoir une planification réutilisable pour l'extraction des métadonnées, vous pouvez la sélectionner dans une liste de planifications existantes ou créer une planification réutilisable différente qui répond à vos besoins.

Créez des planifications personnalisées que vous pouvez attribuer à des ressources spécifiques. Vous pouvez attribuer des planifications distinctes aux ressources pour extraire les métadonnées source et le profilage des métadonnées. Lorsque vous créez une planification, vous pouvez choisir d'avoir une planification sans date de fin ou qui revient jusqu'à une date spécifique.

Types de planifications

Vous pouvez créer des planifications réutilisables ou personnalisées qui répondent aux exigences de fréquence de chaque ressource pour extraire des métadonnées.

Vous pouvez joindre plusieurs ressources à une planification réutilisable. Vous pouvez créer une planification personnalisée si vous avez besoin d'une planification séparée spécifique pour une ressource unique. Reportez-vous au fuseau horaire du serveur lorsque vous sélectionnez ou créez des planifications personnalisées ou réutilisables.

Planifications réutilisables

Les métadonnées des systèmes source peuvent être modifiées à des moments différents. Les modifications peuvent inclure de nouvelles ressources de données à ajouter à la source ou des mises à jour des ressources

de données existantes. Vous pouvez définir une planification réutilisable à attribuer à plusieurs ressources afin de continuer à extraire ces modifications de source à intervalles réguliers.

Planifications personnalisées

Créez une planification personnalisée si aucune des planifications réutilisables existantes ne correspond au calendrier d'extraction des métadonnées de la ressource.

Vous pouvez créer une planification personnalisée lorsque vous créez une ressource. Créez une planification personnalisée quotidienne, hebdomadaire ou mensuelle pour une ressource. Vous pouvez créer une planification personnalisée indéfinie ou une planification qui se termine à une date spécifique.

Création d'une planification

Vous pouvez créer une planification lorsque vous configurez une ressource. Vous pouvez créer une planification réutilisable à l'aide du menu **Nouveau** dans l'en-tête de Catalog Administrator.

1. Cliquez sur **Nouvelle > Planification réutilisable**.
L'assistant **Nouvelle planification réutilisable** s'affiche dans l'espace de travail **Planification**.
2. Entrez un nom et une description facultative pour la planification.
3. Vérifiez que le fuseau horaire du serveur affiché est exact.
4. Cliquez sur le champ **Commence le** pour ouvrir un calendrier et choisissez une date de début pour la planification.
5. Utilisez les champs situés à droite du champ **Commence le** pour définir l'heure de début.
6. Choisissez de créer un horaire quotidien, hebdomadaire ou mensuel.
7. Configurez les paramètres de récurrence, tels que tous les n jours pour une planification quotidienne ou un jour de la semaine pour une planification mensuelle.
Les paramètres de récurrence sont différents en fonction de la fréquence de planification.
8. Choisissez une date de fin ou définissez la planification sans date de fin.
9. Cliquez sur **Enregistrer**.

Affichage de la liste des planifications

Vous pouvez consulter la liste des planifications dans l'espace de travail Espace de travail **Bibliothèque**.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Bibliothèque**.
L'espace de travail **Bibliothèque** s'affiche.
2. Dans le volet de gauche, cliquez sur **Planification**.
La liste des planifications s'affiche dans le volet droit.
3. Pour afficher la fréquence de planification, passez le curseur de la souris sur l'icône au début du nom de la planification.

4. Pour afficher les informations complètes ou modifier la planification, cliquez sur le nom de la planification.

La planification s'ouvre dans l'espace de travail Espace de travail **Planification**.

5. Pour apporter des modifications à la planification, cliquez sur **Modifier**.

CHAPITRE 7

Gestion des attributs

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion des attributs, 97](#)
- [Attributs système, 97](#)
- [Attributs personnalisés, 98](#)
- [Propriétés générales des attributs, 98](#)
- [Propriétés de la configuration de recherche, 99](#)
- [Création d'un attribut personnalisé, 99](#)
- [Modification d'un attribut personnalisé et d'un attribut système, 100](#)
- [Attribution d'attributs personnalisés aux classes et aux ressources, 100](#)

Présentation de la gestion des attributs

Les attributs sont des propriétés de métadonnées que les scanners extraient de différents systèmes source. Les attributs système sont des propriétés prédéfinies que les scanners utilisent pour les types de ressources par défaut. Vous pouvez créer des attributs personnalisés, les configurer et les attribuer à des ressources spécifiques.

En fonction des besoins de l'entreprise, vous pouvez choisir d'attribuer des attributs personnalisés aux ressources. Par exemple, vous pouvez attribuer un glossaire d'entreprise ou une catégorie intitulée Ville ou Département à une ressource. Lorsque vous créez un attribut personnalisé, vous pouvez configurer les propriétés d'attribut de base et le comportement de recherche. Par exemple, vous pouvez sélectionner un type de données spécifique, tel que Données, Décimal, Ville, Département ou Utilisateur. Vous pouvez également utiliser l'attribut comme filtre de recherche dans Enterprise Data Catalog où les utilisateurs recherchent les métadonnées d'entreprise requises.

Attributs système

Les attributs système représentent les différents types de métadonnées que les scanners extraient des systèmes source. Par exemple, Auteur est un attribut système du type de données String que vous pouvez attribuer à une ressource. Vous pouvez configurer ces attributs prédéfinis dans Catalog Administrator.

Vous pouvez utiliser des attributs système pour filtrer les résultats lorsque vous recherchez les métadonnées à l'aide d'Enterprise Data Catalog. Utilisez Catalog Administrator pour configurer le classement de recherche

d'un attribut système en fonction de la configuration requise. Vous pouvez également configurer l'attribut système afin qu'Enterprise Data Catalog inclue l'attribut dans les filtres de recherche.

Attributs personnalisés

Vous pouvez créer des attributs personnalisés en fonction des filtres de recherche à utiliser dans Enterprise Data Catalog où vous recherchez des métadonnées. Les attributs personnalisés vous permettent de trouver rapidement des métadonnées spécifiques. Par exemple, vous pouvez créer un attribut personnalisé nommé Emplacement du centre de données dans Catalog Administrator et l'attribuer à certaines ressources. Vous pouvez ensuite utiliser cet attribut personnalisé Emplacement du centre de données dans Enterprise Data Catalog pour filtrer rapidement les ressources associées à un emplacement spécifique.

Vous devez spécifier un nom et un type de données lorsque vous créez un attribut personnalisé. Vous pouvez choisir un type de données de base, tel que Décimal, Entier, Date et Booléen ou un type de données étendu tel que Utilisateur.

Propriétés générales des attributs

Les propriétés générales des attributs système et personnalisés constituent les propriétés de base, telles que le nom et la description.

Le tableau suivant décrit les propriétés générales des attributs système et des attributs personnalisés :

Propriété	Description
Nom	Nom de l'attribut système ou de l'attribut personnalisé.
Description	Texte descriptif sur l'attribut.
Type de données	Type de données de base ou étendu pour l'attribut. Les exemples sont des types de données de base, tels que Chaîne et Booléen, et étendus, tels que Utilisateur et CSV.
Autoriser la sélection de plusieurs valeurs	Affiche une liste à valeurs multiples pour l'attribut lorsque vous utilisez l'attribut pour la recherche de métadonnées dans Enterprise Data Catalog. Vous pouvez sélectionner simultanément plusieurs valeurs dans la liste. Remarque: Cette propriété ne s'affiche pas pour le type de donnée Booléen ou ne s'y applique pas.

Propriétés de la configuration de recherche

Les propriétés de la configuration de recherche définissent la manière dont Enterprise Data Catalog utilise des attributs dans la recherche de métadonnées.

Le tableau suivant décrit les propriétés de configuration de recherche pour les attributs système et les attributs personnalisés :

Propriété	Description
Rang de recherche	Indique le niveau de classement de la recherche associé à l'attribut. Ce paramètre détermine la position de l'attribut dans les résultats de la requête de recherche d'Enterprise Data Catalog.
Autoriser le filtrage	Détermine si Enterprise Data Catalog peut utiliser l'attribut comme filtre de recherche.
Nom de l'analyseur	Nom de l'analyseur associé aux valeurs de chaîne. Remarque: S'applique uniquement au type de données String.

Création d'un attribut personnalisé

Créez des attributs personnalisés que les utilisateurs d'Enterprise Data Catalog pourront ajouter aux filtres de recherche. Ajouter des filtres de recherche basés sur des attributs personnalisés dans Enterprise Data Catalog afin de classer rapidement les résultats de recherche de métadonnées.

1. Cliquez sur **Gérer > Attributs**.
L'espace de travail **Attributs** s'affiche.
2. Dans le menu **Actions**, sélectionnez **Nouveau**.
La boîte de dialogue **Nouvel attribut personnalisé** s'affiche.
3. Entrez le nom et la description de l'attribut personnalisé.
4. Dans la liste **Type de données**, sélectionnez un type de données, tel que Integer, String, Boolean ou Date.
Le type de données détermine le type de valeurs valide pour l'attribut personnalisé.
5. Dans le champ **Rang de recherche**, choisissez le niveau de classement de recherche pour l'attribut personnalisé.
6. Indiquez si vous devez afficher l'attribut personnalisé en tant que filtre de recherche dans Enterprise Data Catalog.
7. Vous pouvez éventuellement choisir le nom de l'analyseur pour un type de données Chaîne. Le nom de l'analyseur détermine la méthode d'analyse que le moteur de recherche d'Enterprise Data Catalog utilise lorsqu'il effectue l'indexation des valeurs de chaîne de l'attribut. Vous pouvez choisir **STRING**, **TEXT_GENERAL**, ou **TEXT_TECHNICAL**.
8. Ensuite, sélectionnez les types d'objet que vous voulez attribuer à l'attribut personnalisé.
9. Cliquez sur **OK** pour enregistrer les modifications.

Vous pouvez également effectuer cette tâche à l'aide de l'API REST. Pour plus d'informations, reportez-vous à la section [Create Custom Attributes](#) du *Guide de référence de l'API REST d'Enterprise Data Catalog*.

Modification d'un attribut personnalisé et d'un attribut système

Vous pouvez apporter des modifications aux propriétés de la configuration de recherche des attributs personnalisés et des attributs système. Vous ne pouvez pas modifier les propriétés restantes, telles que le nom, la description et le type de données.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Gérer > Attributs**.
L'espace de travail **Attributs** s'ouvre.
2. Cliquez sur l'onglet **Attributs personnalisés** ou **Attributs système** selon l'attribut que vous souhaitez modifier.
3. Dans le panneau **Attributs personnalisés** ou **Attributs système**, sélectionnez l'attribut puis cliquez sur **Modifier**.
Les champs de la section **Configuration de recherche** s'affichent dans le mode d'édition.
4. Effectuez les modifications requises dans les propriétés Rang de recherche, Autoriser le filtrage et Nom de l'analyseur.
5. Cliquez sur **Enregistrer**.

Vous pouvez également mettre à jour et supprimer un attribut personnalisé à l'aide des API REST. Pour plus d'informations, reportez-vous aux sections [Update Custom Attribute](#) et [Delete Custom Attribute](#) du *Guide de référence de l'API REST d'Enterprise Data Catalog*.

Attribution d'attributs personnalisés aux classes et aux ressources

Vous pouvez attribuer des attributs personnalisés aux ressources et aux classes existantes dans le modèle de ressource lorsque vous créez ou modifiez des attributs personnalisés. Vous pouvez attribuer des attributs personnalisés à des classes et des ressources spécifiques, ou à toutes les classes et ressources du catalogue.

Pour attribuer des attributs personnalisés à des ressources et des classes lorsque vous modifiez un attribut personnalisé existant, procédez comme suit :

1. Dans le menu Catalog Administrator, cliquez sur **Gérer > Attributs**. La page **Attributs personnalisés** s'affiche.
2. Dans le volet **Attributs personnalisés**, sélectionnez l'attribut personnalisé que vous souhaitez associer aux ressources et aux classes.
3. Cliquez sur **Modifier**.
4. Dans la section **Attribution de classe**, sélectionnez l'une des options suivantes :
 - Toutes les classes. Attribue l'attribut personnalisé à toutes les classes.
 - Classes spécifiques. Attribue l'attribut personnalisé à des classes spécifiques. Cliquez sur **Afficher les types d'objet dans une hiérarchie** pour afficher les types d'objet inclus dans les classes.

5. Dans la section Attribution de ressource, sélectionnez l'une des options suivantes :
 - Toutes les ressources. Attribue l'attribut personnalisé à toutes les ressources et ressources de référence.
 - Ressources spécifiques. Attribue l'attribut personnalisé à des ressources et ressources de référence spécifiques. Utilisez l'option **Afficher la ressource sélectionnée** pour basculer entre les ressources sélectionnées et toutes les ressources.
 - Aucune. N'associe l'attribut personnalisé à aucune ressource.
6. Cliquez sur **Enregistrer**.

CHAPITRE 8

Attribution de connexions

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'attribution de connexions, 102](#)
- [Connexions attribuées automatiquement, 102](#)
- [Connexions attribuées par l'utilisateur, 103](#)
- [Gestion des connexions, 103](#)
- [Rapport des liens manquants , 104](#)
- [Colonnes du rapport des liens manquants , 105](#)
- [Génération du rapport des liens manquants , 105](#)

Présentation de l'attribution de connexions

Lorsque vous exécutez une analyse des ressources de certains types de ressources, vous devez vous assurer que la connexion source correspond de manière précise aux schémas de la ressource. Enterprise Data Catalog peut détecter automatiquement la façon dont les schémas de base de données sont attribués aux ressources après l'exécution d'une analyse des ressources. Vous pouvez attribuer des schémas de ressource aux connexions ou annuler l'attribution, en fonction de vos besoins. Les tâches de gestion de connexion que vous effectuez dans Catalog Administrator s'appliquent uniquement aux types de ressources SAP Business Objects, plate-forme Informatica, PowerCenter, IBM Cognos, Informatica Cloud Services et Cloudera Navigator.

Si les informations de lignage de ressource de données ne sont pas exactes dans Enterprise Data Catalog, vous pouvez dépanner les connexions attribuées et non attribuées et effectuer les corrections requises dans Catalog Administrator. Vous pouvez ensuite vérifier si le flux de lignage est exact dans Enterprise Data Catalog. Lorsque vous créez des ressources dans Catalog Administrator, vous pouvez choisir d'attribuer automatiquement des schémas de base de données aux ressources. Vous pouvez également attribuer manuellement les schémas à des connexions spécifiques.

Connexions attribuées automatiquement

Lorsque vous créez une ressource pour Informatica Platform, SAP Business Objects, IBM Cognos, les services Informatica Cloud, Cloudera Navigator et une source PowerCenter, vous pouvez choisir de lui attribuer automatiquement les schémas de base de données. Vous pouvez afficher la liste des schémas

attribués automatiquement et leurs connexions pour chaque ressource. Vous pouvez attribuer des schémas dans les connexions attribuées automatiquement ou y annuler l'attribution.

Remarque: Lorsque vous réanalysez une ressource avec l'option d'attribution automatique désactivée, les connexions qui ont été attribuées automatiquement lors de la première analyse sont conservées dans l'onglet **Connexions attribuées automatiquement**. Toutes les autres connexions sont transférées vers l'onglet **Connexions attribuées par l'utilisateur**. Pour plus d'informations sur les connexions attribuées par l'utilisateur, consultez la section ["Connexions attribuées par l'utilisateur" à la page 103](#).

Connexions attribuées par l'utilisateur

Après avoir créé et analysé une ressource pour une plate-forme Informatica, SAP Business Objects, IBM Cognos, les services Informatica Cloud, Cloudera Navigator et une connexion source PowerCenter, vous pouvez afficher la ressource en tant que connexion attribuée par l'utilisateur dans Catalog Administrator. Vous pouvez attribuer manuellement des schémas aux ressources ou annuler des attributions en fonction de vos besoins.

Lorsque vous attribuez manuellement des connexions ou annulez des attributions, le statut de la connexion change en **En cours**. Vous pouvez actualiser l'espace de travail **Attribution de connexion** pour afficher le statut le plus récent.

Remarque: Lorsqu'une connexion pour une ressource PowerCenter est extraite en tant que `<connectionname>.<schemaname>` dans l'espace de travail **Attribution de connexion**, vous devez attribuer une base de données à la connexion.

Gestion des connexions

Vous pouvez attribuer ou annuler l'attribution de connexions une à la fois ou sélectionner plusieurs connexions pour effectuer les modifications.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Gérer > Attribution de connexion**.
L'espace de travail **Attribution de connexion** s'ouvre. L'onglet **Connexions attribuées par l'utilisateur** s'affiche.
2. Utilisez les filtres en haut de la page pour afficher les connexions requises en fonction du type de ressource et du type d'attribution.
3. Pour attribuer un schéma à une ressource, sélectionnez la connexion et cliquez sur **Attribuer** dans le menu système.
La boîte de dialogue **Attribution de connexion** s'affiche.
4. Sélectionnez le schéma à attribuer, puis cliquez sur **Sélectionner**.
Le statut du type d'attribution prend la valeur **En cours**.
5. Dans le menu système en haut de la page, cliquez sur **Actualiser** pour afficher le statut le plus récent dans la colonne Type d'attribution. Vous pouvez cliquer sur **Actualiser** dans le menu système pour afficher les dernières connexions attribuées par l'utilisateur et automatiquement.
6. Pour annuler l'attribution d'une connexion, sélectionnez la connexion, puis cliquez sur **Annuler l'attribution**.
La boîte de dialogue **Annuler l'attribution de la connexion** s'affiche.

7. Cliquez sur **OK**.
8. Pour réattribuer une connexion à un autre schéma, sélectionnez la connexion, puis cliquez sur **Réattribuer**.
9. Pour attribuer ou annuler l'attribution de plusieurs connexions, sélectionnez les connexions, puis sélectionnez **Gérer plusieurs connexions** dans le menu système en haut de la page.
La boîte de dialogue **Connexions attribuées** s'affiche.
10. Apportez les modifications requises aux connexions, puis cliquez sur **OK**.

Attribution de connexions aux ressources de référence

Vous pouvez attribuer des connexions à partir de ressources de référence vers des ressources ou d'autres ressources de référence à l'aide de l'espace de travail **Attribution de connexion**.

Pour attribuer des connexions, suivez les étapes répertoriées dans la section *Gestion des connexions*. Vous pouvez générer un rapport de liens manquants après l'attribution de connexions pour afficher les détails des objets source et cible qui ne sont pas liés. Pour plus d'informations, reportez-vous à la section *Rapport des liens manquants*.

Vous pouvez également attribuer automatiquement des connexions entre la ressource de référence et une ressource à partir de l'onglet **Connexions attribuées automatiquement**. Après l'attribution automatique des connexions, vous pouvez afficher la liste des connexions attribuées pour attribuer manuellement des connexions ou annuler l'attribution.

Lorsque vous effectuez l'attribution de connexions, vous pouvez sélectionner l'option **Conserver les actifs de référence non résolus** de la boîte de dialogue **Attribuer une connexion** pour conserver les actifs de référence non résolus dans le catalogue après l'attribution des connexions. Les actifs de référence non résolus comprennent les tables, les colonnes et les autres actifs qui ne sont pas présents dans la ressource actuelle, mais qui doivent être conservés pour afficher l'intégralité du lignage. Les actifs de référence non résolus s'affichent sous la ressource pour laquelle l'attribution de connexions est effectuée pour la ressource de référence.

Après avoir annulé l'attribution de connexions entre les ressources de référence et les ressources, Enterprise Data Catalog conserve les ressources et les actifs de référence dans le catalogue avec les attributs et les termes d'entreprise personnalisés que vous avez associés aux actifs de référence avant l'attribution de connexions. Une fois l'annulation de l'attribution de connexions effectuée, Enterprise Data Catalog ne conserve pas les attributs ni les termes d'entreprise personnalisés propagés depuis l'actif de référence vers l'actif.

Rapport des liens manquants

Le rapport Liens manquants permet d'identifier les liens de connexion manquants après l'attribution de schémas d'une ressource à des connexions. Les liens manquants peuvent être dus à des attributions de connexion incorrectes et à des ressources contenant des métadonnées obsolètes.

Vous pouvez générer un rapport Liens manquants dans une feuille Excel et afficher les détails d'objets source et cible qui ne sont pas liés. Vous pouvez également identifier les liens de connexion manquants résolus en actifs de référence dans le rapport Liens manquants.

Colonnes du rapport des liens manquants

Le rapport **Liens manquants** contient des détails sur tous les liens manquants après l'attribution de connexions à la ressource. Il inclut des colonnes telles que Connexion de l'objet source, ID de l'objet source, Connexion de l'objet de destination et Connexions de l'objet de destination.

Le tableau suivant décrit les colonnes du rapport **Liens manquants** :

Colonne	Description
Connexion de l'objet source	Nom de la connexion de l'objet source.
ID de l'objet source	ID de l'objet source.
Nom de l'objet source	Nom de l'objet source dans la ressource.
Objet source résolu	Indique si l'objet source est disponible dans le catalogue.
Source de référence	Indique si l'objet source est résolu en un actif de référence.
Connexion de l'objet de destination	Nom de la connexion de l'objet cible.
ID de l'objet de destination	ID de l'objet cible.
Nom de l'objet de destination	Nom de l'objet cible dans la ressource.
Objet de destination résolu	Indique si l'objet cible est disponible dans le catalogue.
Destination de référence	Indique si l'objet cible est résolu en un actif de référence.
Type	Type de flux de données entre les objets source et cible.

Génération du rapport des liens manquants

Générez un rapport des liens manquants pour rechercher les liens de connexion manquants après l'attribution de connexions à la ressource.

1. Ouvrez l'espace de travail **Attribution de connexion**.
2. À partir de l'onglet **Connexions attribuées par l'utilisateur** ou **Connexions attribuées automatiquement**, sélectionnez une ressource pour laquelle vous souhaitez générer le rapport.
 - Vous pouvez éventuellement sélectionner une ressource, puis sélectionner les connexions pour lesquelles vous souhaitez générer le rapport.
3. Cliquez sur **Générer le rapport des liens manquants**.

Catalog Administrator télécharge le rapport sur votre ordinateur.

Remarque: Vous pouvez générer un rapport de liens manquant pour les connexions assignées uniquement.

CHAPITRE 9

Configuration des paramètres réutilisables

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la configuration réutilisable, 106](#)
- [Propriétés générales de configuration, 106](#)
- [Propriétés de la connexion du service d'intégration de données, 107](#)
- [Configuration d'un service d'intégration de données réutilisable, 107](#)

Présentation de la configuration réutilisable

Vous devez configurer les paramètres du service d'intégration de données pour une ressource de façon à extraire les métadonnées de profil des systèmes source. Vous pouvez créer une configuration réutilisable pour les scanners afin d'extraire des métadonnées de profil que vous pouvez réutiliser pour plusieurs ressources.

Une configuration réutilisable vous permet de configurer rapidement plusieurs ressources pour l'extraction des métadonnées de profil. Spécifiez des paramètres, tels que le nom de domaine, le nom du service d'intégration de données et les informations d'identification de l'utilisateur.

Propriétés générales de configuration

Les propriétés générales de la configuration réutilisable incluent le nom, la description et le type de configuration de profilage.

Le tableau suivant décrit les propriétés générales d'une configuration réutilisable :

Propriété	Description
Nom	Nom de la configuration réutilisable pour l'extraction de métadonnées de profil.
Description	Texte descriptif concernant la configuration réutilisable.
Profilage	Indique la configuration du service d'intégration de données pour le profilage.

Propriétés de la connexion du service d'intégration de données

Les propriétés de la connexion du service d'intégration de données incluent les informations de domaine, les informations d'utilisateur de domaine, les informations du service d'intégration de données et les informations du référentiel modèle.

Le tableau suivant décrit les propriétés du service d'intégration de données pour une configuration globale réutilisable :

Propriété	Description
Nom du domaine	Nom du domaine. Le nom ne doit pas dépasser 128 caractères et doit être en ASCII 7 bits. Il ne peut pas contenir d'espaces ni les caractères spéciaux suivants : ` % * + ; " ? , < > \ /
Service d'intégration de données	Nom du service d'intégration de données associé au service de catalogue.
Nom d'utilisateur	Mot de passe pour l'accès au service de référentiel modèle.
Mot de passe	Mot de passe pour l'accès au Service de référentiel modèle.
Domaine de sécurité	Nom du domaine de sécurité auquel appartient l'utilisateur du domaine Informatica.
Hôte	Nom d'hôte du nœud exécutant le service de référentiel modèle.
Port	Numéro de port du nœud exécutant le service de référentiel modèle.

Configuration d'un service d'intégration de données réutilisable

Utilisez le menu **Gérer** pour créer une configuration réutilisable afin d'extraire les métadonnées de profil des systèmes source.

1. Dans l'en-tête de Catalog Administrator, cliquez sur **Gérer > Configuration réutilisable**.
L'espace de travail **Configuration réutilisable** s'ouvre.
2. Dans le menu système, cliquez sur **Nouveau**.
La boîte de dialogue **Nouvelle configuration réutilisable** s'affiche.
3. Entrez les propriétés générales, telles que le nom et la description.
L'option DISOptions est sélectionnée par défaut dans le champ Profilage. Le champ Profilage indique le type de configuration.
4. Dans la section **Paramètres de connexion au domaine**, configurez les informations du domaine, les informations du service d'intégration de données et les informations du service de référentiel modèle.

CHAPITRE 10

Surveillance d'Enterprise Data Catalog

Ce chapitre comprend les rubriques suivantes :

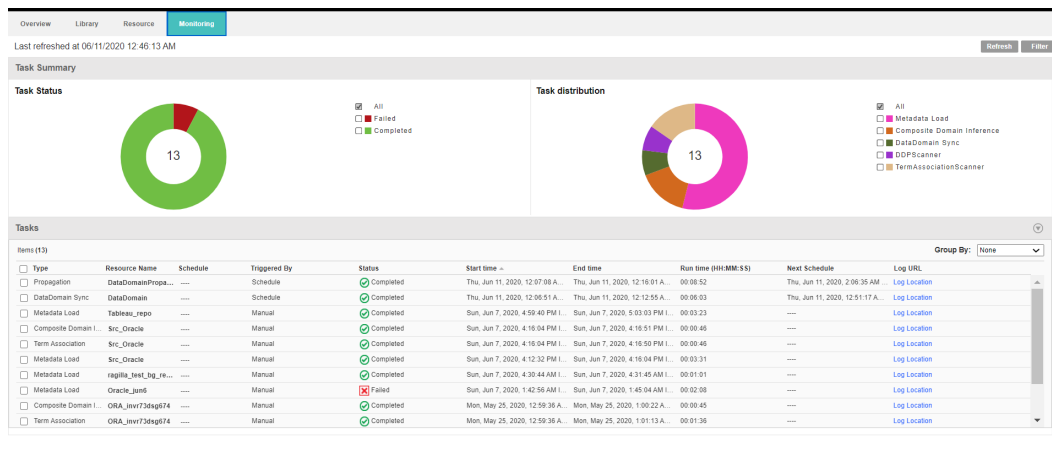
- [Présentation de la surveillance d'Enterprise Data Catalog, 108](#)
- [Statuts de tâches, 109](#)
- [Distribution des tâches, 109](#)
- [Surveillance par ressource, 110](#)
- [Surveillance par tâche, 110](#)
- [Gestion des tâches , 111](#)
- [Application de filtres pour surveiller les tâches, 111](#)

Présentation de la surveillance d'Enterprise Data Catalog

La surveillance d'Enterprise Data Catalog inclut le suivi du statut et de la planification des tâches. Vous pouvez surveiller la durée des tâches en cours d'exécution. Vous pouvez également surveiller la distribution des ressources en fonction de leur nombre pour chaque type de ressource.

L'espace de travail **Présentation** affiche un aperçu des statistiques de surveillance. Vous pouvez afficher le nombre de ressources pour chaque type de ressource, les informations relatives au statut des tâches ainsi que la planification des tâches. Pour procéder à une analyse détaillée des performances d'Enterprise Data Catalog, vous pouvez ouvrir l'espace de travail **Surveillance**. Le statut de la tâche que vous pouvez surveiller inclut le nombre de tâches et leurs statuts, tels que Terminé, Échec et Exécution. Vous pouvez également afficher le nombre de tâches pour chaque phase d'extraction des métadonnées, comme le chargement des métadonnées, l'exécuteur de profils et l'extraction de résultats de profil. Ouvrez les fichiers journaux pour le dépannage des tâches d'Enterprise Data Catalog et un examen approfondi. Vous pouvez également filtrer et regrouper les tâches en fonction de plusieurs facteurs.

L'image suivante montre l'espace de travail Bibliothèque :



Statuts de tâches

Les tâches peuvent avoir des statuts différents en fonction de l'étape du processus d'extraction des métadonnées dans laquelle elles sont effectuées. Le graphique à secteurs des tâches qui se trouve dans l'espace de travail **Surveillance** représente différents statuts de tâches et le nombre de tâches dans chaque statut. Chaque statut de tâche a une couleur différente dans le graphique.

Placez le pointeur sur les différentes sections du graphique à secteurs pour afficher le nombre de tâches pour chaque statut. Cliquez sur n'importe quelle section du graphique à secteurs pour afficher plus d'informations sur la tâche.

Le graphique de statuts de tâches affiche les statuts de tâches suivants :

- Annulé. Nombre de tâches annulées.
- Échec. Nombre de tâches ayant échoué.
- Mis en file d'attente. Nombre de tâches mises en file d'attente pour exécution.
- Exécution. Nombre de tâches en cours d'exécution.
- Terminé. Nombre de tâches qui se sont terminées correctement.

Distribution des tâches

Le graphique à secteurs de la distribution des tâches de l'espace de travail **Surveillance** affiche un résumé des types de tâche et le nombre de tâches pour chaque type. Les types de tâche sont Chargement des métadonnées, Exécuteur de profil et Récupération des résultats du profil. Chaque type de tâche a une couleur différente dans le graphique.

Placez le pointeur sur les différentes sections du graphique à secteurs pour afficher le nombre de tâches pour chaque type. Cliquez sur n'importe quelle section du graphique à secteurs pour afficher plus d'informations sur la tâche. Utilisez les filtres à droite du graphique à secteurs pour filtrer des types de tâches spécifiques.

Surveillance par ressource

Une fois que les scanners ont commencé à extraire des métadonnées des ressources, vous pouvez surveiller les statistiques de ressource dans l'onglet **Surveillance**. Vous pouvez afficher un résumé consolidé des ressources avec différents statuts de tâche.

Les statistiques de surveillance incluent la planification des ressources, le type de tâche, le statut des tâches et les informations relatives à l'exécution suivante planifiée. Vous voudrez peut-être surveiller des ressources spécifiques en fonction de certaines conditions. Par exemple, afficher les ressources qui ont contribué à l'extraction des métadonnées au cours des 12 dernières heures ou les ressources dont les tâches ont échoué dans le cadre des problèmes de dépannage. Utilisez les filtres de l'onglet **Surveillance** pour afficher des statistiques de ressources spécifiques.

Profils de surveillance

Lorsque vous activez la découverte de données et exécutez une ressource, le scanner de profilage exécute la tâche Exécution du profil. Dans l'onglet Surveillance, lorsque vous sélectionnez la tâche Exécution du profil, la section Journaux affiche les statistiques et la progression de l'exécution du profil.

Lorsque vous exécutez une ressource, le scanner de profilage effectue les étapes de validation suivantes :

Validation de l'intégration de données

1. Il vérifie que le service d'intégration de données configuré pour la ressource est activé.
2. Il vérifie que l'entrepôt de profilage a été correctement configuré.
3. Il vérifie que la connexion de l'entrepôt de profilage est correctement configurée.

Validation de profil

1. Il choisit une source de données aléatoire dans la ressource sur laquelle exécuter le profil. Si l'exécution du profil se déroule correctement, le scanner poursuit l'opération sur les autres sources de données.
2. Si l'exécution du profil échoue sur la première source de données, le scanner de profilage choisit une autre source de données aléatoire sur laquelle exécuter le profil. Si l'exécution du profil se déroule correctement, le scanner poursuit l'opération sur les autres sources de données.
3. Si l'exécution de profil échoue sur la première et la deuxième sources de données, le scanner de profilage choisit une autre source de données aléatoire dans la ressource sur laquelle exécuter le profil. Si l'exécution du profil se déroule correctement, le scanner poursuit l'opération sur les autres sources de données.

Si l'exécution du profil échoue pour les trois sources de données, le scanner arrête la tâche Exécution du profil et la propriété Validation du profil affiche la valeur `Échec de la validation` dans les onglets Statistiques et Progression.

Lorsque la validation échoue, cliquez sur **URL du journal** pour afficher le résumé de validation dans le journal de l'application du scanner de profils pour identifier la cause de cet échec.

Surveillance par tâche

Dans le cadre des activités d'administration d'Enterprise Data Catalog, vous pouvez surveiller les tâches de chaque tâche d'analyse. Pour résoudre efficacement les problèmes, vous devez surveiller les tâches de près.

Vous pouvez également vouloir vous concentrer sur les tâches qui ont échoué ou qui sont en cours d'exécution pendant une longue période.

Les statistiques de surveillance incluent le type de tâche, le nom de la ressource, la planification, le statut des tâches ainsi que l'heure de début et de fin. Pour afficher le fichier journal d'une tâche, cliquez sur **Ouvrir le journal** dans la colonne **URL du journal**.

Vous pouvez utiliser le menu de surveillance de chaque tâche pour l'annuler, la suspendre ou la reprendre. Pour annuler, suspendre ou reprendre plusieurs tâches, sélectionnez les tâches et utilisez le menu de surveillance en haut de la grille **Tâches**.

Gestion des tâches

Vous pouvez gérer les tâches répertoriées dans le panneau **Tâches** de l'espace de travail **Surveillance**.

La gestion des tâches inclut les actions suivantes que vous pouvez effectuer en cliquant sur la liste déroulante en regard de la colonne **URL du journal** pour chaque tâche :

Annuler

Sélectionnez cette option pour arrêter une tâche en cours.

Pause

Sélectionnez cette option pour suspendre une tâche en cours.

Reprendre

Sélectionnez cette option pour reprendre une tâche qui a été suspendue.

Réessayer

Sélectionnez cette option pour réessayer une tâche qui a échoué.

Application de filtres pour surveiller les tâches

La grille **Tâches** répertorie toutes les tâches dans le référentiel d'Enterprise Data Catalog. Utilisez le panneau **Filtre** de l'espace de travail Surveillance pour filtrer la liste des tâches. Vous pouvez filtrer les tâches en fonction de plusieurs critères, tels que le nom de la ressource, le type de la ressource et l'heure de création de la ressource. Vous pouvez également afficher les tâches des ressources créées à une période de temps spécifique.

1. Dans l'onglet espace de travail **Surveillance**, cliquez sur **Filtre** dans le coin supérieur droit.
Le panneau **Filtre** apparaît.
2. Pour filtrer la liste des tâches en fonction d'un nom de ressource, entrez le texte requis dans le champ **Nom de ressource**.
3. Pour filtrer les tâches en fonction d'un ou de plusieurs types de ressources, cochez les cases appropriées en regard du champ **Type de ressource**.
4. Pour filtrer les tâches en fonction d'un utilisateur ayant créé une ressource, entrez le nom du client dans le champ **Créé par**.
5. Pour filtrer les tâches en fonction de l'heure à laquelle une ressource a été créée, spécifiez les critères de filtrage appropriés dans le champ **Heure de création de la ressource**.

6. Cliquez sur **Appliquer** pour afficher la liste des tâches filtrées.
7. Pour effacer la condition de filtre, cliquez sur **Effacer**.
8. Pour actualiser la liste des tâches dans la grille **Tâches**, cliquez sur **Actualiser**.

CHAPITRE 11

Gestion des domaines de données

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion des domaines de données, 113](#)
- [Meilleures pratiques pour la création de domaines de données à l'aide de règles, 114](#)
- [Création d'un domaine de données basé sur des règles, 114](#)
- [Création d'un groupe de domaines de données, 117](#)
- [Affichage des domaines de données et des groupes de domaines de données, 117](#)
- [Filtrage de domaines de données, 118](#)
- [Filtrage de groupes de domaines de données, 118](#)
- [Modification d'un domaine de données ou d'un groupe de domaines de données, 118](#)
- [Autorisations et privilèges, 119](#)
- [Découverte de domaines de données sur le moteur Spark, 120](#)
- [Découverte de domaines de données sur le cluster Databricks, 121](#)

Présentation de la gestion des domaines de données

La découverte de domaines de données est le processus de découverte de la signification fonctionnelle des données dans les sources de données en fonction de la sémantique des données. Pour effectuer une découverte de domaines de données, vous devez créer un domaine de données basé sur des règles.

Pour gérer les domaines de données et les groupes de domaines de données, vous pouvez effectuer les tâches suivantes :

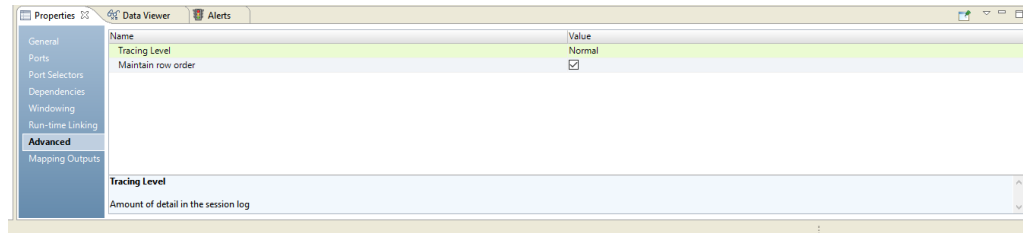
- Créer des domaines de données et des groupes de domaines de données.
- Modifier un domaine de données ou un groupe de domaines de données.
- Ajouter des domaines de données à un groupe de domaines de données.
- Supprimer des domaines de données et des groupes de domaines de données.

Vous devez disposer d'autorisations et de privilèges appropriés pour accéder aux domaines de données et aux groupes de domaines de données dans Enterprise Data Catalog.

Meilleures pratiques pour la création de domaines de données à l'aide de règles

Avant de créer un domaine de données basé sur des règles, vous devez créer des règles dans l'outil Developer tool. Tenez compte des meilleures pratiques suivantes pour les règles que vous définissez dans l'outil Developer tool :

- Si le mapplet contient une transformation Expression, vous devez activer la propriété **Avancé > Maintenir l'ordre des lignes** dans l'outil Developer tool, comme illustré sur l'image suivante :



- La sortie du mapplet doit être de type entier.
- Ne liez pas les ports incompatibles dans le mapplet. Par exemple, ne connectez pas un port de type chaîne à un port de type entier.
- Les transformations d'expressions ne doivent pas contenir plusieurs ports de variable. Au lieu de créer plusieurs ports de variable, vous pouvez utiliser des transformations Java dans le mapplet.
- Vous devez créer un mappage à l'aide de la règle d'expression et exécuter le mappage dans l'environnement d'exécution Spark ou Blaze. Vous pouvez ensuite vérifier les fichiers journaux de mappage avant d'utiliser la règle pour créer un domaine de données dans Catalog Administrator.
- La transformation de sortie doit renvoyer l'entier 1 ou 0 comme sortie au lieu d'une valeur booléenne ou d'une chaîne.

Création d'un domaine de données basé sur des règles






Vous pouvez créer un domaine de données basé sur des règles dans Catalog Administrator. Après avoir créé un domaine de données, vous pouvez découvrir le domaine de données dans les sources de données. Vous pouvez modifier les domaines de données prédéfinis et les domaines de données intelligents dans Catalog Administrator.

1. Cliquez sur **Nouveau > Domaine de données**.
L'assistant **Nouveau domaine de données** s'affiche.
2. Dans le champ **Nom**, entrez un nom pour le domaine de données.
Remarque: Dans le nom du domaine de données, n'incluez aucun espace, nombre ou caractère spécial, à l'exception de l'arobase (@), du signe dièse (#) ou du trait de soulignement (_).
3. Dans le champ **Description**, entrez une brève description du domaine de données.
4. Sélectionnez l'une des options suivantes pour le champ **Spécifier les règles de ce domaine de données** :

- **Oui.** Spécifiez des règles de données, des règles de nom de colonne ou les deux pour le domaine de données.
 - **Non.** Si vous ne spécifiez pas de règles dans le domaine de données, il s'agit d'un domaine de données intelligent. Les domaines de données intelligents sont des domaines de données définis par l'utilisateur qui peuvent être organisés manuellement et propagés à des colonnes similaires.
5. Pour choisir une règle de données pour le domaine de données, cliquez sur **Parcourir**.
 6. Dans la boîte de dialogue **Règle de données**, sélectionnez l'une des règles suivantes :
 - **Sélectionner une table de référence.** Sélectionnez cette option, puis choisissez une table de référence dans la liste des enregistrements. Lorsque vous choisissez une table de référence, les données de la colonne de la table de référence sont utilisées pour associer des domaines de données aux actifs.
 - **Spécifier une expression régulière.** Sélectionnez cette option et entrez une expression régulière pour qu'elle corresponde aux données de chaîne. Une expression régulière est une formule spécialisée pour les chaînes de texte de correspondance qui suivent un modèle. La syntaxe utilisée pour l'expression régulière doit être conforme aux règles de grammaire et aux jetons décrits ici : https://docs.informatica.com/data-catalog/shared-content-for-data-catalog/10-5/transformation-language-reference/functions/reg_extract.html.
 - **Sélectionner une règle.** Sélectionnez l'option, puis choisissez une règle dans la liste des enregistrements. Vous pouvez choisir des règles basées sur des mapplets qui ont été créées à l'aide du client Informatica Analyst ou Informatica Developer. Pour plus d'informations sur l'utilisation de règles basées sur des mapplets, reportez-vous à la section <https://docs.informatica.com/data-engineering/shared-content-for-data-engineering/10-5/developer-mapping-guide/mapplets/mapplet-validation/mapplet-as-a-rule-validation.html>.
 7. Cliquez sur **Sélectionner**.
La règle de données s'affiche dans le champ **Règle de données**.
 8. Pour choisir une règle de nom de colonne pour le domaine de données, cliquez sur **Parcourir**.
 9. Dans la boîte de dialogue **Règle de nom de colonne**, sélectionnez l'une des règles suivantes :
 - **Sélectionner une table de référence.** Sélectionnez cette option, puis choisissez une table de référence dans la liste des enregistrements. Lorsque vous choisissez une table de référence, les données de la colonne de la table de référence sont utilisées pour associer des domaines de données aux actifs.
 - **Spécifier une expression régulière.** Sélectionnez cette option et entrez une expression régulière pour qu'elle corresponde aux données de chaîne. Une expression régulière est une formule spécialisée pour les chaînes de texte de correspondance qui suivent un modèle. La syntaxe utilisée pour l'expression régulière doit être conforme aux règles de grammaire et aux jetons décrits ici : https://docs.informatica.com/data-catalog/shared-content-for-data-catalog/10-5/transformation-language-reference/functions/reg_extract.html.
 - **Sélectionner une règle.** Sélectionnez l'option, puis choisissez une règle dans la liste des enregistrements. Vous pouvez choisir des règles basées sur des mapplets qui ont été créées à l'aide du client Informatica Analyst ou Informatica Developer. Pour plus d'informations sur l'utilisation de règles basées sur des mapplets, reportez-vous à la section <https://docs.informatica.com/data-engineering/shared-content-for-data-engineering/10-5/developer-mapping-guide/mapplets/mapplet-validation/mapplet-as-a-rule-validation.html>.
 10. Cliquez sur **Sélectionner**.
La règle de colonne s'affiche dans le champ **Règle de colonne**.
 11. Dans le champ **Résolution de conflit**, sélectionnez l'une des options suivantes :



- **En cas de conflit de règles, la règle de données remplace la règle de nom de colonne.** Utilisez les résultats de la règle de données pour accepter automatiquement les domaines de données.
- **En cas de conflit de règles, la règle de nom de colonne remplace la règle de données.** Utilisez les résultats de la règle de nom de colonne pour accepter automatiquement les domaines de données.
- **Faire correspondre la règle de données et de nom de colonne.** Les domaines de données sont acceptés automatiquement lorsque les résultats de la règle de données et de la règle de nom de colonne correspondent.
- **Faire correspondre la règle de données et de nom de colonne** Les domaines de données sont acceptés automatiquement lorsque les résultats de la règle de données ou de la règle de nom de colonne correspondent.

Si vous ne sélectionnez pas une option dans le champ **Résolution de conflit**, Enterprise Data Catalog accepte automatiquement les domaines de données selon les résultats de la règle de données.

- Configurez les propriétés de conformité des données suivantes si vous spécifiez des règles pour le domaine de données :
 - **Conformité minimale.** Entrez un pourcentage minimal de conformité pour le domaine de données.
 - **Accepter automatiquement si supérieur à.** Entrez un pourcentage. Enterprise Data Catalog accepte automatiquement le domaine de données si la correspondance de domaine de données dépasse le pourcentage configuré.
 - **Nombre de lignes.** Entrez le nombre de lignes minimales pour la conformité des données.
- Pour ajouter le domaine de données à plusieurs groupes de domaines de données, effectuez les opérations suivantes :
 - Cliquez sur  dans le volet **Groupes de domaines de données**.
La boîte de dialogue **Groupes de domaines de données** s'affiche.
 - Sélectionnez les groupes de domaines de données, puis cliquez sur **Sélectionner**.
La liste des groupes de domaines de données s'affiche dans le volet **Groupes de domaines de données**.
 - Après avoir choisi les groupes de domaines de données pour le domaine de données, vous pouvez également supprimer un groupe de domaines de données. Cliquez sur  pour supprimer un groupe de domaines de données.
- Si vous spécifiez des règles pour le domaine de données, vous pouvez également ajouter plusieurs domaines de données de proximité pour le domaine de données.
Effectuez les étapes suivantes pour ajouter un domaine de données de proximité :
 - Dans la section **Domaines de données de proximité**, cliquez sur .
La boîte de dialogue **Domaines de données** s'affiche.
 - Dans la boîte de dialogue **Domaines de données**, sélectionnez un domaine de données, puis cliquez sur **Sélectionner**.
 - Dans la section **Domaines de données de proximité**, cliquez sur  pour ajouter un autre domaine de données de proximité.
 - Dans la section **Domaines de données de proximité**, cliquez sur  pour supprimer un domaine de données de proximité.
- Cliquez sur **Enregistrer** pour enregistrer le domaine de données.



Création d'un groupe de domaines de données

Vous pouvez créer un groupe de domaines de données et ajouter un ou plusieurs domaines de données au groupe.

1. Cliquez sur **Nouveau > Groupe de domaines de données**.
L'assistant **Nouveau groupe de domaines de données** s'affiche.
2. Dans le champ **Nom**, entrez un nom pour le groupe de domaines de données.
Remarque: Dans le champ **Nom**, n'insérez pas d'espaces ou n'utilisez pas les caractères spéciaux ou les chiffres, à l'exception de l'arobase (@), du dièse (#) ou du trait de soulignement (_) au début du nom du domaine de données dans le nom du groupe de domaines de données.
3. Dans le champ **Description**, entrez une brève description du groupe de domaines de données.
4. Dans le volet **Domaines de données**, cliquez sur  pour ajouter des domaines de données au groupe.
La boîte de dialogue **Domaines de données** s'affiche.
5. Sélectionnez les domaines de données, puis cliquez sur **Sélectionner**.
La liste des domaines de données s'affiche dans le volet **Domaines de données**.
6. Pour supprimer un domaine de données du groupe, sélectionnez le domaine de données dans le volet **Domaines de données**, puis cliquez sur .
7. Cliquez sur **Enregistrer** pour enregistrer le nouveau groupe de domaines de données.

Affichage des domaines de données et des groupes de domaines de données

Vous pouvez afficher la liste des domaines de données et des groupes de domaines de données dans l'espace de travail Bibliothèque. Lorsque vous ouvrez un domaine de données ou un groupe de domaines de données, les propriétés du domaine apparaissent en mode de lecture seule. Lorsque vous affichez un domaine de données, Enterprise Data Catalog affiche les domaines de données composites qui incluent ce domaine de données.

1. L'espace de travail **Bibliothèque** s'affiche.
2. Pour afficher les groupes de domaines de données, cliquez sur **Groupes de domaines de données** dans le volet **Ressources**.
3. Pour afficher les domaines de données, cliquez sur **Domaines de données** dans le volet **Ressources**.
4. Pour afficher un bref résumé du domaine de données ou du groupe de domaines de données, cliquez sur .
5. Pour ouvrir, modifier ou supprimer un groupe de domaines de données ou un domaine de données, sélectionnez le groupe ou le domaine de données, puis cliquez sur le menu **Actions** .
- Les options **Ouvrir**, **Modifier**, **Supprimer** s'affichent.
6. Pour afficher le groupe de domaines de données ou le domaine de données, cliquez sur **Ouvrir**. Vous pouvez également cliquer sur un groupe de domaines de données ou sur un domaine de données pour afficher les détails correspondants.

Filtrage de domaines de données

Vous pouvez filtrer les domaines de données pour ouvrir, modifier ou supprimer des domaines de données spécifiques.

1. Dans la zone d'en-tête, cliquez sur **Bibliothèque**.
Le volet **Actifs** s'affiche dans l'espace de travail **Bibliothèque**.
2. Cliquez sur **Domaines de données** dans le volet **Actifs**.
Les domaines de données s'affichent dans le volet **Domaines de données**. Le filtre **Tout** est sélectionné par défaut dans le volet **Filtrer** et la liste des domaines de données s'affiche dans le volet **Domaines de données**.
3. Pour afficher les domaines de données dans un groupe de domaines de données, sélectionnez un groupe de domaines de données dans la liste **Filtrer > Filtres de domaine**. Cliquez sur **Appliquer**.
Les domaines de données du groupe de données s'affichent dans le volet **Domaines de données**.
4. Pour appliquer un filtre sur les domaines de données, entrez un terme dans le champ **Filtres de domaine**. Cliquez sur **Appliquer**.
Les résultats du filtre s'affichent dans le volet **Domaines de données**.
5. Pour effacer le filtre et les résultats de filtre, cliquez sur **Tout effacer**.

Filtrage de groupes de domaines de données

Vous pouvez filtrer les groupes de domaines de données pour ouvrir, modifier ou supprimer des groupes de domaines de données spécifiques.

1. Dans la zone d'en-tête, cliquez sur **Bibliothèque**
2. Cliquez sur **Groupes de domaines de données** dans le volet **Actifs**.
Les groupes de domaines de données s'affichent dans le volet **Groupes de domaines de données**.
3. Pour appliquer un filtre sur les groupes de domaines de données, entrez un terme dans le champ **Nom du groupe de domaines**. Cliquez sur **Appliquer**.
Les résultats du filtre s'affichent dans le volet **Groupes de domaines de données**.
- 4.

Modification d'un domaine de données ou d'un groupe de domaines de données

Vous pouvez modifier les propriétés d'un domaine de données ou d'un groupe de domaines de données. Vous ne pouvez pas supprimer un domaine de données s'il est inclus dans un domaine de données composite.

1. Dans la zone d'en-tête, cliquez sur **Bibliothèque**.
L'espace de travail **Bibliothèque** s'affiche.

2. Dans le volet **Ressources**, sélectionnez **Domaines de données**.
Les domaines de données s'affichent dans le volet **Domaines de données**.
3. Sélectionnez le domaine de données, puis cliquez sur le menu **Actions** (⌵).
4. Pour modifier les propriétés du domaine de données, cliquez sur **Modifier**.
L'assistant **Domaine de données** s'affiche.
5. Modifiez les propriétés du domaine de données en fonction de vos besoins.
Remarque: Vous ne pouvez pas modifier le nom d'un domaine de données ou d'un groupe de domaines de données.
6. De même, vous pouvez modifier les propriétés du groupe de domaine de données.

Autorisations et privilèges

Vous devez disposer d'autorisations et de privilèges pour accéder aux domaines de données et aux groupes de domaines de données dans Enterprise Data Catalog.

Attribuer des privilèges dans Informatica Administrator pour accéder aux domaines de données

Dans Informatica Administrator, vous devez attribuer des privilèges aux utilisateurs afin qu'ils puissent afficher, modifier ou supprimer des domaines de données et des groupes de domaines de données.

Configurez les privilèges suivants dans l'outil Administrator tool pour accéder aux domaines de données et aux groupes de domaines de données :

Service d'application	Privilège	Description
Service de référentiel modèle	Gérer les domaines de données	L'utilisateur peut créer, modifier et supprimer des domaines de données dans le glossaire de domaine de données. Ce privilège fait partie du groupe de privilèges Administration de domaine de données .
Service de catalogue	Gestion de catalogue : création de domaine	Les utilisateurs peuvent effectuer les actions suivantes : <ul style="list-style-type: none"> - Créer des domaines de données. - Mettre à jour les domaines de données - Afficher les domaines de données - Supprimer les domaines de données
	Gestion de catalogue : conservation d'un domaine	Les utilisateurs peuvent effectuer les actions suivantes : <ul style="list-style-type: none"> - Afficher les domaines de données - Accepter ou rejeter les domaines de données

Service d'application	Privilège	Description
	Gestion des domaines : Admin - ViewDomain et Domaingroup	Les utilisateurs peuvent effectuer les actions suivantes : - Afficher le domaine de données et le groupe de domaines de données
	Gestion des domaines : Admin - EditDomain et Domaingroup	Les utilisateurs peuvent effectuer les actions suivantes : - Créer des domaines de données et des groupes de domaines de données - Mettre à jour un domaine de données et un groupe de domaines de données - Afficher les domaines de données et le groupe de domaines de données - Supprimer un domaine de données et un groupe de domaines de données

Attribuer des autorisations sur des ressources dans Catalog Administrator pour accéder aux domaines de données

Vous pouvez spécifier des autorisations par défaut pour une ou des ressources, un ou des utilisateurs, un ou des groupes d'utilisateurs, ou bien spécifier des ressources, autorisations, utilisateurs et groupes d'utilisateurs selon les besoins.

En tant qu'administrateur du catalogue, vous pouvez attribuer les autorisations suivantes sur des ressources à des utilisateurs et des groupes d'utilisateurs :

Autorisation	Description
Lecture	Permet d'afficher les domaines de données et les groupes de domaines de données dans Enterprise Data Catalog.
Lecture et écriture	Permet d'afficher les ressources des domaines de données et des groupes de domaines de données et d'attribuer des domaines de données aux ressources. Vous pouvez organiser (accepter ou rejeter) les domaines de données.
Toutes les autorisations	Permet d'afficher les domaines de données et les groupes de domaines de données, d'attribuer des domaines de données aux actifs et de créer des domaines de données intelligents. Vous pouvez effectuer l'intégralité des tâches liées aux domaines de données.

Découverte de domaines de données sur le moteur Spark

Lorsque vous exécutez un profil pour effectuer une découverte de données sur le moteur Spark, les tables de référence sont stockées sur le cluster Hadoop. Pour vous assurer que les tables de référence de tous les domaines de données sont stockées sur le cluster, procédez comme suit :

Prérequis :

Vous devez disposer de l'autorisation d'emprunter l'identité de l'utilisateur HDFS lorsque vous effectuez une découverte de domaines de données.

Télécharger les fichiers JAR JDBC

1. Obtenez les fichiers JAR JDBC de la base de données de référence que vous utilisez. Vous pouvez télécharger les fichiers à partir du site Web du fournisseur de la base de données.
2. Copiez les fichiers téléchargés à l'emplacement suivant : `<INFA_HOME>/externaljdbcjars`

Configurer les propriétés personnalisées sur le service d'intégration de données

1. Lancez Informatica Administrator, puis sélectionnez **Service d'intégration de données** dans le **Navigateur de domaine**.
2. Sous l'onglet **Propriétés**, cliquez sur **Propriétés personnalisées**.
3. Définissez les propriétés personnalisées suivantes pour effectuer le stockage intermédiaire des tables de référence pour les domaines de données :

Nom propriété	Valeur propriété
<code>AdvancedProfilingServiceOptions.ProfilingSparkReferenceDataHDFSDir</code>	<code>/tmp/cms</code>
<code>ExecutionContextOptions.SparkRefTableHadoopConnectorArgs</code>	<code>--connect <JDBC thin driver connection URL></code>

4. Assurez-vous que le répertoire `/tmp/cms` existe sur le cluster. Si ce n'est pas le cas, créez le répertoire `/tmp/cms` ou un répertoire personnalisé où vous souhaitez stocker les données. Les données de référence sont stockées dans le répertoire `/tmp/cms` par défaut.
5. Recyclez le service d'intégration de données.
6. Ouvrez Catalog Administrator et assurez-vous que vous exécutez un premier profil avec tous les domaines de données pour stocker les données de référence de manière intermédiaire.

Remarque: Si vous ne sélectionnez pas tous les domaines de données lors de l'exécution du premier profil, puis sélectionnez les domaines de données supplémentaires lors de l'exécution de profil suivante, celle-ci peut échouer.

Découverte de domaines de données sur le cluster Databricks

Utilisez le cluster Databricks pour effectuer la découverte de données sur le moteur Spark. Le cluster Databricks est un environnement d'exécution des travaux Spark. Vous pouvez exécuter un profil pour effectuer la découverte de données pour les sources Azure à l'aide du cluster Databricks.

Vous devez effectuer les étapes suivantes pour vous connecter aux sources Azure dans le cluster Databricks :

Prérequis

Ajoutez les paramètres de configuration Spark avancés suivants pour le cluster Databricks et redémarrez le cluster :

- `fs.azure.account.auth.type OAuth`

- `fs.azure.account.oauth.provider.type`
`org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider`
- `fs.azure.account.oauth2.client.id` `<your-service-client-id>`
- `fs.azure.account.oauth2.client.secret` `<your-service-client-secret-key>`
- `fs.azure.account.oauth2.client.endpoint` `https://login.microsoftonline.com/<directory-ID-of-Azure-AD>/oauth2/token`

Si vous utilisez Azure comme cluster Databricks et exécutez le profil sur une ressource Microsoft Azure Data Lake Store, ajoutez le paramètre de configuration Spark avancé suivant pour le cluster Databricks et redémarrez le cluster : `spark.hadoop.fs.azure.account.key.<<ACCOUNT_NAME>>.dfs.core.windows.net <<VALUE>>`

Télécharger et copier les fichiers JAR pour l'entrepôt de profilage

1. Obtenez les fichiers JAR du pilote JDBC Oracle DataDirect pour l'entrepôt de profilage. Vous pouvez copier les fichiers à partir de l'emplacement suivant : `<INFA_HOME>/services/shared/jars/thirdparty/com.informatica.datadirect-dworacle-6.0.0_F.jar`.
2. Placez les fichiers JAR du pilote JDBC Oracle DataDirect dans les emplacements suivants :
 - `<INFA_HOME>/connectors/thirdparty/informatica.jdbc_v2/spark`
 - `<INFA_HOME>/connectors/thirdparty/informatica.jdbc_v2/common`
 - `<INFA_HOME>/services/shared/hadoop/<DataBricksversion>/runtimeLib`

Télécharger et copier les fichiers JAR pour les objets Delta JBDC

1. Obtenez les fichiers .jar JDBC pour les objets delta JDBC. Vous pouvez télécharger les fichiers sur le site Web du fournisseur de la base de données.
2. Mettez à jour le fichier `genericJDBC.zip` avec les fichiers JAR delta JDBC à l'emplacement suivant : `INFA_HOME/services/CatalogService/ScannerBinaries`.
3. Recyclez le service de catalogue.

Configurer les propriétés personnalisées dans le service d'intégration de données

1. Lancez Informatica Administrator, puis sélectionnez **Service d'intégration de données** dans le **Navigateur de domaine**.
2. Cliquez sur l'option **Propriétés personnalisées** dans l'onglet **Propriétés**.
3. Définissez la propriété personnalisée suivante pour effectuer l'installation automatique des bibliothèques Informatica dans le cluster Databricks :
`ExecutionContextOptions.databricks.enable.infa.libs.autoinstall: true`
4. Recyclez le service d'intégration de données.

Sources prises en charge pour la découverte de domaines de données sur le cluster Databricks

- Delta JDBC
- Azure Data Lake Store Gen2

CHAPITRE 12

Gestion des domaines de données composites

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion des domaines de données composites, 123](#)
- [Création de domaines de données composites, 123](#)
- [Affichage des domaines de données composites existants, 124](#)
- [Filtrage des domaines de données composites, 125](#)
- [Modification des domaines de données composites existants, 125](#)
- [Suppression des domaines de données composites existants, 125](#)

Présentation de la gestion des domaines de données composites

En tant qu'administrateur de catalogue, assurez-vous d'identifier les domaines de données existants, de créer des domaines de données supplémentaires ou des domaines de données composites si nécessaire pour le nouveau domaine de données composites, et de comprendre la combinaison de règles requises pour créer le domaine de données composites. Assurez-vous que les domaines de données ou d'autres domaines de données composites existent avant de créer un domaine de données composite.

La gestion des domaines de données composites implique les tâches suivantes :

- Création de domaines de données composites.
- Modification d'un domaine de données composite existant.
- Suppression de domaines de données composites.

Création de domaines de données composites

Assurez-vous de créer des domaines de données avant de créer des domaines de données composites.


1. Cliquez sur **Nouveau > Domaine de données composites**.



La page **Domaine de données composites** s'affiche.

2. Dans la zone **Nom**, spécifiez le nom du domaine de données composites.


Remarque:

- Assurez-vous de fournir un nom unique pour chaque domaine de données composite.
 - Assurez-vous de ne pas inclure d'espace blanc dans le **Nom** pour le domaine de données composites.
 - Vous ne devez pas commencer le nom du domaine de données composite par des caractères ou des chiffres spéciaux. Vous pouvez inclure les caractères spéciaux suivants dans le nom de domaine de données composite : arobase (@), dièse (#), ou barre de soulignement (_).
3. Dans la zone **Description**, spécifiez la description du domaine de données composite.
 4. Spécifiez les domaines de données que vous souhaitez inclure dans le domaine de données composites

dans la section **Spécifier les critères de correspondance**. Vous pouvez cliquer sur  (**Sélectionner un domaine de données**) pour sélectionner les domaines de données requis dans la boîte de dialogue **Domaines de données**.

- Cliquez sur  (**Ajouter**) pour ajouter des règles qui incluent des domaines de données ou des domaines de données composites à inclure dans la définition de domaine de données composites. Chaque zone de texte dans la section **Spécifier les critères de correspondance** qui apparaît lorsque vous cliquez sur **Ajouter** indique une règle. Enterprise Data Catalog connecte les règles à l'aide de l'opérateur OR.
 - Cliquez sur  (**Supprimer**) pour supprimer les règles de domaine de données de la définition de domaine composite.
5. Cliquez sur **Enregistrer**.

Affichage des domaines de données composites existants

1. L'espace de travail **Bibliothèque** s'ouvre avec l'option **Domaines de données composites** sélectionnée par défaut sous le panneau **Ressources**.
2. Pointez la souris sur le domaine de données composite à afficher dans la liste.
L'option de menu déroulant  s'affiche.
3. Cliquez sur l'option du menu déroulant.
La liste d'actions suivante s'affiche :
 - Ouvrir
 - Modifier
 - Supprimer
4. Cliquez sur **Ouvrir** pour afficher les informations relatives au domaine de données composite sélectionné. Vous pouvez également cliquer sur un domaine de données composite pour en afficher les informations.
Si vous voulez afficher un résumé du domaine de données composite avec le nom et la description, cliquez ► à côté du domaine de données composite.

Filtrage des domaines de données composites

Vous pouvez filtrer les domaines de données composites existants à partir de l'espace de travail Espace de travail **Bibliothèque** pour ouvrir, modifier ou supprimer des domaines de données composites spécifiques.

1. Cliquez sur **Bibliothèque**. L'espace de travail **Bibliothèque** s'ouvre avec l'option **Domaines de données composites** sélectionnée par défaut sous le panneau **Ressources**.
2. Spécifiez le nom du domaine de données composite dans la zone de texte **Nom** de la section **Filtres de domaine de données composite**.
3. Cliquez sur **Appliquer** pour filtrer la liste des domaines de données composites.

Modification des domaines de données composites existants


1. Cliquez sur **Bibliothèque**.
L'espace de travail **Bibliothèque** s'ouvre avec l'option **Domaines de données composites** sélectionnée par défaut sous le panneau **Ressources**.
2. Pointez la souris sur le domaine de données composite à afficher dans la liste.
L'option de menu déroulant (▼) s'affiche.
3. Cliquez sur l'option du menu déroulant.
La liste suivante d'actions que vous pouvez effectuer sur le domaine de données composites s'affiche :
 - Ouvrir
 - Modifier
 - Supprimer
4. Cliquez sur **Modifier** pour modifier les détails du domaine de données composite sélectionné.
5. Mettez à jour les domaines de données et les règles associées au domaine de données composite selon les besoins, puis cliquez sur **Enregistrer**.
Remarque: Vous ne pouvez pas modifier le nom d'un domaine de données composite.

Suppression des domaines de données composites existants

Vous pouvez supprimer les domaines de données composites existants dans l'espace de travail Espace de travail **Bibliothèque**. Lorsque vous supprimez un domaine de données composite, Enterprise Data Catalog supprime ce domaine du référentiel modèle et les résultats de recherche associés du catalogue.

1.
L'espace de travail **Bibliothèque** s'ouvre avec l'option **Domaines de données composites** sélectionnée par défaut sous le panneau **Ressources**.

2. Pointez vers le domaine de données composite à supprimer de la liste des domaines de données composites existants.

L'option de menu déroulant () s'affiche.

3. Cliquez sur l'option du menu déroulant.

La liste suivante d'actions que vous pouvez effectuer sur le domaine de données composites s'affiche :

- Ouvrir
- Modifier
- Supprimer

4. Cliquez sur **Supprimer** pour supprimer le domaine de données composites sélectionné.

L'écran de confirmation de suppression s'affiche.

5. Cliquez sur **Oui** pour confirmer la suppression du domaine de données composites.

CHAPITRE 13

Gérer les définitions de synonymes

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion des définitions de synonymes, 127](#)
- [Chargement de fichiers de définitions de synonymes, 128](#)

Présentation de la gestion des définitions de synonymes

Les définitions de synonymes des ressources du catalogue vous permettent de rechercher des ressources correspondantes à l'aide des synonymes définis.

Vous pouvez définir des synonymes dans un fichier de définitions de synonymes et charger le fichier dans Catalog Administrator. Assurez-vous que le fichier de définitions de synonymes est enregistré en tant que fichier texte avec une extension .txt. Vous devez définir des synonymes séparés par une virgule (,) dans le fichier de définitions de synonymes, comme indiqué dans l'exemple suivant :

social,SSN

security,SSN

uniqueID,SSN

number,SSN

customer,SSN

Remarque:

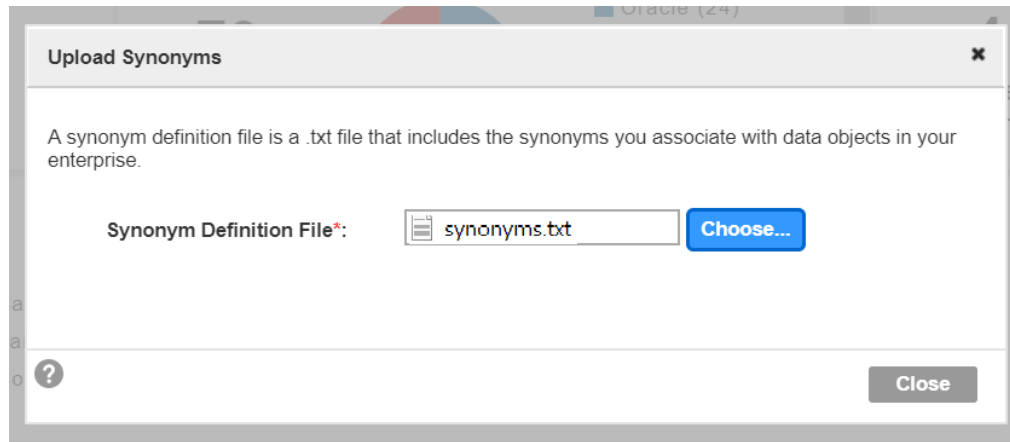
- Assurez-vous de définir les synonymes qui correspondent à chaque ressource dans une ligne distincte du fichier de définition de synonyme.
- Assurez-vous que chaque synonyme que vous définissez est un seul mot sans espaces blancs.
- Vous devez faire précéder un caractère spécial dans un synonyme par une barre oblique inverse (\). Par exemple, si vous voulez définir un synonyme en tant que *credit#number*, vous devez définir le synonyme comme étant *credit\#number*.
- Après avoir chargé le fichier de définition de synonyme, il est inutile d'exécuter les ressources pour rechercher des synonymes dans le catalogue.

Dans l'exemple illustré, les synonymes Social, Sécurité, UniqueID, Nombre et Client correspondent à la ressource nommée SSN dans le catalogue. Vous pouvez rechercher une colonne, une table, un domaine de données ou toute autre ressource nommée SSN à l'aide des synonymes définis.

Chargement de fichiers de définitions de synonymes

1. Cliquez sur **Gérer > Chargement de fichiers de définitions de synonymes**.

La boîte de dialogue **Chargement de synonymes** s'affiche comme illustré dans l'image suivante :



2. Cliquez sur **Choisir**.

La boîte de dialogue **Ouvrir** s'affiche.

Remarque: Vous pouvez également sélectionner un fichier sur votre ordinateur et le placer dans la zone **Fichier de définitions de synonymes** par glisser-déposer.

3. Sélectionnez le fichier de définitions de synonymes que vous avez créé et cliquez sur **Ouvrir**.
4. Cliquez sur **Fermer**.

CHAPITRE 14

Gestion de l'organisation Cloud

Ce chapitre comprend les rubriques suivantes :

- [Présentation de la gestion de l'organisation Cloud , 129](#)
- [Propriétés de l'organisation Cloud , 129](#)

Présentation de la gestion de l'organisation Cloud

Utilisez une organisation pour accéder à Informatica Intelligent Cloud Services depuis Catalog Administrator. Catalog Administrator se connecte à Informatica Intelligent Cloud Services afin d'importer des connexions dans le catalogue. Enterprise Data Catalog utilise les connexions pour prévisualiser les données source et transférer les données de la source vers la cible.

Une organisation est une zone sécurisée du référentiel Informatica Intelligent Cloud Services qui stocke vos licences, comptes utilisateur, ressources d'intégration de données, telles que les mappages et les tâches, ainsi que les informations sur les tâches et la sécurité.

Vous pouvez ajouter, modifier ou supprimer une organisation dans Catalog Administrator. Ajoutez une organisation existante pour configurer le provisionnement de données lorsque vous créez une ressource. Vous pouvez ajouter plusieurs organisations.

Propriétés de l'organisation Cloud

Après avoir ajouté une organisation, Catalog Administrator affiche ses propriétés. Les propriétés de l'organisation sont basées sur celles que vous avez spécifiées pour une organisation dans Informatica Intelligent Cloud Services.

Le tableau suivant décrit les propriétés de l'organisation :

Propriété	Description
Nom de l'organisation	Nom de l'organisation Informatica Intelligent Cloud Service.
ID d'organisation	ID unique attribué à l'organisation lors de sa création dans Informatica Intelligent Cloud Services.
Nom d'utilisateur	Nom d'utilisateur de connexion à Informatica Intelligent Cloud Services.

Ajout d'une organisation Informatica Intelligent Cloud Services

Ajoutez une organisation dans Catalog Administrator pour accéder à Informatica Intelligent Cloud Services.

1. Dans l'en-tête Catalog Administrator, cliquez sur **Gérer > Organisations IICS**.
L'espace de travail **Organisations IICS** s'affiche.
2. Dans le menu Actions, cliquez sur **Nouveau**.
La boîte de dialogue **Ajouter une organisation IICS** s'affiche.
3. Entrez le nom d'utilisateur et le mot de passe que vous avez utilisés pour vous connecter à Informatica Intelligent Cloud Services.
4. Cliquez sur **Valider**.
Catalog Administrator valide les informations d'identification.
5. Cliquez sur **OK**.

Modification de l'organisation Informatica Intelligent Cloud Services

Modifiez les propriétés de l'organisation afin de mettre à jour le mot de passe de l'instance d'Intelligent Cloud Services instance.

1. Dans l'espace de travail **Organisations IICS**, sélectionnez l'organisation pour laquelle vous voulez mettre à jour le mot de passe.
2. Dans la section Propriétés de l'organisation, entrez le nouveau mot de passe de l'organisation.
3. Cliquez sur **Valider**.
Catalog Administrator valide le nouveau mot de passe de l'organisation.
4. Cliquez sur **Enregistrer**.

CHAPITRE 15

Présentation de l'intégration de métadonnées personnalisées

Ce chapitre comprend les rubriques suivantes :

- [Présentation de l'intégration de métadonnées personnalisées, 131](#)
- [Flux de travail d'intégration de métadonnées personnalisées, 132](#)
- [Téléchargement des fichiers de définition de modèle pour la validation, 133](#)
- [Création de modèles personnalisés, 134](#)
- [Mise à jour des modèles personnalisés, 134](#)
- [Exportation de modèles, 135](#)
- [Désapprobation de modèles personnalisés, 135](#)
- [Vue d'ensemble du type de ressource personnalisé, 135](#)
- [Création de types de ressources personnalisés, 136](#)
- [Création de ressources personnalisées, 136](#)
- [Aperçu de l'ingestion des métadonnées, 138](#)
- [Personnaliser les icônes et configurer les vues de relations pour les ressources personnalisées, 143](#)
- [Lignage de contexte pour les ressources personnalisées, 155](#)

Présentation de l'intégration de métadonnées personnalisées

L'intégration de métadonnées personnalisées signifie l'ingestion de métadonnées à partir de sources de données personnalisées dans le catalogue.

Vous pouvez utiliser l'intégration de métadonnées personnalisées pour ingérer des métadonnées à partir de sources pour lesquelles Enterprise Data Catalog ne fournit aucune ressource. En fonction du modèle de la source de données, Enterprise Data Catalog extrait les métadonnées de la source de données. Par défaut, Enterprise Data Catalog fournit des modèles pour plusieurs sources de données à partir desquelles vous pouvez extraire des métadonnées. Ces modèles sont appelés modèles système. Pour ingérer des métadonnées personnalisées dans le catalogue, la première étape consiste à définir un modèle personnalisé.

Les métadonnées personnalisées représentent les métadonnées que vous définissez pour les sources de données personnalisées. Les métadonnées personnalisées incluent des propriétés personnalisées, des relations, des attributs, des classes et des informations de contexte dans la source de données

personnalisées. Pour plus d'informations sur la configuration d'une ressource personnalisée pour extraire des informations de contexte associées au lignage, consultez la section ["Lignage de contexte pour les ressources personnalisées" à la page 155](#).

Vous pouvez définir un modèle personnalisé dans l'administrateur de catalogue. Le modèle définit la structure des métadonnées stockées dans le catalogue. Vous pouvez réutiliser un modèle système et mettre à jour le modèle pour créer et définir un modèle personnalisé. Vous pouvez également créer votre propre modèle.

Après avoir défini le modèle personnalisé, vous devez créer un type de ressource personnalisé dans l'administrateur de catalogue en fonction du modèle personnalisé. Un type de ressource personnalisé peut réutiliser des modèles existants ou vous souhaitez peut-être créer un nouveau modèle. Vous pouvez associer plusieurs modèles, des modèles système et personnalisés à la fois, à un type de ressource personnalisé. En fonction du type de ressource personnalisée, vous pouvez créer une ressource personnalisée dans l'administrateur de catalogue. Vous pouvez configurer les icônes personnalisées afin de représenter les actifs de données provenant d'une ressource personnalisée dans Enterprise Data Catalog. Vous pouvez configurer la page **Vues de relations** dans Enterprise Data Catalog pour les ressources personnalisées. Dans le cadre de la configuration, vous pouvez définir un ensemble de configurations pour les classes dans le modèle personnalisé. Selon les définitions, vous pouvez filtrer ou regrouper des objets associés pour chaque type de classe et afficher les objets sur la page **Vues de relations**.

Vous pouvez configurer une ressource personnalisée afin d'extraire le lignage détaillé ou récapitulatif à partir d'une source ETL.

L'intégration des métadonnées personnalisées implique les étapes avancées suivantes:

1. Créez le modèle personnalisé.
2. Créez le type de ressource personnalisé.
3. Créez la ressource personnalisée.
4. Fournissez les métadonnées que vous extrayez de la source sous la forme d'un fichier ZIP lorsque vous configurez la ressource personnalisée.

Flux de travail d'intégration de métadonnées personnalisées

Le flux de travail d'intégration de métadonnées personnalisé commence par la définition du modèle personnalisé.

Le flux de travail inclut les tâches suivantes :

1. Vérifiez que vous pouvez réutiliser un modèle existant. Par exemple, pour une source de données relationnelle, vous pouvez utiliser n'importe lequel des modèles relationnels inclus dans Enterprise Data Catalog.
2. La création de modèles personnalisés implique les étapes suivantes :
 - Définissez le modèle de métadonnées source dans un fichier XML.
 - Utilisez le fichier de définition de modèle `custom_model.xsd` pour créer et valider le fichier XML du modèle que vous avez créé.
 - Spécifiez les classes, attributs et associations à ingérer dans le catalogue.

- Configurez les options de recherche, de tri et de filtre pour les objets de métadonnées personnalisés ingérés.

Remarque: Les options de recherche, de tri et de filtrage sont des opérations qui exigent d'importantes ressources. Veillez à configurer les options dont vous avez besoin.


3. Enregistrez le modèle personnalisé avec l'administrateur du catalogue.
4. Créez un type de ressource personnalisé.
5. Sélectionnez les modèles que vous souhaitez assigner au modèle de métadonnées personnalisé.
6. Utilisez les fichiers CSV dans le modèle de type de ressource personnalisée afin de remplir les détails de métadonnées et d'associations de la source de données. Vous pouvez également remplir les détails de mappage de la transformation si vous prévoyez d'afficher les détails de la transformation entre plusieurs sources de données du diagramme de lignage.
7. Créez une ressource personnalisée et téléchargez un fichier zip qui inclut les fichiers CSV avec des données lorsque vous configurez la ressource.
8. Exécutez la ressource.

Lorsque la ressource s'exécute, Enterprise Data Catalog traite les métadonnées des fichiers CSV et ingère les métadonnées dans le catalogue.

Téléchargement des fichiers de définition de modèle pour la validation


Enterprise Data Catalog fournit des fichiers de définition de modèle au format XSD que vous pouvez utiliser pour créer et valider les fichiers XML que vous créez.

Pour télécharger les fichiers de définition de modèle, effectuez les étapes suivantes :

1. Cliquez sur **Gérer > Modèles**. La page **Modèles** s'affiche.
2. Cliquez sur  dans le volet **Modèles**.
Les options suivantes s'affichent :
 - **Nouveau modèle personnalisé**
 - **Télécharger le fichier de définition de modèle**
 - **Exporter le canevas de contenu de modèle**
3. Cliquez sur **Télécharger le fichier de définition de modèle**. Le fichier `SchemaDefinitions.zip` est téléchargé sur votre machine.
Le fichier ZIP inclut les fichiers suivants que vous pouvez utiliser pour créer et valider les fichiers XML que vous créez :
 - `custom_model.xsd`. Utilisez ce fichier pour créer et valider le fichier XML du modèle personnalisé que vous créez.
 - `view_configuration.xsd`. Utilisez ce fichier pour créer et valider le fichier `ViewConfiguration.xml` que vous créez pour définir un ensemble de configurations pour les classes du modèle personnalisé.

Création de modèles personnalisés

Vérifiez d'avoir créé et validé le fichier XML de modèle avant d'effectuer les étapes suivantes de création d'un modèle personnalisé :

1. Cliquez sur **Gérer > Modèles**.
La page **Modèles** s'affiche.
2. Cliquez sur  dans le volet **Modèles**.
Le menu s'affiche avec les options suivantes :
 - **Nouveau modèle personnalisé**
 - **Télécharger le fichier de définition de modèle**
 - **Exporter le canevas de contenu de modèle**
3. Cliquez sur **Nouveau modèle personnalisé**.
La boîte de dialogue **Charger le fichier de définition de modèle personnalisé** s'affiche.
4. Cliquez sur **Parcourir** et sélectionnez le fichier XML de modèle personnalisé que vous aviez créé et enregistré.
Le modèle personnalisé est enregistré dans le catalogue.

Mise à jour des modèles personnalisés

Vous pouvez apporter des modifications à un modèle personnalisé en fonction de vos besoins.

- Veillez à incrémenter par un le numéro de version du modèle.
- Vérifiez que vous avez exporté le fichier XML de modèle personnalisé sur votre ordinateur, effectué les modifications requises et enregistré le fichier.
- Consultez les sections suivantes pour en savoir plus sur les propriétés que vous ne devez pas modifier lorsque vous mettez à jour des modèles personnalisés :
 - *Propriétés de l'association*
 - *Propriétés de l'attribut*
 - *Propriétés de la classe*

1. Cliquez sur **Gérer > Modèles**.
La page **Modèles** s'affiche.
2. Sélectionnez le modèle personnalisé que vous souhaitez modifier et cliquez sur **Mettre à jour**.
La boîte de dialogue **Charger le modèle personnalisé** s'affiche.
3. Cliquez sur **Parcourir** pour sélectionner le fichier XML de modèle personnalisé modifié.

Les mises à jour des modèles personnalisés prennent effet au cours de l'analyse suivante par la ressource.

Vous pouvez également effectuer cette tâche à l'aide des API REST. Pour plus d'informations, reportez-vous à la section [Update the Class](#) du *Guide de référence de l'API REST d'Enterprise Data Catalog*.

Exportation de modèles

Vous pouvez exporter un modèle système ou un fichier XML de modèle personnalisé sur votre ordinateur. Vous pouvez ensuite utiliser le fichier XML exporté pour créer un modèle personnalisé.

1. Cliquez sur **Gérer > Modèles**.
La page Modèles s'affiche.
2. Sélectionnez le modèle que vous souhaitez exporter et cliquez sur **Exporter**.
Enterprise Data Catalog télécharge le fichier XML du modèle sur votre ordinateur.

Désapprobation de modèles personnalisés

Vous pouvez désapprouver un modèle personnalisé. Après avoir désapprouvé un modèle personnalisé, vous ne pouvez pas créer de types de ressources personnalisés en fonction du modèle personnalisé déconseillé ou associer le modèle personnalisé déconseillé aux types de ressources personnalisés existants. Lorsque vous désapprouvez un modèle personnalisé, Enterprise Data Catalog place le modèle dans la section **Modèles déconseillés** de la page **Modèles**. Vérifiez que le modèle personnalisé n'est associé à aucun type de ressource personnalisé. Si un type de ressource personnalisé est associé au modèle personnalisé que vous souhaitez désapprouver, supprimez l'association et associez le type de ressource à un autre modèle.

1. Après avoir désapprouvé un modèle personnalisé, vous ne pouvez pas revenir à un modèle que vous pouvez utiliser pour créer des types de ressources personnalisés.
1. Cliquez sur **Gérer > Modèles**.
La page **Modèles** s'affiche.
2. Sélectionnez le modèle personnalisé que vous souhaitez déconseiller et cliquez sur **Déconseiller**.
La boîte de dialogue de confirmation s'affiche.
3. Cliquez sur **OK** pour déconseiller le modèle personnalisé.
Le modèle personnalisé apparaît dans la section **Modèles déconseillés** de la page **Modèles**.
Après avoir déconseillé un modèle personnalisé, Enterprise Data Catalog continue d'analyser et d'ingérer les métadonnées dans le catalogue à l'aide de la ressource personnalisée que vous avez créée en fonction du modèle personnalisé.

Vue d'ensemble du type de ressource personnalisé

Le type de ressource personnalisé représente le type de ressource que vous souhaitez créer en fonction du modèle personnalisé que vous avez défini.

Dans un type de ressource personnalisé, vous définissez les modèles que vous souhaitez associer au type de ressource personnalisé. En outre, vous spécifiez également les associations dans les modèles sélectionnés que vous souhaitez associer au type de ressource personnalisé.

Après avoir créé un type de ressource personnalisé, vous pouvez créer une ressource personnalisée basée sur le type de ressource personnalisé.

Ressources ETL personnalisées


Pour afficher la logique de transformation entre plusieurs sources de données dans le diagramme de lignage, vous pouvez créer et configurer une ressource ETL personnalisée.

Vous pouvez configurer la ressource personnalisée pour afficher le lignage détaillé ou récapitulatif. Par exemple, vos données provenant d'une table Oracle sont transformées dans Informatica PowerCenter avant d'être chargées dans une autre base de données Oracle. Pour ce scénario, vous pouvez configurer une ressource ETL PowerCenter personnalisée afin de capturer les détails de transformation des données dans le diagramme de lignage.

Vous pouvez utiliser le modèle système, **com.infa.idm.etlcore** dans Enterprise Data Catalog pour créer un modèle personnalisé de la ressource ETL.

Création de types de ressources personnalisés

Suivez les instructions fournies ci-dessous pour créer un type de ressource personnalisé basé sur le modèle personnalisé que vous avez créé :

1. Sélectionner **Gérer > Types de ressources personnalisés**.
La boîte de dialogue **Types de ressources personnalisés** s'affiche.
2. Cliquez sur  (Ajouter).
La boîte de dialogue **Nouveau type de ressource personnalisé** s'affiche.
3. Fournissez les détails suivants pour le nouveau type de ressource personnalisé :
 - Nom. Vérifiez que le nom commence par une lettre. Le nom peut inclure des lettres, des chiffres, un trait de soulignement (_), un tiret (-), un point (.) et un espace blanc.
 - Description.
 - Modèle. Cliquez sur **Parcourir**. La boîte de dialogue **Sélectionner les modèles** s'affiche. Sélectionnez le modèle personnalisé que vous avez créé et cliquez sur **OK**.
Remarque: Vous pouvez sélectionner plusieurs modèles système ou personnalisés que vous souhaitez associer au type de ressource personnalisé.
 - Types de connexion Cliquez sur **Sélectionner**. La boîte de dialogue **Sélectionner les types de connexion** s'affiche. Sélectionnez les classes que vous pouvez attribuer en tant que points de terminaison de connexion dans le type de ressource personnalisé et cliquez sur **OK**.
4. Cliquez sur **OK**.
Le type de ressource personnalisé est créé en fonction du modèle personnalisé que vous avez défini.

Création de ressources personnalisées

Suivez les instructions fournies ci-dessous pour créer une ressource personnalisée :

1. Cliquez sur **Nouveau > Ressource**.
L'assistant **Nouvelle ressource** s'affiche.
2. Entrez un nom et une description facultative pour la ressource.

3. Cliquez sur **Fichier > Sélectionner** pour ouvrir la boîte de dialogue **Sélectionner le type de ressource**.
4. Sélectionnez la ressource personnalisée dans la section **Ressource personnalisée**, puis cliquez sur **Sélectionner**.

D'autres champs apparaissent dans l'assistant en fonction du type de ressource que vous avez créé.

5. Configurez les options suivantes sous la section **Propriétés de la connexion** :

- a. **Ressource ETL**. Si vous configurez une ressource ETL pour afficher le lignage de transformation, sélectionnez **Oui**.

Veillez à inclure le fichier `lineage.csv` avec les détails du lignage pour la source de données personnalisée dans le fichier ZIP. Reportez-vous à ["Saisie des détails de la transformation pour les ressources ETL personnalisées" à la page 141](#) pour plus d'informations.

- b. **Attribuer automatiquement des connexions**. Cette option s'affiche si vous aviez sélectionné **Oui** pour l'option **Ressource ETL**. Pour attribuer automatiquement des connexions à la ressource, sélectionnez cette option.

- c. **Exécuter le script**. Si vous avez automatisé le processus pour collecter les métadonnées à partir des sources de données et générer le fichier ZIP à l'aide d'un script ou d'une séquence de commandes, choisissez **Oui**.

Fournissez les détails suivants si vous avez sélectionné **Oui** pour l'option **Exécuter le script** :

- Répertoire de script. Chemin absolu du répertoire à partir duquel vous voulez exécuter les commandes.
- Script ou commandes. Spécifiez le chemin d'accès complet au script à exécuter avec tous les arguments. Vous pouvez également spécifier une séquence de commandes séparées par des points virgules (;) pour générer le fichier ZIP.
- Fichier de sortie. Chemin complet et nom du fichier ZIP à générer par le script.

- d. **Fichiers présents au chemin distant**. Pour sélectionner le fichier ZIP qui inclut les métadonnées personnalisées provenant d'un emplacement de fichier distant, choisissez **Oui**. Spécifiez le chemin absolu du fichier dans la zone de texte **Chemin de fichier distant**.

- e. **Fichier**. Cette option n'est pas applicable si vous avez sélectionné **Oui** pour l'option **Fichiers présents au chemin distant**. Cliquez sur **Choisir** pour sélectionner le fichier ZIP que vous avez créé pour la source de données personnalisée. Vous pouvez également placer le fichier dans la zone **Fichier** à l'aide de la procédure glisser-déposer.

Vérifiez les prérequis suivants pour le fichier ZIP :

- Le fichier se trouve sur un nœud de cluster.
Remarque: Le chemin d'accès au fichier doit être le même sur tous les nœuds du cluster.
- Les autorisations d'accès requises sont configurées pour le fichier.
- Le fichier est accessible par la ressource personnalisée.

6. Cliquez sur **Suivant** pour passer à la page **Paramètres de chargement des métadonnées**.
7. Sélectionnez **Activer les métadonnées source** pour extraire les métadonnées du fichier ZIP.
8. Sélectionnez l'une des valeurs de **Mémoire** pour spécifier la mémoire requise pour exécuter la tâche de scanner :
 - Faible
 - Moyenne
 - Élevée

Sélectionnez la mémoire requise en fonction de la taille du jeu de données que vous souhaitez importer. Reportez-vous à l'article de la bibliothèque de guides pratiques *Tuning Enterprise Data Catalog Performance* pour plus d'informations sur les valeurs de mémoire.

Remarque: Enterprise Data Catalog restreint la taille du fichier `lineage.csv` en fonction de la taille de mémoire que vous spécifiez :

- Faible. La taille maximale de fichier est 500 Mo.
- Mémoire moyenne. La taille maximale de fichier est 1 024 Mo.
- Mémoire élevée. La taille maximale de fichier est 2 048 Mo.

Pour remplacer la restriction imposée sur la taille du fichier, utilisez l'option de JVM - `DcustomScannerMaxFileSize`. Si vous indiquez une taille de fichier très élevée, une erreur de manque de mémoire Java risque de se produire.

9. Cliquez sur **Suivant** pour passer à la page **Attributs personnalisés**, et configurez les paramètres d'attribut. Vous pouvez sélectionner les attributs personnalisés à associer à la ressource.
10. Cliquez sur **Suivant** pour accéder à la page **Planification**.
11. Vous pouvez éventuellement sélectionner les planifications pour le chargement des métadonnées source et les métadonnées de profilage. Vous pouvez créer une planification globale si nécessaire.
12. Cliquez sur **Enregistrer**, ou cliquez sur **Enregistrer et exécuter**.

Après avoir créé une ressource personnalisée, veillez à actualiser la recherche d'Enterprise Data Catalog avant de rechercher des métadonnées personnalisées dans le catalogue.

Aperçu de l'ingestion des métadonnées

Pour ingérer des métadonnées personnalisées dans le catalogue, vous devez saisir les données de la source de données personnalisée dans un modèle suivant la structure du modèle. La ressource personnalisée utilise les données fournies dans le modèle pour extraire et ingérer les métadonnées dans le catalogue.

Après avoir créé un type de ressource personnalisé basé sur le modèle personnalisé, vous devez suivre les instructions suivantes pour ingérer les métadonnées dans le catalogue :

1. Exportez le fichier zip du type de ressource personnalisé.
2. Décompressez le fichier zip pour en extraire les fichiers CSV.
3. Remplissez les détails dans les fichiers CSV en fonction de la structure.
4. Comprimez les fichiers CSV mis à jour en un fichier zip.
5. Fournissez le fichier zip lorsque vous configurez la ressource personnalisée.

Exportation du modèle de type de ressource personnalisé

Vous pouvez utiliser le modèle de type de ressource personnalisé pour ajouter les détails de l'objet d'une classe spécifique définie dans le modèle. Enterprise Data Catalog utilise les détails saisis dans le modèle pour ingérer les métadonnées dans le catalogue.

1. Cliquez sur **Gérer > Types de ressources personnalisées**.
La page Types de ressources personnalisés s'affiche.
2. Sélectionnez la ressource personnalisée pour laquelle vous souhaitez ajouter des détails à partir de la section **Types de ressources personnalisés** et cliquez sur **Exporter le modèle**.

Le modèle de type de ressource personnalisé est téléchargé sur votre ordinateur en tant que fichier zip.

3. Extrayez les fichiers suivants du fichier zip sur votre ordinateur et saisissez les détails requis :
 - `objects.csv`. Utilisez ce fichier pour saisir les données basées sur les détails de la classe de la source de données dans le fichier CSV.
 - `links.csv`. Saisissez les détails de l'association des objets dans la source de données dans le fichier CSV.
 - `lineage.csv`. Applicable si vous voulez configurer une ressource ETL personnalisée afin d'afficher les détails de la transformation entre plusieurs sources de données dans le diagramme de lignage. Entrez les détails de mappage de la transformation dans le fichier CSV.
4. Remplacez les fichiers CSV dans le fichier zip par les fichiers CSV mis à jour.

Vous pouvez inclure plusieurs fichiers `objects.csv`, `links.csv` et `lineage.csv` dans le fichier ZIP.

Veillez à nommer les fichiers au format suivant :

- `objects*.csv`
- `links*.csv`
- `lineage*.csv`

Remarque: * indique les noms de fichiers personnalisés à définir après le nom obligatoire en cas de chargement de plusieurs fichiers. Par exemple, plusieurs fichiers `objects.csv` peuvent se trouver dans le fichier ZIP, tels que `objects1.csv`, `objectscritical.csv`, et `objects12July.csv`.

Saisie des détails de l'association

Pour saisir les détails de l'association pour les objets de la source de données personnalisée, suivez les instructions ci-dessous :

Veillez à ne pas inclure d'espaces blancs dans les en-têtes.

1. Ouvrez le fichier `links.csv` dans n'importe quel éditeur de texte.
2. Saisissez les détails de l'association pour les données dans la source de données personnalisée au format suivant :

`association, from object, to object`

Consultez le tableau suivant pour plus d'informations sur les détails de l'association que vous devez saisir :

Champ	Description
association	<p>Obligatoire. Représente le nom de l'association dans la source de données. Pour saisir des données d'association d'objet pour les objets d'une seule ressource, vérifiez que vous saisissez les entrées <code>association</code> dans le format suivant :</p> <pre>com.example.packageName.associationName</pre> <p>Vous pouvez utiliser le scanneur de lignage personnalisé pour spécifier les détails de l'association avec d'autres types de ressources.</p>
fromObjectIdentity	<p>Obligatoire. Identité unique de l'objet du côté <code>From</code> de l'association. Veillez à saisir le chemin d'accès complet à l'objet.</p> <p>Assurez-vous de spécifier l'identité de l'objet dans un format parent-enfant. Vérifiez que l'identité de l'objet correspond à l'identité de l'objet fourni dans le fichier <code>objects.csv</code>.</p> <p>Par exemple, pour faire référence à une colonne, fournissez l'identité avec le chemin d'accès complet qui inclut le schéma, le tableau et la colonne comme indiqué : <code>Schéma/Table/Colonne</code></p> <p>La saisie du chemin d'accès complet aux objets garantit qu'Enterprise Data Catalog affiche la lignée des objets.</p>
toObjectIdentity	<p>Obligatoire. Identité unique de l'objet du côté <code>To</code> de l'association.</p> <p>Assurez-vous de spécifier l'identité de l'objet dans un format parent-enfant. Vérifiez que l'identité de l'objet correspond à l'identité de l'objet fourni dans le fichier <code>objects.csv</code>.</p> <p>Par exemple, pour faire référence à une colonne, fournissez l'identité avec le chemin d'accès complet qui inclut le schéma, le tableau et la colonne comme indiqué : <code>Schéma/Table/Colonne</code></p> <p>La saisie du chemin d'accès complet aux objets garantit qu'Enterprise Data Catalog affiche la lignée des objets.</p>

Remarque: Veillez à saisir une entrée par ligne.

3. Enregistrez le fichier CSV avec les mises à jour.

Saisie des détails de la classe

Pour saisir les détails de la classe pour les données de la source de données personnalisée, suivez les instructions ci-dessous :

1. Ouvrez le fichier `Fichier objects.csv` dans tout éditeur de texte.
2. Entrez les détails de la classe pour les données de la source de données personnalisée au format suivant :

```
class, identity, core.name, core.description, attributes of the class
```

Pour plus d'informations sur les détails de la classe que vous devez entrer, consultez le tableau suivant :

Champ	Description
class	Obligatoire. Type de classe de l'objet au format suivant: <code>com.example.packageName.ClassName</code>
identity	Obligatoire. Identité unique de la classe avec le chemin d'accès à la classe. Par exemple, pour faire référence à une colonne, fournissez l'identité de l'objet au format parent/enfant: <code>Schema/Table/Column</code>
core.name	Obligatoire. Nom de l'objet.
core.description	Facultatif. Description de l'objet.
com.infa.ldm.etlcore.transformationType	Facultatif. Pour afficher le type de transformation dans le diagramme de lignage, définissez le type de transformation.
attributes of the object	Facultatif. Données pour les attributs associés à l'objet, séparés par des virgules.

Remarque: Veillez à saisir une entrée par ligne.

3. Enregistrez le fichier CSV avec les mises à jour.

Saisie des détails de la transformation pour les ressources ETL personnalisées

Pour entrer les détails de la transformation d'une ressource ETL personnalisée, procédez comme suit :

1. Ouvrez le fichier `lineage.csv` dans tout éditeur de texte.
2. Entrez les détails de la transformation au format suivant :

`association, from connection, to connection, from object, to object`

Pour plus d'informations sur les détails de la classe que vous devez entrer, consultez le tableau suivant :

Champ	Description
Association	Obligatoire. Représente le nom de l'association de flux de données dans le modèle de base. Pour entrer les données d'associations d'objets dans une ressource unique, veillez à entrer les entrées d'associations au format suivant : <code>com.example.packageName.associationName</code> . Remarque: Veillez à ne fournir que les associations de flux de données et pas d'autres types d'association, tels que des associations parent-enfant.
Connexion source	Facultatif. Nom de la source de la transformation.
Connexion cible	Facultatif. Nom de la cible de la transformation.

Champ	Description
Objet source	Obligatoire. Identité unique de l'objet du côté From de l'association. Vérifiez que l'identité de l'objet correspond à l'identité de l'objet fourni dans le fichier <code>objects.csv</code> .
Objet cible	Obligatoire. Identité unique de l'objet du côté to de l'association (transformation appliquée à l'objet dans la colonne Objet source). Vérifiez que l'identité de l'objet correspond à l'identité de l'objet fourni dans le fichier <code>objects.csv</code> . Remarque: Vous pouvez utiliser la variable <code>\$etlRes</code> au début de l'identité.
<code>com.infa.ldm.etl.ETLContext</code>	Facultatif. Utilisez les variables suivantes situées dans la colonne Association du fichier CSV afin d'afficher le lignage détaillé ou récapitulatif des transformations : <ul style="list-style-type: none"> - <code>core.DataSetDataFlow</code>. Lie les objets source et cible au niveau du jeu de données. Pour une source relationnelle, par exemple, le flux de données se trouve au niveau de la table. - <code>core.DirectionaDataFlow</code>. Lie les objets source et cible au niveau de l'élément de données. Pour une source relationnelle, par exemple, le flux de données se trouve au niveau de la colonne. Ajoutez les variables dans la colonne Association en fonction des objets source et cible pour lesquels afficher le lignage détaillé ou récapitulatif. Entrez True dans la colonne <code>com.infa.ldm.etl.ETLContext</code> des variables pour afficher le lignage détaillé des transformations. Ne renseignez pas l'option ou entrez False pour afficher le lignage récapitulatif des transformations. Dans l'exemple suivant, le lignage détaillé est activé pour les tables <code>table source SUPPLIER</code> et <code>table cible SALES</code> : Exemple 1 <pre>core.DataSetDataFlow,Oracle_source,Oracle_wembley,SUPPLIER,SAL ES,TRUE</pre> Dans l'exemple suivant, le lignage détaillé est activé pour les colonnes source et cible <code>ITEMSID</code> et <code>SALESID</code> dans les tables source et cible : Exemple 2 <pre>core.DirectionaDataFlow,Oracle_source,Oracle_wembley,SUPPLIER /ITEMSID,SALES/SALESID,TRUE</pre>

Pour les champs **From Object** et **To Object**, vous pouvez utiliser la variable `$etlRes://` au début de l'identité pour représenter la ressource ETL personnalisée que vous avez créée. Enterprise Data Catalog remplace `$etlRes` par le nom de la ressource ETL personnalisée afin de former le chemin complet de l'identité de l'objet. Par exemple, vous pouvez spécifier l'identité par `$etlRes://SUPPLIER/QUANTITY`. Si vous avez créé une ressource ETL personnalisée nommée `test_etl_res`, Enterprise Data Catalog remplace la variable, comme indiqué dans l'exemple suivant : `test_etl_res://SUPPLIER/QUANTITY`.

Remarque: Veillez à saisir une entrée par ligne.

3. Enregistrez le fichier CSV avec les mises à jour.

Personnaliser les icônes et configurer les vues de relations pour les ressources personnalisées

Vous pouvez configurer les icônes personnalisées afin de représenter les actifs de données provenant des ressources personnalisées dans Enterprise Data Catalog. Vous pouvez également configurer la page **Vues de relations** dans Enterprise Data Catalog pour les ressources personnalisées.

Après avoir configuré les icônes personnalisées, elles s'affichent dans Enterprise Data Catalog pour représenter les actifs de données provenant d'une ressource personnalisée. Vous pouvez utiliser les icônes personnalisées afin de différencier les actifs de données des ressources personnalisées dans Enterprise Data Catalog.

Dans le cadre de la configuration de la page **Vues de relations**, vous pouvez définir un ensemble de configurations pour les classes dans le modèle personnalisé.

Les étapes d'inclusion des icônes personnalisées pour la ressource personnalisée et de définition des configurations des classes sont intégrées au processus de création des modèles personnalisés. Les étapes sont facultatives lorsque vous définissez le modèle personnalisé. Vous pouvez configurer les fonctionnalités en fonction de vos besoins.

Flux de travail


Le flux de travail de personnalisation des icônes pour les actifs de données des ressources personnalisées et de configuration de la page **Vues de relations** dans Enterprise Data Catalog pour les ressources personnalisées inclut les étapes suivantes :

1. Exportation du canevas de modèle. Cette étape implique le téléchargement des fichiers suivants que vous pouvez utiliser pour créer le modèle personnalisé, associer les icônes personnalisées et configurer la page **Vues de relations** pour les ressources personnalisées :
 - `model.xml`
 - `ViewConfiguration.xml`
 - `icons.zip`
2. Création du modèle de ressource personnalisée dans le fichier `model.xml`, comme décrit dans la section ["Création d'un fichier XML de modèle personnalisé" à la page 144](#).
3. Facultatif. Placement des icônes personnalisées dans le fichier `icons.zip`, comme décrit dans la section ["Création du fichier ZIP d'icônes" à la page 144](#).
4. Facultatif. Création du fichier `ViewConfiguration.xml` pour définir un ensemble de configurations pour les classes dans le modèle personnalisé, comme décrit dans la section ["Création du fichier XML de configuration de vue" à la page 145](#).
5. Enregistrement du modèle de ressource personnalisée, comme décrit dans la section ["Création du modèle personnalisé avec les icônes personnalisés" à la page 154](#).
6. Préparation des métadonnées à ingérer, comme décrit dans la section ["Aperçu de l'ingestion des métadonnées" à la page 138](#).
7. Création de la ressource personnalisée, comme décrit dans la section ["Création de types de ressources personnalisés" à la page 136](#).

Exportation du canevas de modèle

Pour télécharger le canevas afin de configurer les icônes personnalisés, effectuez les étapes suivantes :

1. Connectez-vous à Catalog Administrator.
2. Cliquez sur **Gérer > Modèles**. La page **Modèles** s'affiche.
- 3.

Cliquez sur  dans le volet **Modèles**.
Les options suivantes s'affichent :

- **Nouveau modèle personnalisé**
- **Télécharger le fichier de définition de modèle**
- **Exporter le canevas de contenu de modèle**

4. Cliquez sur **Exporter le canevas de contenu de modèle**.

Le fichier `ModelContentTemplate.zip` qui inclut les fichiers suivants est téléchargé sur la machine:

- `model.xml`
- `ViewConfiguration.xml`
- `icons.zip`

Création d'un fichier XML de modèle personnalisé

Utilisez le fichier `model.xml` afin de créer un modèle personnalisé qui définit la structure des métadonnées ingérées dans le catalogue par la ressource personnalisée.

Pour plus d'informations sur les modèles personnalisés et leur création, consultez les sections "[Flux de travail d'intégration de métadonnées personnalisées](#)" à la page 132 et "[Création de modèles personnalisés](#)" à la page 134.

Vous pouvez utiliser le fichier `custom_model.xsd` pour créer et valider le fichier `model.xml`.

Création du fichier ZIP d'icônes

Pour créer le fichier `icons.zip` qui inclut les icônes personnalisées, effectuez les étapes suivantes :

1. Décompressez le fichier `icons.zip` sur la machine locale.
2. Renommez le répertoire nommé `default` dans l'emplacement non compressé avec le nom du modèle. .
Si, par exemple, le nom du modèle est **com.infa.access**, renommez le répertoire en `com_infa_access`.
Veillez à remplacer le point (.) dans le nom du package par le trait de soulignement (_) lorsque vous renommez le répertoire.
3. Vérifiez que les fichiers d'icônes sont nommés en fonction des classes dans le modèle. Si le modèle inclut, par exemple, des classes telles que `Tables`, `Vues` et `Rapports`, nommez les icônes `Tables.svg`, `Views.svg` et `Reports.svg`.
Remarque: Veillez à faire correspondre les noms d'icônes et les noms de classes à des fins de sensibilité à la casse.
4. Placez les fichiers d'icônes au format `.svg` dans le répertoire que vous avez renommé.
5. Comprimez le fichier et nommez le fichier `icons.zip`.

Création du fichier XML de configuration de vue

Le fichier `ViewConfiguration.xml` permet de définir un ensemble de configurations pour les classes dans le modèle personnalisé.

Vous pouvez utiliser le fichier `ViewConfiguration.xsd` pour créer et valider le fichier XML. Pour plus d'informations sur le téléchargement du fichier, consultez la section ["Téléchargement des fichiers de définition de modèle pour la validation" à la page 133](#).

Pour identifier et utiliser les balises que vous pouvez utiliser dans le fichier XML, consultez le tableau suivant :

Balise	Description	Exemple
<code><ViewConfigurationCollection></code> <code></ViewConfigurationCollection></code>	Élément racine du fichier <code>viewConfiguration.xml</code> . Incluez toutes les informations de configuration dans la balise <code><viewConfigurationCollection></code> <code></viewConfigurationCollection></code> .	<code><ViewConfigurationCollection> <viewConfiguration</code> <code><viewConfiguration> <classId>com.infa.db.Table</</code> <code><associationConfigs> <entry> <key>core.DataSet</</code> <code><value> <name>Views</name> </value> </entry> </</code> <code><associationConfigs> </viewConfiguration> </</code> <code>viewConfigurations> </ViewConfigurationCollectio</code>
<code><viewConfigurations> </viewConfigurations></code>	Balise parente qui inclut un ensemble de balises <code><viewConfiguration></code> <code></viewConfiguration></code> que vous utilisez pour définir les configurations des classes dans le modèle.	Consultez l'exemple dans la ligne précédente.
<code><viewConfiguration> </viewConfiguration></code>	Définit les configurations requises pour la classe à personnaliser. Comme indiqué dans l'exemple, les configurations sont définies dans les balises spécifiées à l'intérieur des balises <code><viewConfiguration></code> <code></viewConfiguration></code> .	<code><viewConfiguration> <classId>com.infa.db.Table</</code> <code><associationConfigs> <entry> <key>core.DataSet</</code> <code><value> <name>Views</name> </value> </entry> </</code> <code><associationConfigs> </viewConfiguration></code>
<code><classId> </classId></code>	Balise permettant de spécifier la classe pour laquelle vous voulez définir les configurations. Comme indiqué dans l'exemple, <code>com.infa.db.Table</code> est le nom de la classe pour laquelle vous définissez les configurations.	<code><classId>com.infa.db.Table</classId></code>
<code><associationConfigs> </associationConfigs></code>	Balise qui définit une configuration pour chaque association dans laquelle vous voulez inclure la classe. Comme indiqué dans l'exemple, l'association configurée est <code>core.DataSet</code> et la classe incluse est <code>com.infa.db.Table</code>	<code><viewConfiguration> <classId>com.infa.db.Table</</code> <code><associationConfigs> <entry> <key>core.DataSet</</code> <code><value> <name>Views</name> </value> </entry> </</code> <code><associationConfigs> </viewConfiguration></code>
<code><entry> </entry></code>	La balise permettant de définir la configuration d'une association spécifique dans laquelle la classe spécifiée peut être incluse.	<code><entry> <key>core.DataSet</key> <value> <name>Vi</code> <code></value> </entry></code>
<code><key> </key></code>	Balise permettant de définir l'association pour laquelle vous voulez définir la configuration.	<code><key>core.DataSet</key></code>

Balise	Description	Exemple
<value> </value>	Balise permettant de définir les détails de configuration de l'association.	<entry> <key>core.DataSet</key> <value> <name>Vi </value> </entry>
<name> </name>	Balise permettant de créer un groupe en fonction du nom spécifié. Selon le type d'association avec l'objet d'amorce dans la classe, Enterprise Data Catalog affiche tous les objets du nom de groupe sur la page Vues de relations . Le nom spécifié pour le grouè dans l'exemple est <i>Views</i> .	<classId>com.infa.db.Table</classId> <associatio <entry> <key>core.DataSet</key> <value> <name>Vi </value> </entry> </associationConfigs>
<type> </type>	Balise permettant de définir tous les objets à collecter en fonction du type de classe et de l'association configurés. Dans l'exemple, tous les objets de type com.infa.ldm.relational.View associés à la classe infa.ldm.relational.Table via l'association core.DataSetDataFlow sont collectés.	<viewConfiguration><classId>com.infa.ldm.relatio classId<associationConfigs> <entry> <key>core.DataSetDataFlow</key> <value><name>Vie name><type>com.infa.ldm.relational.View</type> < entry></associationConfigs></viewConfiguration>

Balise	Description	Exemple
<code><classTypes> </classTypes></code>	<p>Utilisez la balise pour configurer des associations pour chaque type de classe associé à une autre classe via une association spécifique. Par exemple, un domaine de données peut être associé à des tables et à des vues relationnelles dans le catalogue. La condition requise est de classer des objets de type tables relationnelles et vues relationnelles associées au domaine de données séparément dans le diagramme de relation.</p> <p>Dans l'exemple, la classe qui représente les objets de type domaine de données est <code>com.infa.ldm.profiling.DataDomain</code>. Les objets du type tables sont représentés par le type de classe <code>com.infa.ldm.relational.Table</code>. Les objets de type vues sont représentés par la classe <code>com.infa.ldm.relational.View</code>. L'association entre le domaine de données et les tables et les vues est représentée par l'association <code>com.infa.ldm.profiling.DataDomainTable</code>.</p>	<pre> <viewConfiguration> <classId>com.infa.ldm.profiling.DataDomain</classId> <associationConfigs> <entry> <key>com.infa.ldm.profiling.DataDomainTable</key> <classTypes> <entry> <key>com.infa.ldm.relational.Table</key> <value>Tables</value> </entry> <entry> <key>com.infa.ldm.relational.View</key> <value>Views</value> </entry> </classTypes> </value> </entry> </associationConfigs> </viewConfiguration> </pre>
<code><incomingName> </incomingName></code> <code><outgoingName> </outgoingName></code>	<p>Utilisez les balises <code><incomingName></code> et <code><outgoingName></code> pour classer les sous-ensembles d'objets à partir d'un groupe, selon les types d'association spécifiques avec l'objet d'amorce.</p> <p>Dans l'exemple; les objets avec les types d'association suivants sont associés à la classe <code>com.infa.ldm.relational.Table</code> de l'objet d'amorce :</p> <ul style="list-style-type: none"> - <code>core.LookupRelatedDataSets</code> - <code>core.JoinRelatedDataSets</code> - <code>core.PkFkRelatedDataSets</code> <p>Tous les objets associés à l'objet d'amorce avec les types d'association définis s'affichent dans le nom Tables.</p> <p>Vous pouvez utiliser les balises <code><incomingName></code> et <code><outgoingName></code> pour configurer les noms spécifiques des sous-ensembles d'objets selon les types d'association avec la classe. Les noms des sous-ensembles s'affichent dans le nom de groupe Tables.</p>	<pre> <viewConfiguration> <classId>com.infa.ldm.relational.Table</classId> <associationConfigs> <entry> <key>core.LookupRelatedDataSets</key> <value> <name>Tables</name> <incomingName>Lookup</incomingName> <outgoingName>Lookup</outgoingName> </value> </entry> <key>core.JoinRelatedDataSets</key> <value> <name>Tables</name> <incomingName>Join</incomingName> <outgoingName>Join</outgoingName> </value> </entry> <key>core.PkFkRelatedDataSets</key> <value> <name>Tables</name> <incomingName>Primary Key Tables</incomingName> <outgoingName>Foreign Key Tables</outgoingName> </value> </entry> </associationConfigs> </viewConfiguration> </pre>

Balise	Description	Exemple
	<p>Les noms spécifiques attribués pour classer les sous-ensembles d'objets avec les types d'association spécifiques sont les suivants :</p> <ul style="list-style-type: none"> - core.LookupRelatedDataSets.Lookup - core.JoinRelatedDataSets.Join - core.PkFkRelatedDataSets.Primary Key Tables, Foreign Key Tables <p>Les noms spécifiques s'affichent dans le nom Tables.</p> <p>Remarque: Vous pouvez utiliser la balise <code><incomingName></code> <code></incomingName></code> si les objets de l'association sont du même type. Par exemple, les colonnes d'une table associées à d'autres dans une autre table.</p>	

Exemple de fichier XML de configuration de vue

```
<?xml version="1.0" encoding="UTF-8"?>
<ViewConfigurationCollection>
  <ViewConfigurations>
    <viewConfiguration>
      <classId>com.infa.ldm.relational.Table</classId>
      <associationConfigs>
        <entry>
          <key>core.DataSetDataFlow</key>
          <value>
            <name>Views</name>
          </value>
        </entry>
      </associationConfigs>
    </viewConfiguration>
    <viewConfiguration>
      <classId>com.infa.ldm.relational.Table</classId>
      <associationConfigs>
        <entry>
          <key>core.DataSetDataFlow</key>
          <value>
            <name>Views</name>
            <type>com.infa.ldm.relational.View</name>
          </value>
        </entry>
      </associationConfigs>
    </viewConfiguration>
    <viewConfiguration>
      <classId>com.infa.ldm.profiling.DataDomain</classId>
      <associationConfigs>
        <entry>
          <key>com.infa.ldm.profiling.DataDomainTable</key>
          <value>
            <classTypes>
              <entry>
                <key>com.infa.ldm.relational.Table</key>
                <value>Tables</value>
              </entry>
              <entry>
                <key>com.infa.ldm.relational.View</key>

```

```

        <value>Views</value>
      </entry>
    </classTypes>
  </value>
</entry>
</associationConfigs>
</viewConfiguration>
<viewConfiguration>
  <classId>com.infa.ldm.relational.Table</classId>
  <associationConfigs>
    <entry>
      <key>core.LookupRelatedDataSets</key>
      <value>
        <name>Tables</name>
        <incomingName>Lookup</incomingName>
        <outgoingName>Lookup</outgoingName>
      </value>
    </entry>
    <entry>
      <key>core.JoinRelatedDataSets</key>
      <value>
        <name>Tables</name>
        <incomingName>Join</incomingName>
        <outgoingName>Join</outgoingName>
      </value>
    </entry>
    <entry>
      <key>core.PkFkRelatedDataSets</key>
      <value>
        <name>Tables</name>
        <incomingName>Primary Key Tables</incomingName>
        <outgoingName>Foreign Key Tables</outgoingName>
      </value>
    </entry>
  </associationConfigs>
</viewConfiguration>
</ViewConfigurations>
</ViewConfigurationCollection>

```

Exemples de cas d'utilisation et diagrammes de relations

Pour plus d'informations sur la définition de balises dans le fichier `ViewConfiguration.xml` en fonction de cas d'utilisation spécifiques, consultez les sections suivantes. Chaque section inclut également un diagramme de relation à la fin de la section pour afficher le mode d'affichage des noms de groupes en fonction des définitions.

Utilisation de la balise `<name>`

Dans le scénario de cas d'utilisation, l'objectif est de regrouper tous les objets de type synonyme associés à la table sous le nom de groupe `Custom Synonyms`.

Les détails de la table, des synonymes et de l'association sont fournis dans la liste suivante :

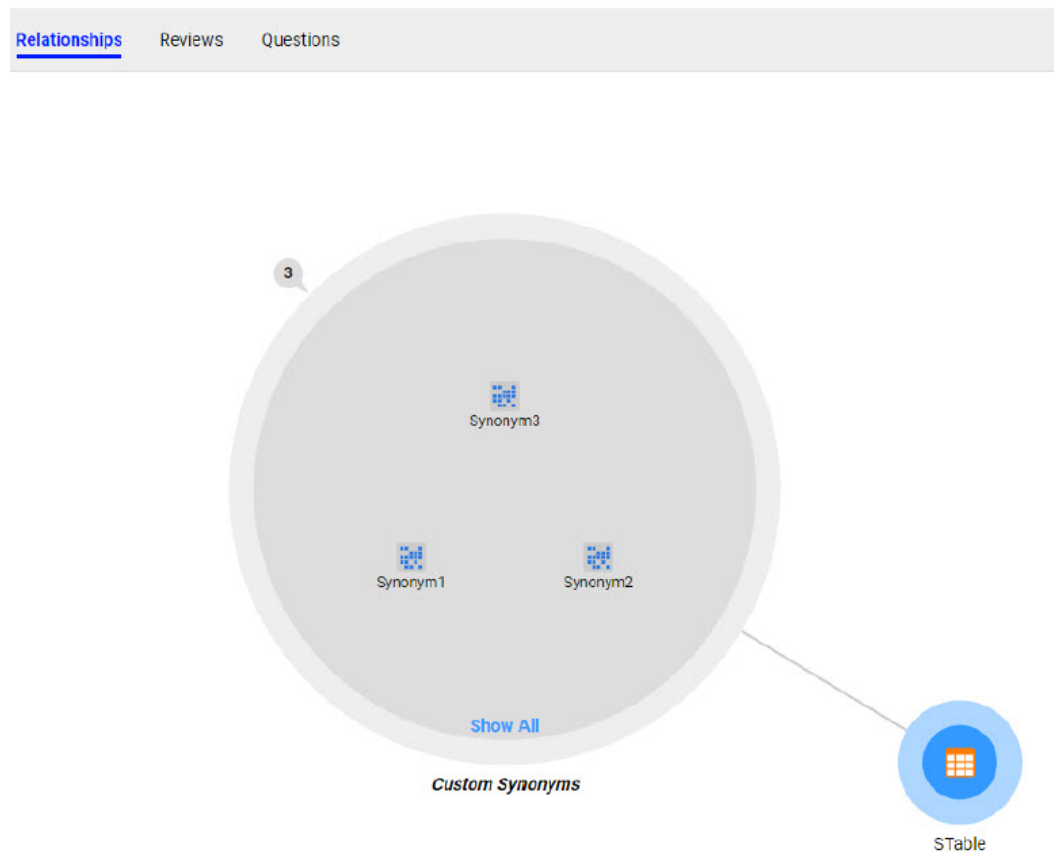
- `STable`, un objet de table provenant d'une base de données relationnelle comprise dans la classe ayant l'ID `org.custom.relational.Table`
- Synonymes nommés `Synonym1`, `Synonym2` et `Synonym3` associés à `STable` via l'association, `core.TableSynonym`.
- Le nom du groupe à définir afin d'inclure tous les objets de type synonymes avec le type d'association spécifique ayant la classe est `Custom Synonyms`.

Pour traiter le cas d'utilisation, vous pouvez ajouter les balises suivantes dans le fichier

ViewConfiguration.xml :

```
<viewConfiguration>
    <classId>org.custom.relational.Table</classId>
    <associationConfigs>
        <entry>
            <key>core.TableSynonym</key>
            <value>
                <name>Custom
Synonyms</name>
            </value>
        </entry>
    </associationConfigs>
</viewConfiguration>
```

Après avoir exécuté la ressource personnalisée et affiché la page d'onglets **Vues de relations** de la ressource de données `STable`, vous pouvez voir tous les synonymes dans le nom de groupe `Custom Synonyms`, comme illustré sur l'image suivante :



Utilisation de la balise `<type>`

Dans le scénario du cas d'utilisation, l'objectif est d'afficher uniquement les types spécifiques d'objets faisant partie d'un lignage avec la table. Celle-ci peut être associée à d'autres tables et vues. Dans l'exemple, la condition requise est de n'afficher que les vues associées à la table sous le nom de groupe `Custom Views`.

Les détails de la table, des types de vue et de l'association sont fournis dans la liste suivante :

- `TableA`, un objet de table provenant d'une base de données relationnelle comprise dans la classe ayant l'ID `org.custom.relational.Table`

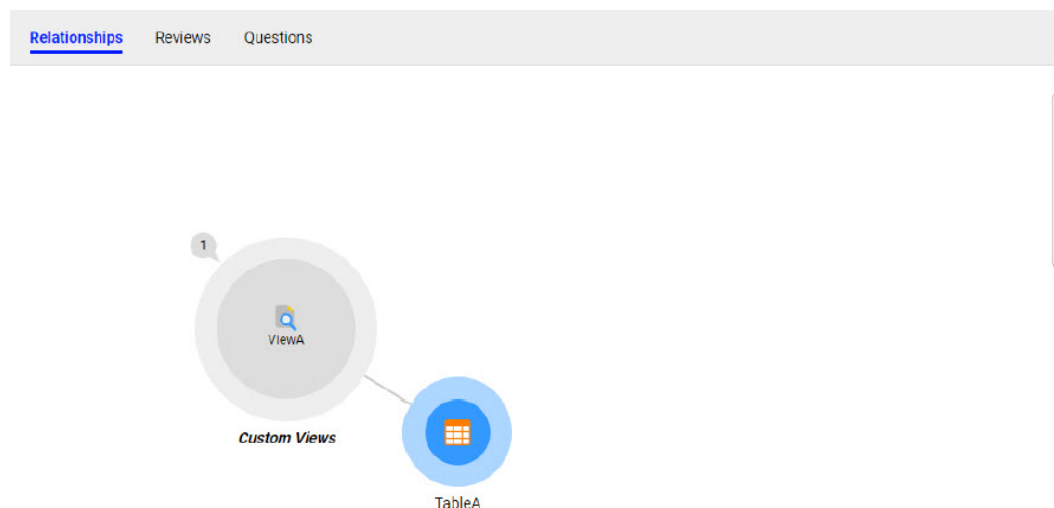
- La vue ViewA est associée à TableA via l'association, core.DataSetDataFlow.
- Nom du groupe à définir afin d'inclure tous les objets d'un type spécifique Custom Views

Pour traiter le cas d'utilisation, vous pouvez ajouter les balises suivantes dans le fichier

ViewConfiguration.xml :

```
<viewConfiguration>
    <classId>org.custom.relational.Table</classId>
    <associationConfigs>
        <entry>
            <key>core.DataSetDataFlow</key>
            <value>
                <name>Custom Views</name>
            </value>
        </entry>
    </associationConfigs>
</viewConfiguration>
```

Après avoir exécuté la ressource personnalisée et affiché la page d'onglets **Vues de relations** de la ressource de données TableA, vous pouvez voir toutes les vues dans le nom de groupe Custom Views, comme illustré sur l'image suivante :



Utilisation de la balise <classTypes>

Dans le scénario de cas d'utilisation, pour une table, des vues et des tables sont associées à la table à l'aide du même type d'association. Nous voulons regrouper toutes les tables faisant partie d'un lignage avec la table sous le nom de groupe Tables et toutes les vues associées à un lignage avec la table sous le nom de groupe Views.

Les détails des ressources de données, de la table spécifique et de l'association sont fournis dans la liste suivante :

- TableM, un objet de table provenant d'une base de données relationnelle comprise dans la classe ayant l'ID org.custom.relational.Table
- La vue ViewO est associée à TableM via l'association, core.DataSetDataFlow.
- La table TableN est associée à TableM via l'association, core.DataSetDataFlow.
- Les noms de groupes qui représentent les tables spécifiques associées à TableM sont Tables et les vues associées à la table sont Views.

Pour traiter le cas d'utilisation, vous pouvez ajouter les balises suivantes dans le fichier

ViewConfiguration.xml :

```
<viewConfiguration>
    <classId>org.custom.relational.Table</classId>
    <associationConfigs>
        <entry>
            <key>core.DataSetDataFlow </key>
            <value>
                <classTypes>
                    <entry>

<key>com.infa.ldm.relational.Table</key>

<value>Tables</value>

</entry>
</entry>

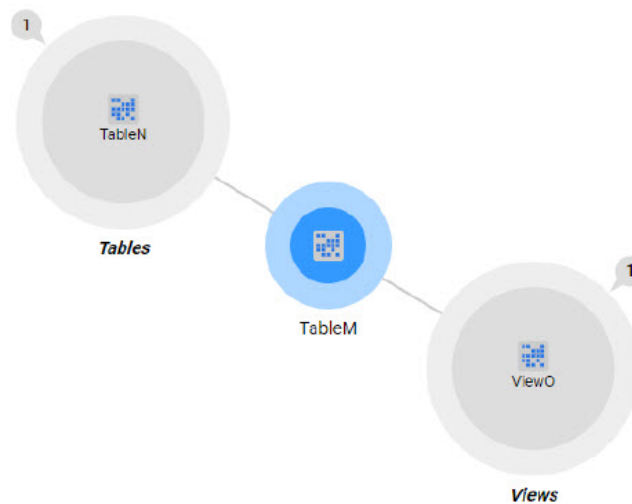
<key>com.infa.ldm.relational.View</key>

<value>Views</value>

</entry>
</classTypes>
</value>
</entry>
</associationConfigs>
</viewConfiguration>
```

Après avoir exécuté la ressource personnalisée et affiché la page d'onglets **Vues de relations** de la ressource de données TableM, vous pouvez voir toutes les tables et vues dans les noms de groupes Tables et Views, comme illustré sur l'image suivante :

Relationships



Utilisation des balises <incomingName> et <outgoingName>

Dans le scénario de cas d'utilisation, une table et un ensemble de tables sont associés à la table avec les associations Jointure, Recherche et clé primaire-clé étrangère. Nous voulons joindre toutes les tables sous le nom de groupe Tables, mais nous voulons regrouper les tables spécifiquement selon l'association, telle que Jointure, Recherche ou clé primaire-clé étrangère. L'exemple répertorie les étapes d'affichage des tables avec l'association clé primaire-clé étrangère.

Les détails des tables et de l'association sont fournis dans la liste suivante :

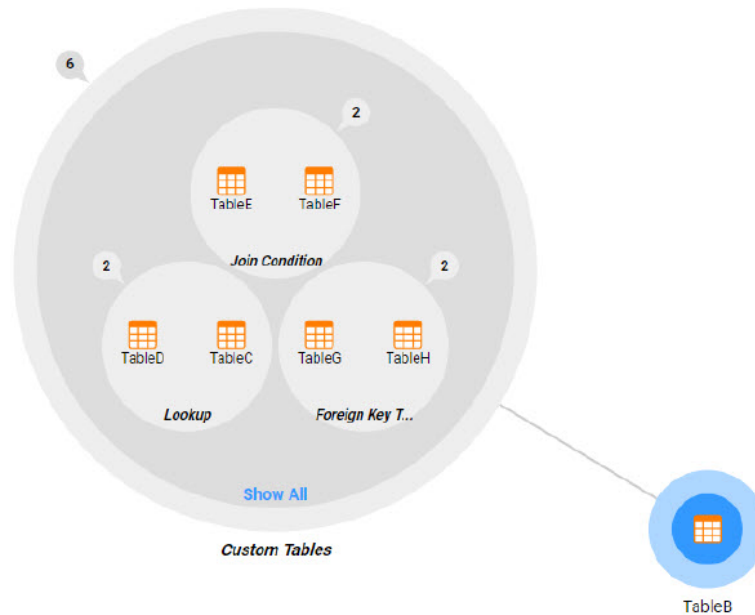
- TableB, un objet de table provenant d'une base de données relationnelle comprise dans la classe ayant l'ID `org.custom.relational.Table`
- L'élément TableB est associé à TableG via l'association clé primaire-clé étrangère, `core.PkFkRelatedDataSets`. TableB contient la clé primaire.
- L'élément TableB est associé à TableH via l'association clé primaire-clé étrangère, `core.PkFkRelatedDataSets`. TableB contient la clé primaire.
- Le nom du groupe qui représente l'association spécifique lorsque TableB est l'objet d'amorce ou l'objet source dans un diagramme de relation est `Foreign Key Tables`.
- Le nom du groupe qui représente l'association spécifique lorsque TableG ou TableH est l'objet d'amorce ou l'objet cible dans un diagramme de relation est `Primary Key Tables`.

Pour traiter le cas d'utilisation, vous pouvez ajouter les balises suivantes dans le fichier

`ViewConfiguration.xml` :


```
<viewConfiguration>
    <classId>org.custom.relational.Table</classId>
    <associationConfigs>
        <entry>
            <key>core.PkFkRelatedDataSets</
key>
            <value>
                <name>Custom Tables</
name>
            <incomingName>Primary Key Tables</incomingName>
            <outgoingName>Foreign Key Tables</outgoingName>
        </value>
        </entry>
    </associationConfigs>
</viewConfiguration>
```

Après avoir exécuté la ressource personnalisée et affiché la page d'onglets **Relations** pour la ressource de données TableB, vous pouvez voir que `Custom Tables` est le groupe qui inclut TableG et TableH sous le sous-groupe `Foreign Key Tables`, comme illustré sur l'image suivante :



Création du modèle personnalisé avec les icônes personnalisés

Après avoir créé les fichiers `model.xml`, `ViewConfiguration.xml` et `icons.zip`, incluez les fichiers dans un fichier ZIP et procédez comme suit pour créer le modèle personnalisé :

1. Cliquez sur **Gérer > Modèles**. La page **Modèles** s'affiche.
2. Cliquez sur  dans le volet **Modèles**.
Le menu s'affiche avec les options suivantes :
 - **Nouveau modèle personnalisé**
 - **Télécharger le fichier de définition de modèle**
 - **Exporter le canevas de contenu de modèle**
3. Cliquez sur **Nouveau modèle personnalisé**. La boîte de dialogue **Charger le fichier de définition de modèle personnalisé** s'affiche.
4. Cliquez sur **Parcourir** et sélectionnez le fichier ZIP qui inclut les fichiers suivants :
 - `model.xml`.
 - `ViewConfiguration.xml`.
 - `icons.zip`.

Le modèle personnalisé est enregistré dans le catalogue. Pour afficher les détails de l'association des ressources personnalisées et les icônes personnalisées, l'enregistrement du modèle personnalisé doit être obligatoirement réussi.

Lignage de contexte pour les ressources personnalisées

Le lignage pour un actif tel qu'un mappage ou une procédure stockée dépend des instances d'exécution. Un contexte est une instance d'exécution d'un actif. L'affichage des détails de ces instances dans le catalogue garantit la précision du lignage et de l'analyse d'impact, et maintient les dépendances appropriées entre les différents actifs du catalogue.

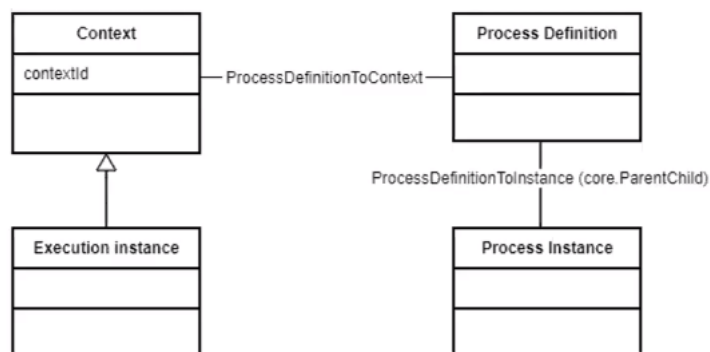
Par exemple, le tableau suivant présente plusieurs jeux de données consommés par la même procédure stockée dans différentes instances d'exécution :

Jeux de données	Mappage	Instance d'exécution	Contexte
Jeu de données A	Procédure stockée 1	Instance d'exécution A	Contexte 1
Jeu de données B	Procédure stockée	Instance d'exécution B	Contexte 2

Le jeu de données A est consommé par la procédure stockée 1 dans l'instance d'exécution A et le jeu de données B est consommé par la procédure stockée 1 dans l'instance d'exécution B. Dans le diagramme Lignage et impact, vous pouvez voir tous les contextes dans lesquels la procédure stockée est exécutée. Dans cet exemple, contexte 1 fait référence à l'instance d'exécution A et contexte 2 à l'instance d'exécution B. Vous pouvez définir le contexte pour chaque instance d'exécution afin que le diagramme Lignage et impact répertorie les contextes pour que vous puissiez identifier les instances.

Vous pouvez utiliser Enterprise Data Catalog pour afficher le contexte d'exécution d'une définition de processus dans le diagramme Lignage et impact à l'aide de ressources personnalisées. Les paramètres qui définissent le flux de données de plusieurs actifs source et cible sont définis en tant que définition de processus. Les types de définition de processus incluent les définitions de procédure stockée ou de mappage. L'exécution d'une définition de processus est définie en tant qu'instance de processus.

L'image suivante présente les liens ou associations qui doivent être inclus dans le modèle personnalisé pour afficher les métadonnées de contexte dans le lignage :



Vous devez utiliser les classes abstraites Définition de processus et Instance de processus pour définir des classes personnalisées, et la classe Contexte pour définir des contextes d'exécution. Pour plus d'informations, consultez la section [“Structure de fichier CSV de ressource de lignage de contexte” à la page 157](#).

Ingestion de métadonnées de contexte à l'aide de ressources personnalisées

Pour ingérer des métadonnées de contexte à l'aide de ressources personnalisées dans Catalog Administrator, procédez comme suit :

1. Dérivez un modèle personnalisé prêt à l'emploi du modèle système `com.infa.ldm.processflow`. Incluez les classes `processDefinition` et `processInstance` du modèle système `com.infa.ldm.processflow` en tant que parent ou superclasses du modèle personnalisé dérivé.
La création de modèle personnalisé inclut les étapes suivantes :

- Définissez le modèle de métadonnées source dans un fichier XML.
- Utilisez le fichier de définition de modèle `custom_model.xsd` pour créer le fichier XML du modèle que vous avez créé.
- Spécifiez les classes, attributs et associations pour lesquels vous souhaitez ingérer les métadonnées de contexte dans le catalogue. Pour plus d'informations sur l'ingestion de métadonnées de contexte, consultez les sections ["Aperçu de l'ingestion des métadonnées" à la page 138](#) et ["Structure de fichier CSV de ressource de lignage de contexte" à la page 157](#).

Pour plus d'informations sur les modèles personnalisés et leur création, consultez les sections ["Flux de travail d'intégration de métadonnées personnalisées" à la page 132](#) et ["Création de modèles personnalisés" à la page 134](#).

2. Enregistrez le modèle personnalisé avec l'administrateur du catalogue. Pour plus d'informations sur l'enregistrement d'un modèle personnalisé, consultez la section ["Création du modèle personnalisé avec les icônes personnalisés" à la page 154](#).
3. Créez un type de ressource personnalisé. Pour plus d'informations, consultez la section ["Création de types de ressources personnalisés" à la page 136](#).
4. Sélectionnez les modèles que vous souhaitez attribuer au modèle de métadonnées personnalisé.
5. Utilisez les fichiers CSV dans le modèle de type de ressource personnalisée afin de remplir les détails de métadonnées et d'associations de la source de données. Vous pouvez également remplir les détails de mappage de la transformation si vous prévoyez d'afficher les détails de la transformation entre plusieurs sources de données du diagramme de lignage.
6. Créez une ressource personnalisée et téléchargez un fichier ZIP qui inclut les fichiers CSV avec des données lorsque vous configurez la ressource. Pour plus d'informations, consultez la section ["Création de ressources personnalisées" à la page 136](#).
7. Exécutez la ressource.

Remarque: Pour afficher la classe `processInstance` dans le modèle personnalisé `processDefinition`, vous devez dériver `processInstance` du `core.DataSet`.

Pour plus d'informations sur l'affichage des informations de contexte dans le diagramme de lignage, consultez le *Guide de l'utilisateur d'Informatica 10.4.1 Enterprise Data Catalog*.

Structure de fichier CSV de ressource de lignage de contexte

Vous pouvez configurer une ressource personnalisée pour afficher les informations de contexte associées à une définition de processus ou une définition de mappage.

Le tableau suivant décrit les éléments qui doivent être inclus dans le fichier `objects.csv` pour afficher les informations de contexte dans le lignage :

Élément	Description
Classes	Incluez les classes suivantes dans le modèle personnalisé : <ul style="list-style-type: none">- com.infa.ldm.processflow.ProcessDefinition. Associe les objets source et cible au niveau du flux de processus. Vous devez définir une instance ou une sous-classe de cette classe dans le modèle personnalisé.- com.infa.ldm.processflow.ProcessInstance. L'exécution d'une définition de processus est appelée une instance de processus. Les liens de flux de processus sont établis au niveau de l'instance de processus.- com.infa.ldm.processflow.Context et com.infa.ldm.processflow.ExecutionContext. Contient tous les objets de contexte d'exécution.
com.infa.ldm.processflow.contextId	Obligatoire. Spécifiez l'ID de contexte pour tous les objets de contexte. La longueur de l'ID de contexte doit être de 32 caractères et unique. Vous pouvez utiliser l'algorithme de hachage MD5 pour générer des ID de contexte uniques. Remarque: Les informations de contexte ne sont pas affichées dans le lignage détaillé si l'ID de contexte n'est pas attribué aux objets de contexte.

Pour plus d'informations sur l'inclusion des classes dans la source de données personnalisées, consultez la section ["Saisie des détails de la classe" à la page 140](#).

Le tableau suivant décrit les associations qui doivent être incluses dans le fichier `links.csv` pour afficher les informations de contexte dans le lignage :

Association	Description
com.infa.ldm.processflow.ProcessDefinitionToInstance	Associe les définitions de processus aux instances de processus.
com.infa.ldm.processflow.ProcessDefinitionToContext	Associe les objets de contexte aux définitions de processus.

Pour plus d'informations sur l'inclusion des associations dans la source de données personnalisées, consultez la section ["Saisie des détails de l'association" à la page 139](#).

Le tableau suivant décrit les éléments qui doivent être inclus dans le fichier `lineage.csv` pour afficher les informations de contexte dans le lignage :

Élément	Description
Association	<p>Incluez les associations de flux de données suivantes :</p> <ul style="list-style-type: none"> - core.DataSetDataFlow. Facultatif. Associe les objets source et cible au niveau du jeu de données. - com.infa.ldm.etl.DetailedDataSetDataFlow. Associe les jeux de données et les flux de processus ou les objets de transformation, et crée les liens de lignage entre les jeux de données et les flux de processus. - com.infa.ldm.etl.DetailedDataFlow. Associe l'élément de données et les flux de processus au niveau de la colonne. <p>Incluez les associations de flux de processus suivantes :</p> <ul style="list-style-type: none"> - com.infa.ldm.processflow.DataSetContextAttributeProjection. Projette les attributs de jeu de données tels que le nom de la ressource source et le nom de la ressource cible en tant qu'informations de contexte dans le lignage. - com.infa.ldm.processflow.DataSetToContext. Associe les objets source aux objets de contexte. - com.infa.ldm.processflow.ContextToDataSet. Associe les objets de contexte aux objets cibles.
core.targetAttribute	<p>Un jeu de données source ou cible crée deux liens de flux de données vers le contexte.</p> <p>Pour un jeu de données source, l'élément <code>core.targetAttribute</code> doit inclure <code>com.infa.ldm.processflow.sourceName</code> et <code>com.infa.ldm.processflow.sourceResourceName</code> en tant que liens de flux de données de contexte.</p> <p>Pour un jeu de données cible, l'élément <code>core.targetAttribute</code> doit inclure <code>com.infa.ldm.processflow.targetName</code> et <code>com.infa.ldm.processflow.targetResourceName</code> en tant que liens de flux de données de contexte.</p> <p>Remarque: Spécifiez <code>core.targetAttribute</code> uniquement pour l'association <code>com.infa.ldm.processflow.DataSetContextAttributeProjection</code>.</p>
	<p>Obligatoire. Spécifiez l'ID de contexte pour tous les liens de flux de données.</p> <p>Remarque: Les informations de contexte ne sont pas affichées dans le lignage détaillé si l'ID de contexte n'est pas attribué aux liens de flux de données.</p> <p>Vous devez inclure l'ID de contexte pour les associations suivantes :</p> <ul style="list-style-type: none"> - <code>com.infa.ldm.etl.DetailedDataFlow</code> - <code>core.DataSetDataFlow</code> - <code>core.DirectionDataFlow</code> - <code>com.infa.ldm.etl.DetailedDataElementDataSetDataFlow</code> - <code>com.infa.ldm.etl.DetailedDataSetDataElementDataFlow</code>

Pour plus d'informations sur l'inclusion des transformations dans la source de données personnalisées, consultez la section ["Saisie des détails de la transformation pour les ressources ETL personnalisées" à la page 141](#).

CHAPITRE 16

Utilitaires de sauvegarde de catalogue et de collecte de journaux

Ce chapitre comprend les rubriques suivantes :

- [Présentation, 159](#)
- [Accès aux utilitaires, 160](#)
- [Exécution d'une sauvegarde de catalogue, 160](#)
- [Utilitaires de collecte de journaux, 161](#)
- [, 162](#)

Présentation

Les utilitaires vous aident à sauvegarder le catalogue et à collecter les fichiers journaux requis pour analyser et dépanner les problèmes que vous pouvez rencontrer lors de l'utilisation d'Enterprise Data Catalog.

Sauvegarde du catalogue

Vous pouvez sauvegarder les données du catalogue si vous souhaitez résoudre les problèmes en fonction des données du catalogue sans arrêter le service de catalogue.

Journaux de service

Les journaux des services d'Enterprise Data Catalog (service de catalogue, service de cluster Informatica, historique EBF et PhoneHome), MongoDB, Apache Solr, service Nomad et travaux d'analyse, Apache ZooKeeper (ZK) et PostgreSQL vous aident à analyser et à résoudre les problèmes liés à l'ingestion de métadonnées dans le catalogue, les problèmes liés à la recherche ou les problèmes liés aux pannes et à la stabilité des services.

Journaux de ressource

Les journaux de ressources vous aident à résoudre les problèmes liés à l'échec de ressources spécifiques.

Accès aux utilitaires

Vous pouvez utiliser les utilitaires de sauvegarde et de collecte de journaux après avoir installé et configuré Enterprise Data Catalog. Veillez à utiliser un compte utilisateur configuré avec les privilèges d'Informatica Administrator pour accéder aux utilitaires.

Pour ce faire, procédez comme suit :

1. Cliquez sur le menu Aide, puis sur **Interface de journal**.
2. Sur la page de connexion, entrez le nom d'utilisateur et le mot de passe. La page **Détails du travail de surveillance** s'affiche par défaut.
3. Cliquez sur **Sauvegarder le catalogue** pour accéder à la sauvegarde du catalogue, ou cliquez sur **Collecter les journaux de service** ou **Collecter les journaux de ressources** pour accéder aux utilitaires de collecte de journaux de service et de ressources.

Exécution d'une sauvegarde de catalogue

Vous pouvez sauvegarder le catalogue sans arrêter le service de catalogue. En outre, vous pouvez toujours utiliser Enterprise Data Catalog lorsque vous effectuez une sauvegarde.

Vous pouvez sauvegarder le catalogue dans un répertoire de votre choix. Vous pouvez accéder à l'utilitaire depuis la page **Paramètres de sauvegarde du catalogue**.

Lorsque vous effectuez une sauvegarde, vous pouvez afficher les données de catalogue et les détails de configuration de catalogue à l'aide des applications Enterprise Data Catalog. Toutefois, vous ne pouvez pas apporter de modifications à la configuration, la mettre à jour ni ajouter de nouvelles ressources ou de nouveaux actifs dans Enterprise Data Catalog. Les tâches limitées lors de la sauvegarde incluent l'exécution de ressources, la création de ressources, l'ajout d'enrichissement aux ressources, la création d'attributions de connexion et l'importation en bloc.

Effectuez les étapes suivantes pour sauvegarder le catalogue :

1. Dans le panneau de gauche, cliquez sur **Sauvegarder le catalogue**. La page **Paramètres de la sauvegarde du catalogue** s'affiche.
2. Configurez les propriétés facultatives suivantes, selon les besoins :
 - a. Dans la zone de texte **Chemin de sauvegarde**, entrez le chemin vers le répertoire dans lequel vous souhaitez enregistrer la sauvegarde du catalogue. Par défaut, la sauvegarde est enregistrée dans `<INFA_HOME>/logs/<NODE_NAME>/services/CatalogService/<SERVICE_NAME>/JobsStore/`.
Remarque: Le contenu du répertoire de sauvegarde par défaut est conservé uniquement pendant 24 heures. En outre, le contenu du répertoire par défaut est perdu lorsque vous recyclez le service de catalogue. Veillez à spécifier un répertoire distinct pour stocker la sauvegarde ou téléchargez celle-ci à partir du répertoire par défaut.
 - b. Sélectionnez le magasin à sauvegarder.
3. Cliquez sur **Télécharger** pour démarrer le travail de sauvegarde. La page **Détails de la tâche de surveillance** s'affiche depuis l'emplacement à partir duquel vous pouvez télécharger la sauvegarde.

Remarque: Lorsque vous effectuez une sauvegarde de catalogue, le statut des analyses de ressources en cours passe à Annulé par le système. L'analyse reprend à la fin du travail d'analyse.

Le fichier de sauvegarde est chiffré à l'aide de l'algorithme Advanced Encryption Standard (AES).

Utilitaires de collecte de journaux

Enterprise Data Catalog fournit des utilitaires que vous pouvez utiliser pour collecter des journaux à partir du service de catalogue, du service de cluster Informatica et de ressources spécifiques.

Vous pouvez collecter des journaux pour analyser et dépanner des problèmes relatifs aux services et aux ressources s'ils ne fonctionnent pas comme attendu. Les utilitaires n'ont pas besoin que vous arrêtiez les services ou les ressources lors de la collecte des journaux.

Après avoir exécuté les utilitaires pour collecter les journaux, assurez-vous de télécharger les journaux dans les 24 heures. Les journaux sont supprimés du répertoire après 24 heures. En outre, si vous recyclez le service de catalogue après avoir exécuté les API REST, les journaux sont supprimés du répertoire.

Collecter les journaux de service Enterprise Data Catalog

Vous pouvez collecter des journaux de service Enterprise Data Catalog pour identifier et dépanner des problèmes relatifs aux services Enterprise Data Catalog. Vous pouvez utiliser la page **Collecter les journaux de service** pour collecter les journaux de service.

Pour collecter les journaux de service, procédez comme suit :

1. Dans le panneau de gauche, cliquez sur **Collecter les journaux de service**. La page **Collecter les journaux de service** s'affiche.
2. Dans **Sélectionner les journaux**, choisissez l'une des options suivantes pour lesquelles vous souhaitez collecter les journaux :

MONGO

Collecte les journaux MongoDB lors de l'ingestion de données.

SOLR

Collecte les journaux Apache Solr pour résoudre les problèmes liés à la recherche d'actifs de données.

NOMAD

Collecte les journaux liés au service d'orchestration Nomad et les travaux d'analyse.

ZK

Collecte les journaux liés au service Apache ZooKeeper.

POSTGRES

Collecte les journaux liés à la base de données PostgreSQL.

SERVICE

Collecte les journaux liés au service de catalogue, au service de cluster Informatica, à l'historique EBF et à PhoneHome.

ALL

Collecte les journaux pour tous les éléments répertoriés ci-dessus.

3. Dans la zone **Date de début**, sélectionnez la date à partir de laquelle vous souhaitez collecter les journaux de service.
4. Dans la zone **Date de fin**, sélectionnez la date jusqu'à laquelle vous souhaitez collecter les journaux de service.
5. Cliquer sur **Télécharger**. La page **Détails de la tâche de surveillance** s'affiche depuis l'emplacement à partir duquel vous pouvez télécharger le fichier journal.

Collecter les journaux de ressource

Vous pouvez collecter des journaux Nomad de ressources spécifiques et des journaux d'intégration de données pour le profilage afin d'identifier et de résoudre les problèmes liés aux ressources. Vous pouvez utiliser la page **Collecter les journaux de ressource** pour collecter les journaux.

Pour collecter les journaux de ressource, procédez comme suit :

1. Dans le panneau de gauche, cliquez sur **Collecter les journaux de ressource**. La page **Collecter les journaux de ressource** s'affiche.
2. Sélectionnez **SCANNER**, **PROFILING**, ou les deux options pour la collecte des journaux.
3. Dans la zone **Date de début**, sélectionnez la date à partir de laquelle vous souhaitez collecter les journaux de ressource.
4. Dans la zone **Ressource**, cliquez sur **Ajouter un élément** pour spécifier les noms des ressources pour lesquelles vous souhaitez collecter les fichiers journaux.
5. Cliquez sur **Télécharger**. La page **Détails de la tâche de surveillance** s'affiche depuis l'emplacement à partir duquel vous pouvez télécharger le fichier journal.

Après avoir téléchargé et extrait les fichiers journaux, ceux-ci s'affichent dans le répertoire <resource name> dans le répertoire des journaux Informatica.

Remarque: Si aucune ressource ne correspond au nom de la ressource que vous avez spécifié, le travail de collecte du journal des ressources échoue. Cliquez sur **Emplacement du journal** pour télécharger le fichier journal.

Vous pouvez surveiller le statut des exécutions de l'utilitaire de sauvegarde de catalogue et de collecte de journaux que vous avez lancé, sur la page **Détails de la tâche de surveillance**. Vous pouvez également télécharger les journaux et la sauvegarde à partir de cette page.

Cliquez sur **Surveillance** dans le panneau de gauche pour lancer la page **Détails de la tâche de surveillance**. La page répertorie le statut des exécutions de l'utilitaire à l'aide des étiquettes suivantes :

- Soumis
- Exécution
- Terminé
- Échec

Vous pouvez cliquer sur **Actualiser** pour actualiser la page **Détails de la tâche de surveillance** avec le statut mis à jour.

Pour télécharger la sauvegarde ou le journal requis, cliquez sur le lien **Télécharger** présent dans la colonne **Résultats** pour la tâche.

Si une tâche échoue, le statut **Échec** s'affiche. Vous pouvez cliquer sur le lien **Emplacement du journal** dans la colonne **URL de journal** pour la tâche afin d'afficher le journal généré pour l'échec de la tâche.

ANNEXE A

Enregistrement de types de données inconnus pour l'exécution de profils

Cette annexe comprend les rubriques suivantes :

- [Présentation, 163](#)
- [Enregistrement d'un type de données inconnu comme type de données pris en charge, 164](#)

Présentation

Lorsque vous activez la découverte de données et exécutez une ressource, le scanner de profilage exclut les colonnes ayant des types de données inconnus, tels que les types de données définis par l'utilisateur, et exécute le profil sur les colonnes source restantes dont les types de données sont pris en charge. Les colonnes exclues n'apparaissent pas dans les résultats de profilage. Pour atténuer ce problème pour les sources relationnelles, vous pouvez enregistrer le type de données inconnu en tant que type de données varchar. Une fois le type de données enregistré, le scanner de profilage l'utilise lors de l'exécution de la ressource.

Vous pouvez mapper les types de données pour les sources de données relationnelles telles qu'Oracle, Microsoft SQL Server, IBM DB2 et Teradata. Vous ne pouvez pas enregistrer les types de données BLOB, CLOB, image et binary dans le fichier XML. Vous ne pouvez pas enregistrer de types de données personnalisés dans la source de données Oracle.

Exemple

Dans la table Client d'une source de données Microsoft SQL Server, vous concaténez les colonnes FirstName et LastName pour créer une nouvelle colonne nommée CustomerFullName. Vous configurez un type de données défini par l'utilisateur nommé user_name pour la colonne CustomerFullName. Vous souhaitez exécuter le profil de colonne sur la table Client. Lorsque vous exécutez la ressource, le scanneur de profilage exclut la colonne CustomerFullName, car la colonne a un type de données défini par l'utilisateur. Pour résoudre ce problème, vous pouvez enregistrer le type de données user_name comme type de données varchar.

Pour enregistrer le type de données inconnu, effectuez les opérations suivantes :

1. Accédez à l'emplacement `<Enterprise_Data_Catalog Installation Directory>/services/CatalogService/ScannerBinaries` dans le répertoire d'installation d'Enterprise Data Catalog.
2. Décompressez le fichier `catalogMapping.zip`. Le fichier ZIP contient un exemple de fichier **catalog-mapping.xml**.
3. Dans le fichier `catalog-mapping.xml`, entrez l'attribut `source` comme `user_name` et l'attribut cible comme `varchar`.

L'extrait de code suivant montre les modifications que vous devez apporter au fichier :

```
<ColumnDataTypes>
  <!--<item source="CustomerFullName" target="varchar"/>-->
</ColumnDataTypes>
```

4. Créez un fichier ZIP nommé **catalogMapping.zip**.
5. Chargez le fichier vers l'emplacement `<Enterprise_Data_Catalog Installation Directory>/services/CatalogService/ScannerBinaries`.
6. Accédez à l'emplacement `<Enterprise_Data_Catalog Installation Directory>/.../CatalogService/ScannerBinaries/CustomDeployer/`.
7. Dans le fichier `scannerDeployer.xml`, entrez l'extrait de code suivant :

```
<ExecutionContextProperty isLocationProperty="true"
  dependencyToUnpack="catalogMapping.zip">
  <PropertyName>CATALOG_MAPPING_LOC</PropertyName>
  <PropertyValue>scanner_agents/profileMapping</PropertyValue>
</ExecutionContextProperty>
```
8. Enregistrez le fichier.
9. Recyclez le service de catalogue dans Informatica Administrator.
10. Exécutez la ressource dans Catalog Administrator.

Le scanner de profilage utilise le type de données `varchar` pour la colonne `CustomerFullName` pendant l'exécution du profil.

Enregistrement d'un type de données inconnu comme type de données pris en charge

Vous pouvez enregistrer des types de données inconnus comme type de données `varchar` dans Enterprise Data Catalog. Le scanner de profilage utilise les types de données enregistrés pour les colonnes ayant des types de données inconnus.

1. Décompressez le fichier `catalogMapping.zip` situé dans le dossier `<Enterprise_Data_Catalog Installation Directory>/services/CatalogService/ScannerBinaries` du répertoire d'installation d'Enterprise Data Catalog.
Le fichier ZIP contient un exemple de fichier `catalog-mapping.xml`.
2. Ouvrez le fichier `catalog-mapping.xml`.

Le fichier XML contient les éléments XML suivants :

Élément XML	Description
<CatalogMapping> </CatalogMapping>	Élément XML racine.
<Connection> <Name> </Name> </Connection>	Entrez les informations de mappage de type de données pour une source relationnelle. Vous pouvez entrer un type de connexion générique, tel qu'ODBC, ou un type de connexion propre à une base de données relationnelle, tel qu'Oracle.
<source> <Name> </Name> </source>	Si vous entrez un type de connexion générique, vous devez entrer la valeur global pour toutes les bases de données relationnelles que vous pouvez utiliser pour le mappage. Pour le type de connexion propre à une base de données, entrez le nom de la base de données relationnelle. Par exemple, pour le type de connexion Oracle, entrez Oracle comme valeur de base de données source.
<ExcludeDataTypes> </ExcludeDataTypes>	Entrez le type de données que vous souhaitez exclure du mappage.
<ColumnDataTypes> </ColumnDataTypes>	Entrez le mappage de type de données.
<item source=" " target=" " />	Entrez les attributs de type de données source et cible entre guillemets. Vous pouvez entrer un ou plusieurs éléments.

- Entrez les attributs de type de données source et cible comme **varchar**. Vous pouvez entrer plusieurs attributs source et cible.
- Enregistrez le fichier XML et créez un fichier ZIP nommé catalogMapping.zip.
- Chargez le fichier ZIP vers l'emplacement <Enterprise_Data_Catalog Installation Directory>/services/CatalogService/ScannerBinaries du répertoire d'installation d'Enterprise Data Catalog.
- Accédez à l'emplacement <Enterprise_Data_Catalog Installation Directory>/.../CatalogService/ScannerBinaries/CustomDeployer/.
- Ouvrez le fichier scannerDeployer.xml.
- Entrez le contenu suivant dans le fichier :

```
<ExecutionContextProperty isLocationProperty="true"
  dependencyToUnpack="catalogMapping.zip">
  <PropertyName>CATALOG_MAPPING_LOC</PropertyName>
  <PropertyValue>scanner_agents/profileMapping</PropertyValue>
</ExecutionContextProperty>
```

Si le contenu existe déjà dans un élément de commentaire, supprimez ce dernier pour l'élément

```
<ExecutionContextProperty isLocationProperty="true"
  dependencyToUnpack="catalogMapping.zip"> ... </ExecutionContextProperty>
```

L'élément

```
<ExecutionContextProperty isLocationProperty="true"
  dependencyToUnpack="catalogMapping.zip"> ... </ExecutionContextProperty>
```

permet au scanner de profilage de lire le fichier ZIP.

- Enregistrez et fermez le fichier.

10. Recyclez le service de catalogue dans Informatica Administrator.

Remarque: Après avoir modifié ou mis à jour le fichier XML, recyclez le service de catalogue et exécutez la ressource dans Catalog Administrator.

ANNEXE B

Gestion de l'importation à partir de ServiceNow

Cette annexe comprend les rubriques suivantes :

- [Présentation de la gestion de l'importation à partir de ServiceNow, 167](#)
- [Importation de connexions à partir de ServiceNow , 167](#)

Présentation de la gestion de l'importation à partir de ServiceNow

Catalog Administrator se connecte à ServiceNow pour importer des connexions et extraire les métadonnées de configuration dans le catalogue. Les métadonnées de configuration importées incluent le nom d'instance, le nom d'hôte, le numéro de port et le nom de la base de données. Vous pouvez importer les métadonnées de configuration pour les types de ressources Oracle, IBM DB2 et Microsoft SQL Server.

Après l'importation, Catalog Administrator crée une ressource pour chaque connexion importée. Le nom de la ressource est identique au nom de connexion importé.

Importation de connexions à partir de ServiceNow

Importez des connexions à partir de ServiceNow pour extraire des métadonnées de configuration dans le catalogue.

1. Cliquer sur **Gérer > Importation de ServiceNow**.

La fenêtre **Importer des connexions à partir de ServiceNow** s'affiche.

2. Entrez les propriétés suivantes pour vous connecter à ServiceNow :

- **URL de service.** URL pour se connecter à l'instance ServiceNow.
- **Nom d'utilisateur.** Nom d'utilisateur pour l'instance ServiceNow.
- **Mot de passe.** Mot de passe pour l'instance ServiceNow.

3. Cliquez sur **Suivant**.

La fenêtre **Importer des connexions à partir de ServiceNow** affiche les connexions disponibles à partir de ServiceNow.

4. Sélectionnez les connexions que vous souhaitez importer à partir de la liste connexions disponibles, puis cliquez sur **Importer**. Vous pouvez éventuellement cliquer sur le nom de la connexion pour le modifier.

Catalog Administrator affiche une notification sur le statut de la tâche d'importation.

Remarque: La fonctionnalité Importation à partir de ServiceNow est disponible pour l'évaluation technique. La fonctionnalité d'évaluation technique est prise en charge mais elle n'est pas justifiée, ni prête pour la production. Informatica vous recommande d'utiliser cette fonctionnalité uniquement dans des environnements autres que de production.

ANNEXE C

Fichier CSV de ressource de lignage personnalisée

- [Fichier CSV de ressource de lignage personnalisée , 169](#)

Fichier CSV de ressource de lignage personnalisée

Créez une ressource de lignage personnalisée pour afficher le lignage de données à partir de sources de données, telles que les bases de données relationnelles et les bases de données de fichiers plats. La ressource de lignage personnalisée utilise un fichier CSV pour l'extraction des informations de lignage. Les éléments du fichier CSV incluent Association, Connexion de, Connexion à, Objet de et Objet à.

Le tableau suivant décrit les éléments du fichier CSV :

Élément	Description
Association	<p>Indique le type d'association. Spécifiez les valeurs correctes dans la colonne Association en fonction du type d'objet source et cible à lier. Voici quelques types d'associations de flux de données :</p> <ul style="list-style-type: none">- core.DataSourceDataFlow. Pour lier l'objet source et cible au niveau de la source de données. Le flux de données se trouve au niveau du schéma pour les sources relationnelles.- core.DataSetDataFlow. Pour lier l'objet source et cible au niveau du jeu de données. Le flux de données se trouve au niveau de la table pour les sources relationnelles.- core.DirectionaDataFlow. Pour lier l'objet source et cible au niveau de l'élément de données. Le flux de données se trouve au niveau de la colonne pour les sources relationnelles.- core.DirectionaControlFlow. Pour lier l'objet source et cible au niveau de l'élément de données. Le flux de contrôle se trouve au niveau de la colonne pour les sources relationnelles.- core.DataSetControlFlow. Pour lier l'objet source et cible au niveau du jeu de données. Le flux de contrôle se trouve au niveau de la table pour les sources relationnelles.- core.DataSourceControlFlow. Pour lier l'objet source et cible au niveau de la source de données. Le contrôle se trouve au niveau du schéma pour les sources relationnelles. <p>Remarque: Si vous n'entrez pas de valeurs dans la colonne Association, la ressource de lignage personnalisée extrait uniquement les informations de lignage entre les sources relationnelles et de fichiers.</p>
Connexion de	Paramètre de connexion de l'objet source.
Connexion à	Paramètre de connexion de l'objet cible.

Élément	Description
Objet de	ID partiel de l'objet source après l'exécution de l'attribution de la connexion.
Objet à	ID partiel de l'objet cible après l'exécution de l'attribution de la connexion.

Exemples de fichiers CSV de ressource de lignage personnalisée

Vous pouvez entrer des valeurs dans le fichier CSV de la ressource de lignage personnalisée à l'aide des paramètres de connexion et de l'UID d'objet absolu.

Paramètres de connexion

Vous pouvez fournir des valeurs de paramètres de connexion dans le fichier CSV pour l'extraction des informations de lignage. Après avoir exécuté la ressource de lignage personnalisée, vous devez effectuer l'attribution de la connexion pour afficher le lignage dans le catalogue.

L'exemple suivant montre un fichier CSV pour l'importation du lignage entre deux objets relationnels :

```
Association,From Connection,To Connection,From Object,To Object
core.DirectionaDataFlow,<FromConnection>,<ToConnection>,<FromTableName>/
<FromColumnName>,<ToTableName>/<ToColumnName>
core.DataSetDataFlow,<FromConnection>,<ToConnection>,<FromTableName>,<ToTableName>
```

L'exemple suivant montre un fichier CSV pour l'importation du lignage entre deux objets de fichiers structurés :

```
Association,From Connection,To Connection,From Object,To Object
core.DirectionaDataFlow,<FromConnection>,<ToConnection>,<SourceObjectFilePath>/
<FromColumnName>,<TargetObjectFilePath>/<ToColumnName>
core.DataSetDataFlow,<FromConnection>,<ToConnection>,<SourceObjectFilePath>,<TargetObject
FilePath>
```

L'exemple suivant montre un fichier CSV pour l'importation du lignage entre des objets de fichiers relationnels et structurés :

```
Association,From Connection,To Connection,From Object,To Object
core.DirectionaDataFlow,<FromConnection>,<ToConnection>,<SourceObjectFilePath>/
<FromColumnName>,<ToTableName>/<ToColumnName>
core.DataSetDataFlow,<FromConnection>,<ToConnection>,<SourceObjectFilePath>,<ToTableName>
```

UID d'objet absolu

Fournissez l'UID d'objet absolu dans le fichier CSV pour l'extraction des informations de lignage. Pour identifier l'UID d'objet absolu d'une ressource, accédez à la ressource du catalogue, puis copiez l'UID d'objet à partir de l'URL. Par exemple, l'UID d'objet d'une ressource de système de fichiers est `ResourceName://FileServer/<chemin absolu>/File`.

L'exemple suivant montre un fichier CSV pour l'importation du lignage à partir des objets de fichiers relationnels et structurés à l'aide de l'UID d'objet :

```
Association,From Connection,To Connection,From Object,To Object
core.DirectionaDataFlow,,,Flatfile://FileServer/test2/cust2.csv/column1,Oracle_nc://
ORCL/NC/TEST/column1
core.DataSetDataFlow,,,Flatfile://FileServer/test2/cust3.dat,Oracle_nc://ORCL/NC/TEST2
```

Remarque: Si vous souhaitez extraire les informations de lignage à l'aide de l'UID d'objet absolu, vous ne devez pas spécifier de valeurs dans les colonnes FROM CONNECTION (CONNEXION DE) et TO CONNECTION (CONNEXION A) du fichier CSV.

ANNEXE D

Agent Enterprise Data Catalog

Cette annexe comprend les rubriques suivantes :

- [Installation d'Enterprise Data Catalog Agent, 171](#)
- [Installation d'Enterprise Data Catalog Agent en tant que Service Windows, 172](#)
- [Configuration d'Enterprise Data Catalog Agent pour SSL, 173](#)

Installation d'Enterprise Data Catalog Agent

Vous devez installer Enterprise Data Catalog Agent pour extraire les métadonnées de sources de données exécutées sous Microsoft Windows.

Veillez à installer Enterprise Data Catalog Agent pour les sources de données suivantes :

- Service d'intégration de SQL Server
- Cognos
- MSTR
- SAP BO
- Erwin
- SAP Power Designer
- QlikView
- Microsoft SQL Server

Remarque: Vous devez installer Enterprise Data Catalog Agent lorsque vous utilisez l'authentification Windows pour la ressource Microsoft SQL Server.

- Qlik Sense

Remarque: Pour extraire des métadonnées de la ressource Qlik Sense, vous devez installer Enterprise Data Catalog Agent sur le serveur Qlik Sense.

La machine sur laquelle vous installez Enterprise Data Catalog Agent doit disposer d'au moins 512 Mo de RAM, 1 CPU et 600 Mo d'espace disque libre.

Pour les cas d'utilisation à volume élevé tels que l'analyse d'un grand nombre de rapports SAP BO, vous devez disposer de la configuration système minimale requise suivante :

- Configuration de l'agent MITI SAP BO : processeur 32 cœurs, 64 Go de RAM et 500 Go d'espace disque libre.

- Configuration du serveur SAP BO : processeur 8 cœurs, 32 Go de RAM et 200 Go d'espace disque libre.

Pour configurer Enterprise Data Catalog Agent, procédez comme suit :

1. Téléchargez le fichier suivant à partir d'Akamai Download Manager :
EDC_Agent_<version>_Windows.zip
2. Extrayez les fichiers binaires dans un répertoire d'installation sur la machine.
3. Modifiez le numéro de port du connecteur dans le fichier <Installation directory>\tomcat\conf\tomcat.properties. La valeur par défaut est 19980.

Si le port 19980 est utilisé par une autre application, vous pouvez modifier le numéro de port dans le fichier tomcat.properties. Par exemple, vous pouvez définir le port du connecteur sur 19981 et le port d'arrêt sur 19982.
4. Pour démarrer Enterprise Data Catalog Agent, exécutez le fichier startup.bat dans le répertoire suivant : <Installation directory>\tomcat\bin

Enterprise Data Catalog Agent est démarré. Pour redémarrer l'agent Enterprise Data Catalog, exécutez le fichier RestartServerApplication.bat dans le répertoire d'installation.

Lorsque vous configurez une ressource dans Catalog Administrator pour extraire les métadonnées d'une source de données, utilisez le format suivant dans le champ URL d'agent :

http://<hostname>:<connector_port>/MIMBWebServices

Pour extraire les métadonnées de la ressource Cognos, copiez les fichiers JAR du répertoire sdk\java\lib de l'installation du SDK Cognos vers <Répertoire d'installation>\java\CognosRepository. Une fois les fichiers JAR copiés, redémarrez Enterprise Data Catalog Agent.

Installation d'Enterprise Data Catalog Agent en tant que Service Windows

Vous pouvez installer Enterprise Data Catalog Agent en tant que service Windows. Le nom d'affichage du service EDC Agent est le serveur d'applications Meta Integration.

Pour configurer Enterprise Data Catalog Agent, procédez comme suit :

1. Téléchargez le fichier EDC_Agent_<version>_Windows.zip.
2. Extrayez les fichiers binaires dans un répertoire.
3. Exécutez le fichier RestartServerService.bat.

Un service Windows avec le nom d'affichage Apache Tomcat 9.0 MITI et le nom de service MITI est créé.
4. Vérifiez que le service est en cours d'exécution.
5. Accédez à l'URL suivante pour vérifier l'installation : http://<hostname>:19980/MIMBWebServices/mimb.

Si le service est correctement installé, une page contenant les informations suivantes s'affiche :

Threads MIMB en cours d'exécution :

ID	Heure de début	Nom	Fichier journal	Synchronisation
----	----------------	-----	-----------------	-----------------

Ensemble de synchronisations actif : {}

Configuration d'Enterprise Data Catalog Agent pour SSL

Par défaut, Enterprise Data Catalog Agent ne communique pas via une connexion sécurisée. Vous devez utiliser un certificat signé par une autorité de certification (CA) ou auto-signé afin d'activer SSL pour Enterprise Data Catalog Agent. Vous devez également modifier le script `Setup.bat` afin que celui-ci fonctionne comme prévu. Effectuez les étapes suivantes pour activer l'agent pour SSL :

1. Procurez-vous un certificat signé par une autorité de certification (CA) ou créez un certificat auto-signé à l'aide de la commande suivante :

```
keytool -genkey -keyalg RSA -alias edcagent -keystore agent_keystore.jks -storepass  
<keystore password> -validity 365 -keysize 2048
```

2. Modifiez le fichier `Setup.bat` et remplacez la référence au fichier `MMConfiguration.jar` dans la dernière ligne par *.
3. Exécutez le script `Setup.bat` à l'aide de la commande suivante pour installer Enterprise Data Catalog Agent en tant que service Windows et configurer le service pour utiliser le certificat SSL spécifié.

```
.\Setup.bat -s true -sc agent_keystore.jks -sp <keystore password> --tomcat-service true
```

Paramètre	Description
-s	Définissez cette valeur sur true ou false pour activer ou désactiver SSL.
-sc	Spécifiez le fichier X.509 contenant le certificat SSL.
-sp	Spécifiez le mot de passe du fichier de clé SSL.
--tomcat-service	Définissez cette valeur sur true ou false pour installer ou supprimer Tomcat en tant que service

Remarque: Pour afficher la liste des paramètres pouvant être utilisés avec le script, exécutez `.\Setup.bat --help`.

L'agent est démarré et ne peut accepter que les connexions https. Les scanners qui utilisent l'agent doivent être configurés pour utiliser l'URL https pour l'agent.

Important: Si vous utilisez Enterprise Data Catalog Agent avec SAP BO, vous devez lancer l'agent en tant qu'application de bureau standard, et non en tant que service. Pour lancer l'agent en tant qu'application de bureau, exécutez le script `RestartServerApplication.bat`.

ANNEXE E

Lignage entre les ressources PowerCenter et les fichiers plats

Cette annexe comprend les rubriques suivantes :

- [Présentation, 174](#)
- [Configuration et exécution d'une ressource de fichier plat sur une machine Windows, 175](#)
- [Affichage du lignage entre les ressources PowerCenter et les fichiers plats, 176](#)

Présentation

Dans Catalog Administrator, vous pouvez exécuter les ressources PowerCenter et les ressources de fichiers plats pour afficher le lignage qui existe entre elles dans Enterprise Data Catalog.

Lorsqu'un mappage PowerCenter réside sur un ordinateur Windows, il arrive parfois que le lignage entre les ressources PowerCenter et les fichiers plats ne s'affiche pas dans Enterprise Data Catalog. Ce problème peut se produire en raison d'une non concordance du chemin d'accès aux fichiers dans les ressources. Quand Enterprise Data Catalog génère le lignage, le scanneur du système de fichiers utilise le chemin d'accès réseau partagé alors que le scanneur de PowerCenter utilise le chemin d'accès absolu. Pour résoudre ce problème, entrez le chemin d'accès mappé sur l'ordinateur Windows dans le champ **Emplacement physique** de la ressource de système de fichiers. Vous pouvez indiquer le chemin d'accès réseau ou le chemin partagé dans le champ **Chemin d'accès**.

Le champ **Emplacement physique** s'applique aux fichiers plats. Lorsque le lignage est généré entre les fichiers PowerCenter et les fichiers plats, Enterprise Data Catalog utilise le champ **Emplacement physique** à la place du champ **Chemin d'accès** dans le scanneur de système de fichiers. Dans une ressource PowerCenter, si vous choisissez l'option **Attribution automatique des connexions**, Enterprise Data Catalog attribue la connexion en fonction du chemin d'accès absolu.

Exemple

Supposons que les fichiers plats se trouvent dans le dossier `c:\user1\1021\SMB\test_files`. Vous mappez le dossier `c:\user1\1021\SMB` au répertoire partagé `\\SMB\`.

Cet exemple inclut les chemins d'accès aux dossiers suivants :

- `c:\user1\1021\SMB\test_files` est le chemin d'accès absolu.
- `c:\user1\1021\SMB` est le chemin d'accès mappé ou le chemin d'accès absolu du répertoire partagé.
- `\\SMB\` est le répertoire réseau.
- `\\SMB\test_files\` est le chemin d'accès réseau ou le chemin d'accès partagé.

Configuration et exécution d'une ressource de fichier plat sur une machine Windows

Vous devez indiquer le chemin d'accès absolu d'un fichier plat de la ressource de système de fichiers pour afficher le lignage entre les mappages PowerCenter et les sources de fichier plat qui résident dans la machine Windows.

1. Dans Catalog Administrator, cliquez sur **Nouveau > Ressource**.
2. Dans l'assistant **Ressource**, entrez les paramètres suivants dans l'onglet **Général** :

Champ	Description
Nom	Entrez le nom de la ressource.
Description	Facultatif. Entrez une brève description de la ressource.
Type de ressource	Choisissez Système de fichiers comme type de ressource.
Protocole de fichier	Choisissez SMB/CIFS comme protocole de fichier.
Nom d'utilisateur	Entrez le nom d'utilisateur pour accéder à la machine Windows.
Mot de passe	Entrez le mot de passe pour accéder à la machine Windows.
Emplacement physique	S'applique aux fichiers plats. Entrez le chemin d'accès absolu du répertoire partagé ou le chemin d'accès mappé sur la machine Windows.
Hôte	Spécifiez le nom d'hôte ou l'adresse IP de la machine.
Chemin d'accès	Spécifiez le réseau ou le chemin partagé du fichier à partir duquel vous voulez importer les métadonnées dans le catalogue.
Tester la connexion	Cliquez sur Tester la connexion pour vérifier la connexion à l'emplacement spécifié.

L'image suivante montre un exemple d'onglet Général de la ressource de système de fichiers :

FF_Source

General Metadata Load Settings Custom Attributes Schedule Save Cancel

Enter the basic information about the resource.

Name*: FF_Source

Description:

► Additional properties

Resource type*: File System

Connection Properties

File Protocol*: SMB/CIFS

User Name*: Administrator

Password*:

Physical Location: E:\FF_sources\Sub_dir

Host*: 10.65.70.232

Path*: /FF_sources1/Sub_dir

Test Connection

3. Cliquez sur **Suivant**.
4. Dans l'onglet **Paramètres de chargement des métadonnées**, définissez les options **Métadonnées source**, **Découverte de domaines de données composites** et **Découverte de données** en fonction des besoins.
5. Cliquez sur **Suivant** pour ajouter des attributs personnalisés.
6. Cliquez sur **Enregistrer** pour enregistrer la ressource ou sur **Enregistrer et exécuter** pour enregistrer et exécuter la ressource.

Affichage du lignage entre les ressources PowerCenter et les fichiers plats

Après avoir exécuté les ressources PowerCenter et les ressources de fichiers plats, vous pouvez afficher le lignage qui existe entre elles dans Enterprise Data Catalog.

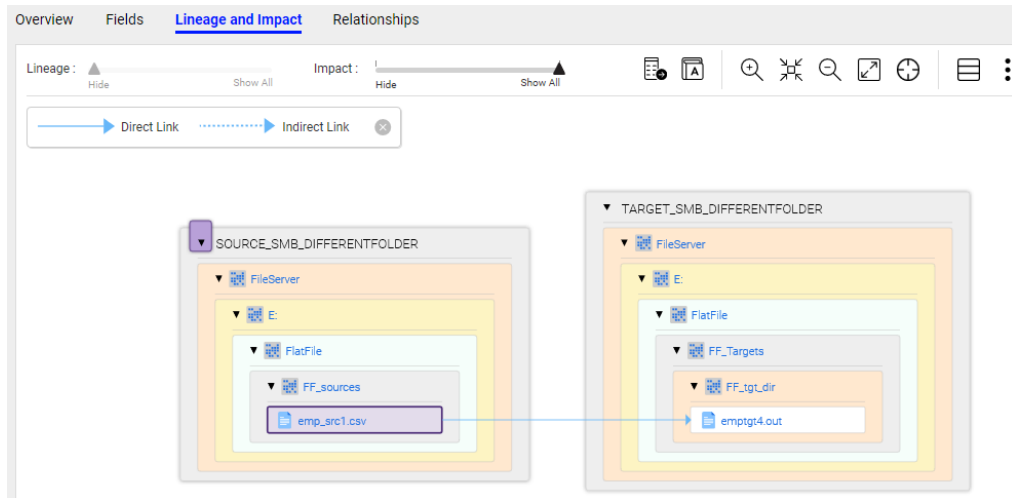
1. Sur la page d'accueil d'Enterprise Data Catalog, entrez le nom de l'actif dans le champ de **recherche**.
2. Cliquez sur l'icône de **recherche**.
Les résultats s'affichent dans la fenêtre **Résultats de la recherche**.
3. Sélectionnez les filtres **Nom de la ressource**, **Type d'actif** et **Type de ressource** appropriés pour affiner les résultats.


4. Cliquez sur le nom de l'actif.

Les détails de l'actif s'affichent dans l'onglet **Présentation**.

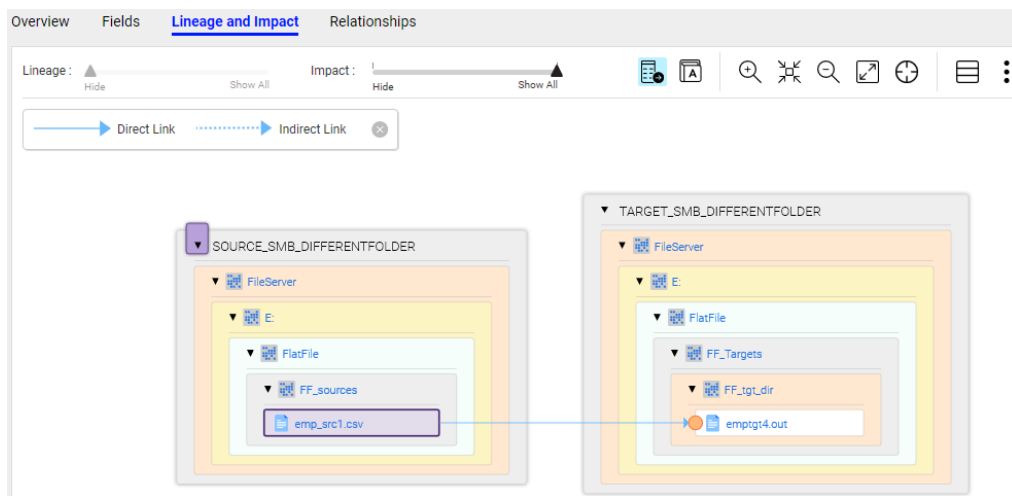
5. Cliquez sur l'onglet **Lignage et impact**.

L'image suivante montre un exemple de lignage d'un actif :




6. Cliquez sur l'icône .

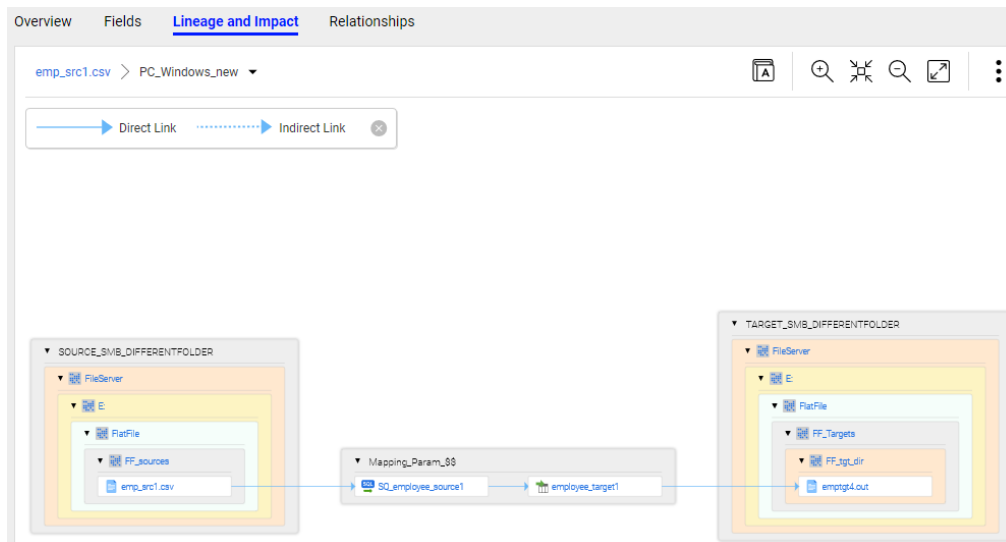
L'image suivante montre un exemple de la vue récapitulative d'un lignage d'actif :



7.

Cliquez sur l'icône .

L'image suivante montre un exemple de la vue détaillée d'un lignage d'actif :



ANNEXE F

Extraction de métadonnées à partir de ressources inaccessibles et hors ligne

Cette annexe comprend les rubriques suivantes :

- [Présentation, 179](#)
- [Processus, 181](#)
- [Composants de l'utilitaire de scanneur, 182](#)
- [Prérequis et meilleures pratiques, 182](#)
- [Exécution du script de scanneur , 183](#)
- [Commande de script, 184](#)
- [Structure du fichier d'archive des métadonnées, 193](#)

Présentation

Informatica fournit un utilitaire de scanneur autonome qui permet d'extraire les métadonnées de ressources hors ligne et inaccessibles. Cet utilitaire contient un script qu'il convient d'exécuter avec les diverses commandes associées dans une séquence.

Catalog Administrator est un outil d'administration qui permet de gérer et de surveiller les ressources, les planifications, les attributs et les connexions. Vous pouvez extraire les métadonnées de l'ensemble des ressources d'une entreprise par le biais de Catalog Administrator. Pour ce faire, vous pouvez créer des connexions, créer et exécuter les ressources, et créer des planifications pour exécuter les ressources à intervalles réguliers. Lorsque les sources de données sont hors ligne ou inaccessibles, l'outil Catalog Administrator tool ne permet pas d'extraire les métadonnées.

Pour extraire les métadonnées de sources externes qui sont inaccessibles au moment de l'exécution ou hors ligne, vous pouvez utiliser l'utilitaire de scanneur autonome. Vous pouvez exécuter le script et ses commandes associées pour extraire les métadonnées de sources de données inaccessibles ou hors ligne, et les publier dans le catalogue. Une fois les métadonnées publiées dans le catalogue, vous pouvez afficher les ressources de données dans Enterprise Data Catalog. Sous Linux, exécutez le script **scanner.sh** et ses commandes associées.

Pour exécuter le script, vous devez disposer des autorisations appropriées sur les dossiers et les fichiers que le script et Enterprise Data Catalog utilisent respectivement pour stocker et publier les métadonnées.

Vous pouvez extraire des métadonnées des sources externes suivantes lorsqu'elles sont inaccessibles au moment de l'exécution ou hors ligne.

Nuage

- Amazon Redshift
- Amazon S3
- Microsoft Azure SQL Data Warehouse
- Microsoft Azure SQL Server
- Glue
- Google BigQuery
- Google Cloud Storage
- Microsoft Azure Blob Storage
- Microsoft Azure Data Lake Store
- Salesforce
- Snowflake
- Workday

Data Engineering

- Apache Atlas
- Cloudera Navigator
- HDFS
- Hive

Business Glossary

- Axon
- Business Glossary

Veille stratégique

- IBM Cognos
- Microstrategy
- Oracle Business Intelligence
- PowerBI
- QlikView
- SAP BW
- SAP BW/4HANA
- Tableau

Intégration des données

- Lignage personnalisé
- Scripts de base de données
- Service Informatica Cloud
- Informatica Platform
- PowerCenter
- Services d'intégration de Microsoft SQL Server

Modélisation de données

- Erwin
- SAP PowerDesigner

Gestion de la base de données

- IBM DB2
- IBM DB2 for z/OS
- IBM Netezza
- JDBC
- Microsoft SQL Server
- Oracle
- SAP HANA
- Sybase
- Teradata

Gestion de fichiers

- Système de fichiers
- OneDrive
- SharePoint

NoSQL

Apache Cassandra

Application

SAP S/4HANA

Remarque: Pour extraire des métadonnées de la source de données SAP S/4HANA en mode hors connexion, spécifiez SAPERP6 comme type de ressource dans la commande scanner.sh generateResourceConfig.

Processus

Vous pouvez exécuter le processus pour extraire et stocker les métadonnées de sources hors ligne ou inaccessibles et les publier dans Enterprise Data Catalog en deux étapes.

Lors de la première étape, un utilisateur d'Enterprise Data Catalog effectue les opérations suivantes :

1. Génération d'un modèle pour le fichier de configuration de ressource en fonction du type de ressource.
2. Chiffre le mot de passe dans le fichier de configuration de la ressource.
3. Modification du fichier pour indiquer les détails de connexion de la source et le mot de passe chiffré.
4. Validation du fichier pour établir une connexion à la source.
5. Une fois la connexion établie, l'utilitaire extrait les métadonnées et les enregistre dans un fichier d'archive des métadonnées.

Lors de la seconde étape, un administrateur d'Enterprise Data Catalog effectue les opérations suivantes :

1. Validation de l'intégrité du fichier d'archive des métadonnées.
2. Publication des métadonnées dans le catalogue.

Une fois les métadonnées publiées dans le catalogue, vous pouvez afficher les ressources de données dans Enterprise Data Catalog.

Composants de l'utilitaire de scanner

Pour exécuter le script et ses commandes associées, vous devez décompresser l'utilitaire puis exécuter le script. Le fichier ZIP du programme d'installation se trouve à l'emplacement `$INFA_HOME/services/CatalogService/ScannerCliUtility/eic-scannercli-utility.zip`. Vous pouvez décompresser le fichier sur la machine sur laquelle vous avez installé le service de catalogue et exécuter le script `scanner.sh` et ses commandes à l'emplacement `$INFA_HOME/services/CatalogService/ScannerCliUtility/`.

Le fichier ZIP du programme d'installation contient les fichiers et répertoires suivants :

Fichiers ou répertoires	Description
/ScannerBinaries	Contient les dépendances tierces et de déploiement de scanner disponibles dans <code>\$INFA_HOME/services/CatalogService/ScannerBinaries</code> .
/ScannerJars	Contient tous les fichiers jar du dossier <code>bin</code> .
/scanner_dependency	Contient la plupart des fichiers jar du fichier <code>access.war</code> .
/lib	Contient les fichiers journaux et les fichiers jar suivants : <ul style="list-style-type: none">- Fichiers jar de l'interface de ligne de commande- <code>log4j.properties</code>- <code>log4j-publisharchive.xml</code>
/bin/scanner.sh	Contient le fichier script principal que vous pouvez exécuter avec ses commandes associées.
/app.properties	Contient les valeurs par défaut des paramètres de commande.

Prérequis et meilleures pratiques

Vous pouvez utiliser les prérequis et meilleures pratiques ci-dessous lorsque vous exécutez le script et ses commandes associées :

Prérequis

Avant d'exécuter le script, vérifiez que les prérequis suivants sont respectés :

- Vous devez disposer des autorisations appropriées pour exécuter les commandes de script **generateResourceConfig**, **encryptPassword**, **validate** et **scan**.
- Assurez-vous de disposer des autorisations de lecture, d'écriture et d'exécution pour le fichier de sortie. La commande **generateResourceConfig** génère le modèle de fichier de configuration de ressource ou le fichier de sortie.
- Assurez-vous de disposer de l'autorisation de lecture pour le fichier d'entrée. Lorsque vous exécutez la commande de script **scan**, celle-ci utilise le fichier de configuration de ressource comme fichier d'entrée.

- L'administrateur d'Enterprise Data Catalog doit disposer des autorisations appropriées pour exécuter les commandes de script **describeArchive** et **publishArchive**.
- Assurez-vous que la machine sur laquelle vous exécutez le script `scanner.sh` dispose d'une taille du tas Java d'au moins 1 Go lorsque vous exécutez les options **generateResourceConfig**, **validate**, **cancel** et **pause**.

Prérequis propres au scanneur

Vérifiez que les prérequis suivants sont respectés pour les ressources externes correspondantes :

Ressources SAP Business Warehouse (SAP BW), SAP BW/4HANA et SAP S/4HANA

- Téléchargez le fichier `sapjco3.jar` et copiez-le à l'emplacement `$CLI_HOME/scanner_dependency`. Par exemple, vous pouvez télécharger et copier le fichier à l'emplacement `<INFA_HOME>/services/CatalogService/ScannerCliUtility/scanner_dependency`.

Remarque: La version de SAP Java Connector prise en charge est la version 3.0.19.

- Incluez le fichier `libsapjco3.so` dans le fichier `SAPJCO.zip` et copiez le fichier `SAPJCO.zip` à l'emplacement `$CLI_HOME/ScannerBinaries`.

Scanneurs avec dépendances externes

Les scanneurs Teradata, Sybase, Redshift et JDBC ont des dépendances externes. Avant d'exécuter le script pour extraire les métadonnées de ces ressources, vous devez extraire les bibliothèques de dépendances externes pour le scanneur et les placer dans les dossiers respectifs sous `$CLI_HOME/ScannerJars/externalDependencies/<scanner>`.

Par exemple, vous pouvez décompresser le fichier `teradataJars.zip` sous `$CLI_HOME/ScannerJars/externalDependencies/TeradataScanner` avant d'exécuter le script pour extraire les métadonnées des ressources Teradata.

Meilleures pratiques

Tenez compte des meilleures pratiques suivantes pour obtenir des performances optimales lors de l'exécution du script :

- Utilisez un seul répertoire de sortie pour chaque exécution du scanneur à la fois. Des erreurs s'affichent si vous utilisez le même répertoire de sortie pour plusieurs exécutions du scanneur en même temps.
- Pour annuler ou suspendre une exécution du scanneur, ouvrez un autre terminal ou une autre commande pour exécuter les commandes.

Exécution du script de scanneur

Pour exécuter le script, vous devez décompresser l'utilitaire et exécuter le script et ses commandes associées dans une séquence afin d'extraire les métadonnées de sources hors ligne ou inaccessibles. Après avoir extrait les métadonnées, vous pouvez les publier dans Enterprise Data Catalog pour afficher les ressources de données.

1. Accédez au répertoire `<INFA_HOME>/services/CatalogService/ScannerCliUtility`.
2. Décompressez le fichier de l'utilitaire `eic-scannercli-utility.zip` dans un dossier local sur votre machine.
3. Exécutez la commande de script **scanner.sh** pour afficher les commandes et les types de ressources pris en charge.

4. Pour afficher la version de l'interface de ligne de commande de la machine, exécutez la commande **scanner.sh version**
5. Pour générer un modèle du fichier de configuration de ressource pour une ressource, exécutez la commande **scanner.sh generateResourceConfig (-resourceType <resource_type>) (-output<path>)**.
6. Pour chiffrer le mot de passe, exécutez la commande **scanner.sh encryptPassword (<password>)**.
7. Pour valider le fichier et établir une connexion à la source, exécutez la commande **scanner.sh validate (-rcf <resource_config_file>)**.
8. Pour analyser, extraire et stocker les métadonnées dans un fichier d'archive des métadonnées, exécutez la commande **scanner.sh scan (-rcf <resource_config_file>) (-output <path>)**.
9. Pour valider le fichier d'archive de métadonnées situé dans le dossier de sortie, exécutez la commande **scanner.sh describeArchive (-arf <path_to_archive_file>)**.
10. Pour publier les métadonnées dans le catalogue, exécutez la commande **scanner.sh publishArchive (-un <isp_user>) (-pd <isp_user_password>) (-sd <isp_security_domain>) (-dh <isp_gateway_host>) (-dp <isp_domain_port>) (-dn <fully_qualified_domain_name>) (-sn <catalog_service_name>) (-rn <resource_name>) (-arf <path_to_archive_file>) (-csh <catalog_service_hostname>) (-csp <catalog_service_port>)**
 - Ajoutez les options **[-dse] [-ts <path_to_truststore>] [-tsp <trust_store_password>]** à la commande **publishArchive...** afin d'activer un domaine SSL pour Enterprise Data Catalog.
 - Ajoutez les options **[-f] [-w] [-v]** pour créer ou mettre à jour une ressource, attendre l'achèvement de la tâche d'ingestion avant le lancement de l'étape suivante du processus et afficher les informations de progression en mode détaillé, respectivement.
11. Ouvrez Enterprise Data Catalog pour rechercher des ressources de données et les afficher.

Commande de script

Scanner.sh est la commande de script que vous pouvez exécuter pour afficher la liste des types de ressource ainsi que la liste des options disponibles. Vous pouvez exécuter le script avec les options et les arguments dans une séquence pour extraire les métadonnées de sources inaccessibles et hors ligne, et les publier dans le catalogue.

La commande scanner.sh utilise la syntaxe suivante :

```
scanner.sh
version
generateResourceConfig (-resourceType <resource_type>) (-output <path>) [-v]
encryptPassword (<password>)
validate (-rcf <resource_config_file>) [-v]
scan (-rcf <resource_config_file>) (-output <path>) [-v]
cancel (-rn <resource_name>) (-output <path>) [-v]
pause (-rn <resource_name>) (-output <path>) [-v]
resume (-rcf <resource_config_file>) (-output <path>) [-v]
describeArchive (-arf <path_to_archive_file>) [-v]
publishArchive (-un <isp_user>) (-pd <isp_user_password>) (-sd <isp_security_domain>) (-dh <isp_gateway_host>) (-dp <isp_domain_port>) (-dn <fully_qualified_domain_name>) (-sn <catalog_service_name>) [-dse] [-ts <path_to_truststore>] [-tsp <trust_store_password>] (-rn <resource_name>) (-arf <path_to_archive_file>) [-f] [-w] [-v] (-csh <catalog_service_hostname>) (-csp <catalog_service_port>)
```


Le tableau suivant décrit les options et les arguments de la commande scanner.sh :

Option	Argument	Description
version	-	Facultatif. Imprime le numéro de version de l'interface de ligne de commande.
generateResourceConfig	(-resourceType <resource_type>) (-output <path>) [-v]	Obligatoire. Génère un modèle du fichier de configuration de ressource.
encryptPassword	(<password>)	Obligatoire. Imprime le mot de passe chiffré.
validate	(-rcf <resource_config_file>) [-v]	Obligatoire. Lit et imprime le fichier de configuration de ressource dans la console, et établit une connexion à la source.
scan	(-rcf <resource_config_file>) (-output <path>) [-v]	Obligatoire. Imprime le statut de la tâche dans la console à intervalles réguliers.
cancel	(-rn <resource_name>) (-output <path>) [-v]	Facultatif. Annule l'analyse de ressource.
pause	(-rn=<resource_name>) (-output=<path>) [-v]	Facultatif. Suspend une analyse de ressource.
resume	(-rcf=<resource_config_file>) (-output=<path>) [-v]	Facultatif. Reprend une analyse de ressource suspendue.
describeArchive	(-arf <path_to_archive_file>) [-v]	Obligatoire. Valide le total de contrôle du fichier d'archive des métadonnées.
publishArchive	(-un <isp_user>) (-pd <isp_user_password>) (-sd <isp_security_domain>) (-dh <isp_gateway_host>) (-dp <isp_domain_port>) (-dn <fully_qualified_domain_name>) (-sn <catalog_service_name>) [-dse] [-ts <path_to_truststore>] [-tsp <trust_store_password>] (-rn <resource_name>) (-arf <path_to_archive_file>) [-f] [-w] [-v] (-csh <catalog_service_hostname>) (-csp <catalog_service_port>)	Obligatoire. Publie les métadonnées dans Enterprise Data Catalog.

Remarque: Le code de sortie du script est **0** pour Réussite et **-1** pour Erreur fatale.

generateResourceConfig

Génère un modèle de fichier de configuration de ressource basé sur le type de ressource. Ce modèle comprend toutes les options de scanner qui s'appliquent au scanner sélectionné.

La commande scanner.sh generateResourceConfig utilise la syntaxe suivante :

```
scanner.sh generateResourceConfig
-resourceType <resourceType>
-output <path>
-v
```

Le tableau suivant décrit les options et les arguments de la commande scanner.sh generateResourceConfig :

Option	Argument	Description
-resourceType	resourceType	Requis. Entrez un type de ressource en respectant la convention d'affectation de noms suivante : <ResourceType> Par exemple, Oracle, MSSqlServer ou Teradata. Remarque: Pour la ressource SAP S/4HANA, spécifiez SAPERP6 comme type de ressource.
-output	chemin	Requis. Chemin d'accès au répertoire de sortie où le fichier de configuration des ressources est généré. Par exemple, vous souhaitez générer le fichier de configuration de scanner dans l'emplacement \$HOME/resources. Entrez le chemin sous \$HOME/resources. Remarque: Une invite [O/N] s'affiche si le fichier cible existe déjà. Sélectionnez O si vous voulez écraser le fichier.
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

Dépannage

Quelles sont les erreurs susceptibles de s'afficher lors de l'exécution de la commande generateResourceConfig ?

Comment puis-je les résoudre ?

La liste suivante décrit les erreurs possibles ainsi que les actions correctives recommandées :

- Le modèle relatif au scanneur n'existe pas. Pour plus d'informations, contactez l'administrateur d'Enterprise Data Catalog.
- Vous ne disposez pas des autorisations appropriées sur le fichier de sortie. Pour attribuer les autorisations nécessaires, contactez l'administrateur d'Enterprise Data Catalog.

encryptPassword

Imprime le mot de passe chiffré dans la console. Le mot de passe chiffré dans le fichier de configuration de la ressource est utilisé pour d'autres commandes.

La commande scanner.sh encryptPassword utilise la syntaxe suivante :

```
scanner.sh
encryptPassword (<password>)
```

Dans la commande **scanner.sh encryptPassword (<password>)**, entrez le mot de passe et exécutez la commande.

validate

Lorsque vous exécutez la commande **scanner. sh validate**, les tâches suivantes sont exécutées :

1. Déchiffrement du mot de passe et test de validité de la connexion à la source de données. Si le test de la connexion échoue, l'exécution du script s'arrête et une erreur s'affiche.
2. Si le test de la connexion réussit, la commande lit le fichier de configuration de ressource et en imprime le contenu sur la console.
3. Une fois le contenu affiché, la commande établit une connexion à la source de données.

La commande **scanner.sh validate** utilise la syntaxe suivante :

```
scanner.sh validate
-rcf <resource_config_file>
-v
```

Le tableau suivant décrit les options et les arguments de la commande **scanner.sh validate** :

Option	Argument	Description
-rcf	resource_config_file	Requis. Entrez le nom du fichier de configuration de ressource.
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

Dépannage

Quelles sont les erreurs susceptibles de s'afficher lors de l'exécution de la commande validate ou de l'échec du test de la connexion ? Comment puis-je les résoudre ?

La liste suivante décrit les erreurs possibles ainsi que les actions correctives recommandées :

- Les informations d'identification ne sont pas valides. Vérifiez que les informations d'identification indiquées sont correctes.
- La source de données n'est pas accessible. Vérifiez les détails de la connexion. Si le problème persiste, contactez l'administrateur de la base de données pour en savoir plus.
- Problèmes liés à la source de données. Pour plus d'informations, contactez l'administrateur de la base de données.
- Le mot de passe n'a pas été chiffré. Exécutez la commande **scanner. sh encrypt Password (<password>)** pour chiffrer le mot de passe. Entrez le mot de passe chiffré dans le fichier de configuration de la ressource, le cas échéant. Exécutez la commande **scanner. sh validate (-rcf <resource_config_file >) [-v]**.

scan

Lorsque vous exécutez la commande **scanner. sh scan**, les tâches suivantes sont exécutées :

1. Exécution de l'analyse de la ressource.
2. Impression périodique du statut de la tâche sur la console.

3. Crée un fichier d'archive intitulé *Archive_\$(resource_name)_\$(version).zip* pour les métadonnées extraites,
4. et l'enregistre dans le dossier de sortie. Le dossier de sortie est situé à l'emplacement que vous avez spécifié à l'aide du paramètre **-output** dans la commande scan.

La commande `scanner.sh scan` utilise la syntaxe suivante :

```
scanner.sh scan
-rcf <resource_config_file>
-output <path>
-v
```

Le tableau suivant décrit les options et les arguments de la commande `scanner.sh scan` :

Option	Argument	Description
-rcf	resource_config_file	Requis. Entrez le nom du fichier de configuration de ressource.
-output	chemin	Requis. Chemin d'accès au répertoire de sortie où le fichier d'archive de métadonnées est généré. Par exemple, vous souhaitez générer le fichier d'archive de métadonnées dans l'emplacement <code>\$HOME/resources</code> . Entrez le chemin sous <code>\$HOME/resources</code> .
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

Dépannage

Quelles sont les erreurs susceptibles de s'afficher lors de l'exécution de la commande scan ? Comment puis-je les résoudre ?

La liste suivante décrit les erreurs possibles ainsi que les actions correctives recommandées :

- Espace disque insuffisant pour enregistrer le fichier d'archive des métadonnées. Augmentez l'espace disque et exécutez la commande.
- Impossible de se connecter à la source de données. Vérifiez les détails de la connexion. Si le problème persiste, contactez l'administrateur de la base de données pour en savoir plus.
- Les informations d'identification ne sont pas valides. Vérifiez que les informations d'identification indiquées sont correctes.

Remarque: Si vous annulez l'analyse, le fichier d'archive des métadonnées ne sera pas créé. Pour plus d'informations sur les erreurs, consultez les fichiers journaux.

cancel

Annule l'analyse de la ressource si elle est en cours et ne crée pas le fichier d'archive des métadonnées.

La commande `scanner.sh cancel` utilise la syntaxe suivante :

```
scanner.sh cancel
-rn <resource_name>
-output <path>
-v
```

Le tableau suivant décrit les options et les arguments de la commande `scanner.sh cancel` :

Option	Argument	Description
-rn	resource_name	Requis. Entrez le nom de la ressource.
-output	chemin	Requis. Chemin d'accès au répertoire de sortie où le fichier d'archive de métadonnées est généré. Par exemple, vous souhaitez générer le fichier d'archive de métadonnées dans l'emplacement <code>\$HOME/resources</code> . Entrez le chemin sous <code>\$HOME/resources</code> .
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

pause

Suspend l'exécution de la ressource et ne crée pas le fichier d'archive des métadonnées. La commande conserve les répertoires d'entrée et de sortie.

La commande `scanner.sh pause` utilise la syntaxe suivante :

```
scanner.sh pause
-rn <resource_name>
-output <path>
-v
```

Le tableau suivant décrit les options et les arguments de la commande `scanner.sh pause` :

Option	Argument	Description
-rn	resource_name	Requis. Entrez le nom de la ressource.
-output	chemin	Requis. Chemin d'accès au répertoire de sortie où le fichier d'archive de métadonnées est généré. Par exemple, vous souhaitez générer le fichier d'archive de métadonnées dans l'emplacement <code>\$HOME/resources</code> . Entrez le chemin sous <code>\$HOME/resources</code> .
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

Dépannage

Quelles sont les erreurs susceptibles de s'afficher lors de l'exécution de la commande `pause` ? Comment puis-je les résoudre ?

La liste suivante décrit les erreurs possibles ainsi que les actions correctives recommandées :

- L'exécution de la ressource n'est pas en cours. Aucune action requise.

resume

Lorsque vous exécutez la commande **scanner.sh resume**, le scanneur reprend l'analyse suspendue de la ressource. Une fois l'exécution de la ressource terminée, la commande crée et enregistre le fichier d'archive des métadonnées.

La commande `scanner.sh resume` utilise la syntaxe suivante :

```
scanner.sh resume
-rcf <resource_config_file>
-output <path>
-v
```

Le tableau suivant décrit les options et les arguments de la commande `scanner.sh resume` :

Option	Argument	Description
-rcf	resource_config_file	Requis. Entrez le nom du fichier de configuration de ressource.
-output	chemin	Requis. Chemin d'accès au répertoire de sortie où le fichier d'archive de métadonnées est généré. Par exemple, vous souhaitez générer le fichier d'archive de métadonnées dans l'emplacement <code>\$HOME/resources</code> . Entrez le chemin sous <code>\$HOME/resources</code> .
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

Dépannage

Quelles sont les erreurs susceptibles de s'afficher lors de l'exécution de la commande `resume` ? Comment puis-je les résoudre ?

La liste suivante décrit les erreurs possibles ainsi que les actions correctives recommandées :

- L'exécution de la ressource est en cours. Aucune action requise.
- L'exécution de la ressource ne peut pas reprendre en raison de circonstances incontournables. Pour atténuer le problème, contactez l'administrateur du catalogue.

describeArchive

Lorsque vous exécutez la commande **scanner.sh describeArchive**, les tâches suivantes sont exécutées :

1. Validation du total de contrôle du fichier d'archive des métadonnées.
2. Impression de la version de l'interface de ligne de commande, des détails de configuration de la ressource et de l'horodatage sur la console.
3. Affichage d'un récapitulatif de l'analyse incluant des statistiques d'analyse.

La commande `scanner.sh describeArchive` utilise la syntaxe suivante :

```
scanner.sh describeArchive
-arf <path_to_archive_file>
-v
```

Le tableau suivant décrit les options et les arguments de la commande `scanner.sh describeArchive` :

Option	Argument	Description
-arf	resource_config_file	Requis. Entrez le chemin du fichier d'archive des métadonnées.
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

publishArchive

Lorsque vous exécutez la commande **scanner.sh publishArchive**, les tâches suivantes sont exécutées :

1. Valide le nom d'utilisateur et le mot de passe de l'administrateur du catalogue et obtient la clé d'API ou le jeton d'autorisation pour les appels d'API suivants.
2. Validation du total de contrôle du fichier d'entrée.
3. Vérification que la version d'Enterprise Data Catalog est identique ou plus récente que celle du scanneur hors ligne.
4. Vérification que la version de configuration de ressource et la version d'Enterprise Data Catalog sont compatibles.
5. Si nécessaire, création d'une ressource dans Enterprise Data Catalog.
6. Vérification que les éléments suivants sont identiques dans la ressource et le fichier de configuration de ressource :
 - a. Configuration de la ressource
 - b. Type et nom de la ressource
7. Ingestion de documents Exchange dans le catalogue.

Vous pouvez afficher sur la console la progression des tâches suivantes :

1. Connexion au domaine et à Enterprise Data Catalog.
2. Création d'un fichier de configuration de ressource ou mise à jour de celle-ci.
3. Chargement de fichier.
4. Analyse de la tâche d'ingestion.
5. Heure de début et heure de fin.
6. URL de surveillance.

La syntaxe de la commande `scanner.sh publishArchive` est la suivante :

```
scanner.sh publishArchive
-un <isp_user>
-pd <isp_user_password>
-sd <isp_security_domain>
-dh <isp_gateway_host>
-dp <isp_domain_port>
-dn <fully_qualified_domain_name>
-sn <catalog_service_name>
[-dse]
[-sse]
[-ts <path_to_truststore>]
[-tsp <trust_store_password>]
-rn <resource_name>
```

```

-arf <path_to_archive_file>
[-f]
[-w]
[-v]
-csh <catalog_service_hostname>
-csp <catalog_service_port>

```

Le tableau suivant décrit les options et les arguments de la commande `scanner.sh publishArchive` :

Option	Argument	Description
-un	isp_user	Obligatoire. Entrez le nom d'utilisateur de l'administrateur du catalogue.
-pd	isp_user_password	Obligatoire. Entrez le mot de passe chiffré de l'administrateur.
-sd	isp_security_domain	Obligatoire. Entrez Natif ou LDAP dans le domaine de sécurité pour se connecter à Enterprise Data Catalog.
-dh	isp_gateway_host	Obligatoire. Entrez le nom d'hôte de passerelle du domaine.
-dp	isp_domain_port	Obligatoire. Entrez le numéro de port du nœud de passerelle.
-dn	fully_qualified_domain_name	Obligatoire. Entrez le nom du nœud Informatica.
-sn	catalog_service_name	Obligatoire. Entrez le nom du service de catalogue.
-dse	<true false>	Facultatif. Spécifiez true pour activer SSL sur le domaine. La valeur par défaut est False.
-ts	path_to_truststore	Facultatif. Entrez le chemin et le nom du fichier truststore qui contient les certificats d'authentification approuvés par le service d'intégration de données. Remarque: Ce champ est requis si vous affectez true à -dse.
-tsp	trust_store_password	Facultatif. Entrez le mot de passe chiffré du fichier truststore. Remarque: Ce champ est requis si vous affectez true à -dse.
-rn	resource_name	Obligatoire. Entrez le nom de la ressource.
-arf	path_to_archive_file	Obligatoire. Entrez le chemin du fichier d'archive des métadonnées.
-f	-	Facultatif. Met à jour ou crée une ressource dans Enterprise Data Catalog.
-w	-	Facultatif. En attente de l'achèvement de la tâche d'ingestion.
-v	-	Facultatif. Affiche les informations en mode détaillé. Le mode détaillé fournit des informations approfondies sur les versions de l'objet, notamment le nom du référentiel, le nom du dossier, le numéro de version et le statut.

Option	Argument	Description
-csh	catalog_service_hostname	Facultatif. Nom de la machine hôte sur laquelle s'exécute le service de catalogue.
-csp	catalog_service_port	Facultatif. Numéro de port du service de catalogue.

Dépannage

Quelles sont les erreurs susceptibles de s'afficher lors de l'exécution de la commande publishArchive ? Comment puis-je les résoudre ?

La liste suivante décrit les erreurs possibles ainsi que les actions correctives recommandées :

- La version de configuration de ressource et les Exdocs ne sont pas compatibles. Implémentez les modifications requises, exécutez les commandes, puis publiez le fichier d'archive dans le catalogue.
- Il existe une incohérence au niveau du type de la ressource ou du nom de la ressource. Implémentez les modifications requises, exécutez les commandes, puis publiez le fichier d'archive dans le catalogue.

Structure du fichier d'archive des métadonnées

Le fichier d'archive des métadonnées ou le fichier Archive.zip contient les fichiers suivants :

Fichier ou répertoire	Description
/EXDOC_STORE	Contient tous les Exdocs, y compris les TOC.
/XMIFiles	Contient les fichiers XML qui se trouvent dans le répertoire \services\CatalogService\XMIFiles*.
/logs	Contient des fichiers journaux tels que les suivants : <ul style="list-style-type: none"> - Stderr - Stdout - Scanner.log - GC.log
/scanner/checkpoints	La tâche d'ingestion crée des points de contrôle à chaque dossier de niveau supérieur pendant l'analyse des tâches. Ces points de contrôle sont enregistrés dans le fichier /scanner/checkpoints.
/scanner/resourceConfig	Fichiers issus de EICRepo/*
/scanner/JobStatus	Fichiers issus de JobStatus/*

Fichier ou répertoire	Description
/config	<p>Contient une copie du fichier de configuration de ressource avec l'extension de fichier <code>.properties</code>. Le fichier a le nom d'utilisateur dummy et le mot de passe dummy.</p> <p>Le dossier contient également le fichier <code>Runtime.log</code> qui comporte les informations suivantes :</p> <ul style="list-style-type: none"> - Variables d'environnement. - Paramètres au niveau du système d'exploitation. - Configuration matérielle, telle que la mémoire physique disponible et l'espace disque. - Environnement d'exécution.
/META-INF/metainfo.json	<p>Le fichier <code>metainfo.json</code> contient des totaux de contrôle qui incluent <code>metadataArchiveChecksum</code>. Le paramètre <code>metadataArchiveChecksum</code> est mis à jour lorsque le script ajoute des entrées au fichier d'archive. Une fois les totaux de contrôle générés, le script génère le fichier <code>metainfo.json</code> et l'ajoute au fichier <code>Archive.zip</code> en dernière entrée.</p> <p>Le fichier <code>metadata.json</code> représente les dates en tant que chaîne (String) au format Temps universel coordonné (UTC).</p> <p>L'exemple suivant montre le contenu d'un fichier <code>metadata.json</code> :</p> <pre> { "createdOn" : 12345678, "scanMode" : "FULL", "cli_version" : "10.2.2", "summary" : { {"Tables" : "10"}. {"Column" : "20" } }, // updateStats() "resourceConfigChecksum" : "abcd", "metadataArchiveChecksum" : "xyz", "startTime" : "YYYY-MM-DDThh:mm:ssZ", "endTime" : "YYYY-MM-DDThh:mm:ssZ", "resourceName" : "OracleResource" ... }</pre>

ANNEXE G

Enterprise Data Catalog Modules complémentaires

Cette annexe comprend les rubriques suivantes :

- [Présentation des modules complémentaires Enterprise Data Catalog, 195](#)
- [Déployer les modules complémentaires, 196](#)
- [Connectez-vous aux modules complémentaires Enterprise Data Catalog, 196](#)
- [Afficher les planifications de ressources et les ressources en double, 196](#)
- [Accepter les domaines de données, 197](#)
- [Associer des termes d'entreprise à partir du fichier de synonymes, 198](#)

Présentation des modules complémentaires Enterprise Data Catalog

Les modules complémentaires Enterprise Data Catalog étendent ou complètent les fonctionnalités disponibles dans l'outil Catalogue et l'outil Catalog Administrator.

Utilisez les modules complémentaires suivants pour effectuer des tâches supplémentaires sur les ressources et les actifs du catalogue :

Rapports d'opérations du catalogue

Affichez les planifications pour toutes les ressources du catalogue et la liste de ressources en double.

CurateAssist

Conservez les actifs de domaines de données en déplaçant les domaines de données spécifiés vers l'état accepté et en rejetant les autres domaines de données en bloc.

Association de termes d'entreprise à partir du fichier de synonymes

Associez des termes d'entreprise d'une ressource Informatica Axon à des colonnes spécifiques dans un actif tabulaire.

Déployer les modules complémentaires

Téléchargez le package d'installation des modules complémentaires Enterprise Data Catalog à partir d'Informatica Marketplace. Le fichier ZIP inclut les fichiers dont vous avez besoin pour déployer les modules complémentaires.

Effectuez les étapes suivantes pour déployer les modules complémentaires :

1. Sur le nœud sur lequel vous souhaitez déployer les modules complémentaires, extrayez le fichier ZIP.
2. Dans le répertoire d'extraction des fichiers, exécutez la commande suivante pour démarrer l'application : `./startup.sh`
Le nom d'hôte et le numéro de port pour accéder à l'application s'affichent au format suivant : `https://<host>:<port>`

Connectez-vous aux modules complémentaires Enterprise Data Catalog

Après avoir déployé les modules complémentaires, effectuez les étapes suivantes pour vous connecter aux modules complémentaires Enterprise Data Catalog :

1. Ouvrez un navigateur.
2. Dans le champ d'adresse, entrez l'URL pour accéder au module complémentaire au format suivant : `https://<host>:<port>`
3. Entrez les propriétés de la connexion pour vous connecter à l'outil Catalog.
Le tableau suivant décrit les propriétés de la connexion :

Propriété	Description
Hôte de catalogue	Nom d'hôte du nœud de passerelle.
Port de catalogue	Numéro de port configuré pour le service de catalogue.
Protocole	Sélectionnez https si vous avez activé une connexion HTTPS sécurisée pour l'outil Catalog.
Nom d'utilisateur	Nom d'utilisateur pour se connecter à l'outil Catalog.
Mot de passe	Mot de passe pour se connecter à l'outil Catalog.

4. Cliquez sur **Se connecter**.

Afficher les planifications de ressources et les ressources en double

Vous pouvez générer le rapport de planifications de ressources pour afficher une liste de toutes les ressources du catalogue et les planifications associées à chaque ressource. Le rapport vous aide à

déterminer la meilleure planification pour minimiser le nombre de travaux d'analyse simultanés des ressources lorsque vous créez une planification. Vous pouvez également générer le rapport de ressources en double pour identifier les ressources de type base de données relationnelle non-JDBC en double. Affichez les types de ressources et le nombre de groupes de ressources en double pour chaque type de ressource à l'aide du rapport.

Afficher les planifications des ressources

1. Accédez aux modules complémentaires Enterprise Data Catalog.
2. Sur la page **Mes services**, cliquez sur **Rapports d'opérations du catalogue**.
3. Sur la page **Informations de catalogue**, vérifiez que l'option Planifications de ressources est sélectionnée comme type de rapport.
4. Cliquez sur **Se connecter** pour afficher le rapport.
Le rapport de planification des ressources affiche les planifications d'analyse dans la vue de calendrier ou de rapport. La valeur par défaut est vue de calendrier. Utilisez l'icône bascule pour basculer entre les vues.

Utilisez la vue du rapport pour afficher une liste de toutes les ressources du catalogue et les planifications d'analyse. Dans la vue du rapport, cliquez sur **Ressources avec planifications** pour afficher uniquement les ressources avec des planifications. Cliquez sur le lien d'accès à la ressource dans la colonne **Détails de la planification** pour afficher les détails de planification de la ressource.

Dans la vue de rapport, vous pouvez trier la liste des ressources, filtrer la liste de ressources, télécharger un rapport contenant la liste de ressources et de planifications dans un fichier CSV et actualiser la liste. Vous pouvez également enregistrer des critères de filtre. Les filtres enregistrés s'affichent dans le menu **Filtres enregistrés**.

Utilisez la vue du calendrier pour afficher les planifications d'analyse des ressources sur un calendrier. Personnalisez la vue à l'aide des options en haut du calendrier.

Afficher les ressources en double

1. Sur la page **Informations de catalogue**, sélectionnez l'option Groupes de ressources en double comme type de rapport.
2. Cliquez sur **Se connecter** pour afficher le rapport.
3. Cliquez sur le nombre correspondant à un type de ressource pour afficher les groupes de ressources en double pour le type de ressource. Les ressources regroupées en tant que doublons présentent des propriétés de connexion identiques.
4. Cliquez sur **Propriétés en double** dans la colonne **Propriétés de la connexion** pour afficher les propriétés de connexion du groupe de ressources. Cette page indique également si un filtre de métadonnées est appliqué à la ressource.

Accepter les domaines de données

Vous pouvez déplacer des domaines de données spécifiés vers l'état accepté et rejeter d'autres domaines de données en bloc à l'aide du module complémentaire Enterprise Data Catalog. Pour effectuer une conservation en bloc, exportez les actifs vers un fichier CSV et spécifiez les domaines de données à déplacer vers l'état accepté. Utilisez le module complémentaire pour analyser le fichier CSV et générer un fichier de sortie. Dans ce dernier, les domaines de données spécifiés passent à l'état accepté et tous les autres

domaines de données passent à l'état rejeté. Importez le fichier CSV dans l'outil Enterprise Data Catalog pour conserver les actifs.

Spécifiez les domaines de données que vous souhaitez faire passer à l'état accepté.

1. Exportez les actifs vers un fichier CSV.
2. Ouvrez le fichier CSV que vous avez exporté.
3. Ajoutez une colonne `AcceptedDataDomains` à la fin de la feuille de calcul.
4. Pour chaque actif à conserver, entrez le domaine de données que vous souhaitez faire passer à l'état accepté dans la colonne `AcceptedDataDomains`.
5. Enregistrez le fichier.

Effectuez une conservation en bloc.

1. Accédez aux modules complémentaires Enterprise Data Catalog.
2. Sur la page **Mes services**, cliquez sur **CurateAssist**.
3. Cliquez sur **Choisir le fichier** et sélectionnez le fichier CSV.
4. Cliquez sur **Conserver** pour analyser le fichier.
5. Cliquez sur le bouton Télécharger pour télécharger le fichier de sortie.
Les domaines de données que vous avez spécifiés dans la colonne `AcceptedDataDomains` passent à l'état accepté et tous les autres domaines de données qui étaient à l'état `Accepté` ou `Inféré` sont rejetés.
6. Importez le fichier CSV que vous avez téléchargé dans l'outil Enterprise Data Catalog pour terminer la conservation du domaine de données pour les actifs.

Associer des termes d'entreprise à partir du fichier de synonymes

Vous pouvez associer des termes d'entreprise d'une ressource Informatica Axon à des colonnes spécifiques dans un actif tabulaire à l'aide du module complémentaire Enterprise Data Catalog. Le module complémentaire prend en charge les actifs tabulaires à partir de ressources de type base de données relationnelle. Pour associer des termes d'entreprise, utilisez un fichier de synonymes au format CSV pour définir les associations. Vous pouvez ensuite spécifier le nom de la ressource, la ressource Informatica Axon qui contient le terme d'entreprise, charger le fichier de synonymes et exécuter l'utilitaire.

Définissez les associations de termes d'entreprise dans un fichier de synonymes.

1. Créez un fichier de synonymes de valeurs séparées par des virgules (CSV) avec les colonnes suivantes : `TableName`, `ColumnName`, `BGMapping`
2. Entrez les noms des colonnes que vous souhaitez associer aux termes d'entreprise et les noms des tables correspondantes.
3. Dans la colonne `BGMapping`, entrez les termes d'entreprise que vous souhaitez associer.
4. Enregistrez le fichier.

Associez des termes d'entreprise à des colonnes spécifiques.

1. Accédez aux modules complémentaires Enterprise Data Catalog.
2. Sur la page **Mes services**, cliquez sur **Association de termes d'entreprise à partir du fichier de synonymes**.
3. Spécifiez la ressource qui contient les colonnes à associer aux termes d'entreprise.

4. Sélectionnez la ressource Informatica Axon qui contient le terme d'entreprise.
5. Choisissez le fichier synonyme.
6. Cliquez sur **Soumettre** pour démarrer l'association de termes d'entreprise.
Le statut du travail s'affiche dans un tableau.
7. Pour télécharger un fichier CSV avec le statut de chaque colonne que vous avez spécifiée, cliquez sur **Télécharger le fichier**. Vous pouvez également afficher les messages d'erreur des associations ayant échoué.

INDEX

A

- accepter
 - domaine de données [198](#)
- Activer
 - provisionnement [84](#)
- administration
 - processus [15](#)
- Aperçu et provisionnement des données [50](#)
- Association de termes d'entreprise à partir du fichier de synonymes [198](#)
- associer
 - terme d'entreprise [198](#)
- attribut
 - personnalisé [98](#)
 - présentation [97](#)
 - système [97](#)
- attribut personnalisé
 - création [99](#)
- attribut système
 - modification [100](#)
- attributs
 - propriétés de la configuration de recherche [99](#)
 - propriétés générales [98](#)

C

- Catalog Administrator
 - Espace de travail Bibliothèque [61](#)
 - espace de travail Configuration [60](#), [62](#)
 - Espace de travail Présentation [60](#)
 - Espace de travail Surveillance [61](#)
 - onglets, affichage [58](#)
 - prérequis [16](#)
 - présentation [58](#)
- catalogue
 - présentation [19](#)
- configuration globale
 - Propriétés du service d'intégration de données [107](#)
 - propriétés générales [106](#)
- configuration réutilisable
 - création [107](#)
 - présentation [106](#)
- connexions
 - attribution [103](#)
 - connexions attribuées par l'utilisateur [103](#)
 - connexions automatiques [103](#)
 - présentation [102](#)
- CurateAssist [198](#)

D

- domaines de données
 - création d'un domaine de données [114](#)

- domaines de données (*a continué*)
 - présentation [21](#)
- Données [50](#)

E

- Enterprise Data Catalog
 - accès à [16](#)
 - concepts [18](#)
 - exemple [20](#)
 - présentation [13](#)
 - présentation de l'administration [15](#)
 - surveillance [108](#)
- Enterprise Unified Metadata
 - architecture [14](#)

F

- Filtre de métadonnées source
 - présentation [71](#)

G

- Gérer l'organisation Cloud
 - Présentation [129](#)

M

- Module complémentaire du rapport de planification des ressources [197](#)
- Modules complémentaires Enterprise Data Catalog [197](#), [198](#)

P

- planification
 - personnalisée [95](#)
 - présentation [20](#)
 - réutilisable [95](#)
- planification réutilisable
 - création [95](#)
- planifications
 - affichage [95](#)
 - présentation [94](#)
- Prérequis [50](#)
- Processus [51](#), [54](#), [56](#)

R

- ressource
 - affichage [87](#)

ressource (*a continué*)

- attributs [64](#)
- création [68](#)
- exécution [86](#)
- modification [85](#)
- planification du scanneur [64](#)
- présentation [19](#), [63](#)
- scanneur [64](#)

ressources de glossaire d'entreprise

- BG_DEFAULT_RESOURCE [49](#)
- synchronisation [49](#)
- tâche de chargement des métadonnées [49](#)

S

scanner.sh

- cancel [188](#)
- describeArchive [190](#)
- encryptPassword [186](#)
- generateResourceConfig [186](#)
- pause [189](#)
- publishArchive [191](#)
- resume [190](#)
- scan [187](#)
- validate [187](#)

scanneur

- présentation [19](#)

surveillance

- filtres [111](#)

Surveillance

- ressources [110](#)
- tableau de distribution des tâches [109](#)

Surveillance (*a continué*)

- tableau de statut de tâches [109](#)
- tâches [111](#)
- synchronisation
- ressources de glossaire d'entreprise [49](#)

T

type de planification

- présentation [94](#)

type de ressource

- présentation [19](#), [69](#)

U

Utilitaire autonome

- Meilleures pratiques [182](#)
- Composants [182](#)
- Exécution du script [183](#)
- Prérequis [182](#)
- Présentation [179](#)
- Processus [181](#)
- scanner.sh [184](#)
- Structure du fichier d'archive des métadonnées [193](#)

V

vue

- planifications de ressources [197](#)