



Informatica®

10.5.1

Guide des données de référence

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.

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.

Informatica et le logo Informatica sont des marques ou des marques déposées d'Informatica LLC aux États-Unis et dans de nombreux autres pays. Une liste actuelle des marques déposées d'Informatica est disponible sur le site <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.

Certaines parties de ce logiciel et/ou de cette documentation sont soumises à des droits d'auteur détenus par des tiers. Les notifications de tiers requises sont incluses avec le produit.

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**

Sommaire

Préface.....	7
Ressources Informatica.	7
Informatica Network.	7
Base de connaissances Informatica.	7
Documentation Informatica.	7
Matrices de disponibilité des produits Informatica.	8
Informatica Velocity.	8
Informatica Marketplace.	8
Support client international Informatica.	8
 Chapitre 1: Introduction aux données de référence.....	 9
Présentation des données de référence.	9
Données de référence Informatica.	10
Données de référence définies par l'utilisateur.	10
Tables de référence.	11
Structure des tables de référence.	12
Privilèges de l'entrepôt de données de référence.	12
Paramètres et tables de référence.	13
Objets de données de référence et contrôle de version.	13
 Chapitre 2: Tables de référence dans l'outil Analyst.....	 14
Présentation des tables de référence de l'outil Analyst.	14
Propriétés des tables de référence.	15
Propriétés générales de la table de référence.	15
Propriétés de la colonne Table de référence.	16
Création d'une table de référence dans l'éditeur de table de référence.	16
Créer une table de référence à partir des données de profil.	17
Créer une table de référence depuis les données de colonne de profil.	17
Création d'une table de référence depuis les modèles de valeurs.	19
Créer une table de référence à partir d'un fichier plat.	20
Propriétés de fichier plat de l'outil Analyst.	20
Création d'une table de référence à partir d'un fichier simple.	21
Créer une table de référence à partir d'une table de base de données.	21
Création d'une table de référence à partir d'une table de base de données.	22
Utilisation avec des tables de référence dans un référentiel modèle avec version.	23
Mises à jour de la table de référence.	23
Gérer les colonnes.	24
Gérer les lignes.	24
Rechercher et remplacer les valeurs.	25
Exportation des données de table de référence.	26

Activer et désactiver les modifications dans une table de référence non gérée.	26
Actualiser les valeurs de la table de référence.	27
Événements suivi d'audit.	27
Voir les événements de suivi d'audit.	28
Règles et directives pour les tables de référence.	28

Chapitre 3: Données de référence dans l'outil Developer..... 30

Présentation des données de référence de l'outil Developer.	30
Données de référence et transformations.	31
Utilisation d'objets de données de référence dans un référentiel modèle avec version.	31
Extraction d'objets de données de référence.	31
Archivage des objets de données de référence.	32
Tables de référence.	32
Propriétés des données de la table de référence.	33
Création d'un objet de table de référence.	33
Création d'une table de référence à partir d'un fichier plat.	34
Créer une table de référence à partir d'une source relationnelle.	35
Ensembles de contenu.	37
Ensembles de caractères.	37
Modèles de classeur.	38
Ensembles de modèles.	38
Modèles probabilistes.	38
Expressions régulières.	39
Ensembles de jetons.	40
Règles et directives pour les modèles probabilistes et les modèles de classeur.	41
Création d'un ensemble de contenu.	42
Création d'un objet de données de référence dans un ensemble de contenu.	43

Chapitre 4: Modèles de classeur..... 44

Présentation des modèles de classeur.	44
Structure du modèle de classeur.	45
Scores de classeur.	45
Exemple de transformation Classeur.	45
Options de modèle de classeur.	46
Données de référence du modèle de classeur.	48
Données de libellé du modèle de classeur.	48
Gestion des libellés du modèle de classeur.	49
Configuration d'un modèle de classeur.	50
Création d'un modèle de classeur.	50
Ajout de données à un modèle de classeur à partir d'une source de données	51
Ajout d'une ligne de données de référence à un modèle de classeur.	52
Ajout d'un libellé à un modèle de classeur.	52
Attribution d'un libellé à des lignes de données de référence.	52

Identification de valeurs de libellés inutilisées.	53
Suppression de ligne d'un modèle de classeur.	53
Suppression d'un libellé d'un modèle de classeur.	53
Compilation d'un modèle de classeur.	53
Opérations de filtre et de recherche.	54
Utilisation d'une valeur de données afin de filtrer les lignes de données de référence.	54
Utilisation d'une valeur de libellé de façon à filtrer les lignes de données de référence.	54
Recherche d'une valeur dans une ligne de données de référence.	54
Opérations de copier/coller.	55
Copie d'un modèle de classeur dans un autre ensemble de contenu.	55
Importation d'un modèle de classeur depuis un autre ensemble de contenu.	55

Chapitre 5: Modèles probabilistes..... 56

Présentation des modèles probabilistes.	56
Structure du modèle probabiliste.	57
Exemple de transformation Libellé.	58
Exemple de transformation Analyseur.	59
Options de modèle probabiliste.	59
Vue Données du modèle probabiliste.	59
Vue Libellé du modèle probabiliste.	61
Données de référence du modèle probabiliste.	62
Données de libellé du modèle probabiliste.	62
Libellé de dépassement.	63
Propriétés du modèle probabiliste.	64
Configuration d'un modèle probabiliste.	64
Création d'un modèle probabiliste vide.	65
Création d'un modèle probabiliste à partir d'un objet de données.	65
Ajout de données à un modèle probabiliste à partir d'une source de données.	66
Ajout d'une ligne de données de référence à un modèle probabiliste.	67
Ajout d'un libellé à un modèle probabiliste.	67
Attribution d'un libellé à une valeur de données de référence.	68
Attribution d'un libellé à plusieurs valeurs de données	68
Suppression de ligne d'un modèle probabiliste.	69
Suppression d'un libellé d'un modèle probabiliste.	69
Compilation du modèle probabiliste.	70
Recherche de lignes de données dans un modèle probabiliste.	70
Filtrage des valeurs de données de référence par l'attribution d'un libellé.	70
Recherche de valeurs de libellés inutilisées.	70
Opérations de copier/coller.	71
Copie d'un modèle probabiliste dans un autre ensemble de contenu.	71
Importation d'un modèle probabiliste depuis un autre ensemble de contenu.	71
Copie des lignes de données de référence dans le presse-papiers.	72

Annexe A: Données de référence et Informatica Data Engineering Integration.....	73
Données de référence et Informatica Data Engineering Integration.	73
Données de référence pour la validation des adresses.	73
Installation des fichiers de données de référence d'adresse.	74
Index.	75

Préface

Consultez le *Guide des données de référence d'Informatica* pour en savoir plus sur les objets de données de référence et les fichiers que vous pouvez utiliser dans Informatica Developer et Informatica Analyst. Vous pouvez utiliser des données de référence pour identifier, analyser, améliorer ou supprimer des valeurs dans vos données source en fonction de vos besoins d'entreprise.

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 rechercher dans la base de connaissances, visitez le site <https://search.informatica.com>. N'hésitez pas à contacter l'équipe Base de connaissances Informatica à l'adresse KB_Feedback@informatica.com pour lui faire part de vos questions, commentaires ou 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 infa_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 aux données de référence

Ce chapitre comprend les rubriques suivantes :

- [Présentation des données de référence, 9](#)
- [Données de référence Informatica, 10](#)
- [Données de référence définies par l'utilisateur, 10](#)
- [Tables de référence, 11](#)
- [Objets de données de référence et contrôle de version, 13](#)

Présentation des données de référence

Les transformations Informatica peuvent utiliser les données de référence pour analyser et mettre à jour des données. Vous pouvez créer des objets de données de référence dans Developer tool et dans l'outil Analyst. Vous pouvez également importer des objets et des fichiers de données de référence dans le référentiel modèle et dans le système de fichiers.

Vous pouvez créer et modifier les types suivants de données de référence :

Tables de référence

Une table de référence contient la version standard et les versions alternatives d'un ensemble de valeurs de données. Vous pouvez ajouter une table de référence à une transformation dans Developer tool afin de vérifier que les valeurs de données sources sont précises et correctement mises en forme.

La plupart des tables de référence contiennent au moins deux colonnes. Une colonne contient la version standard ou préférentielle d'une valeur. Les autres colonnes contiennent des versions alternatives. Lorsque vous ajoutez une table de référence à une transformation, la transformation recherche les données de port d'entrée qui apparaissent également dans la table. Vous pouvez créer des tables avec toute donnée utile au projet de données sur lequel vous travaillez.

Ensembles de contenu

Un ensemble de contenu est un objet du référentiel modèle qui indique les valeurs de données de référence dans le référentiel ou dans un fichier. Lorsque vous ajoutez un ensemble de contenu à une transformation, la transformation recherche les données d'entrée pour les valeurs qui correspondent aux modèles de données dans l'ensemble de contenu.

Vos pouvez télécharger et installer les types suivants de données de référence depuis Informatica :

Fichiers de données de référence d'adresse

Fichiers de données de référence qui contiennent des données pour les adresses livrables d'un pays. La transformation Outil de validation des adresses lit les données de référence. Vous ne pouvez pas créer ni modifier de fichiers de données de référence d'adresse.

Les données de référence d'adresse sont valides pour une période définie et vous devez actualiser vos données de manière régulière, par exemple chaque trimestre.

Fichiers de remplissage d'identité

Fichiers de données de référence qui contiennent des informations sur les identités personnelles, familiales et professionnelles. La transformation Correspondance et la transformation Comparaison utilisent des fichiers de population pour trouver des identités potentielles dans les données d'entrée. Vous ne pouvez pas créer ni modifier de fichiers de remplissage d'identité.

Données de référence Informatica

Vous pouvez acheter et télécharger des données de référence d'adresses et des données de population d'identité à partir d'Informatica.

Vous téléchargez et installez les données de référence indépendamment des applications Informatica. Contactez votre administrateur pour obtenir plus d'informations sur les données de référence à installer sur votre système.

Vous achetez les fichiers de données d'adresses selon un abonnement des pays dont vous voulez vérifier les adresses. Vous pouvez télécharger à tout moment les dernières données d'adresses depuis Informatica pendant la période d'abonnement.

Données de référence définies par l'utilisateur

Vous pouvez utiliser les valeurs d'un objet de données pour créer un objet de données de référence.

Vous pouvez par exemple sélectionner un objet de données ou une colonne de profil qui contient des valeurs spécifiques à un projet ou à une organisation. Créez les objets de données de référence personnalisés depuis les valeurs de colonne.

Vous pouvez construire un objet de données de référence depuis une colonne de données pour vérifier les éléments suivants :

- Les lignes de données des colonnes contiennent le même type d'informations.
- Une valeur source est valide. L'objet de référence peut contenir une liste de valeurs valides ou qui ne le sont pas.

Le tableau suivant répertorie des exemples courants de colonnes de données de projets qui peuvent contenir des données de référence :

Information	Exemple de données de référence
Codes d'unités de conservation de stock (SKU)	Utilisez une colonne SKU pour créer une table de référence de codes SKU valides pour une organisation. Utilisez la table de référence pour trouver des codes SKU corrects ou incorrects dans un jeu de données.
Codes Employé	Utilisez un code Employé ou une colonne d'identifiant Employé pour créer une table de référence de codes Employé valides. Utilisez la table de référence pour rechercher des erreurs dans les données Employé.
Numéros de compte Client	Exécutez un profil sur une colonne de compte Client pour identifier un modèle de numéro de compte. Utilisez le profil pour créer un jeu de jetons de modèles de données incorrects. Utilisez le jeu de jetons pour trouver des numéros de compte qui ne sont pas conformes à la structure de numéro de compte correcte.
Noms Client	Lorsqu'un nom Client contient un prénom, un second prénom et un nom de famille, vous pouvez créer un modèle probabiliste qui définit la structure attendue des chaînes de la colonne. Utilisez le modèle probabiliste pour trouver des chaînes de données qui n'appartiennent pas à la colonne.

Tables de référence

Vous pouvez créer les tables de référence et les mettre à jour dans les outils Analyst tool et Developer tool.

Les tables de référence stockent les métadonnées dans le référentiel modèle. Les tables de référence peuvent stocker des données de colonne dans l'entrepôt de données de référence ou dans une autre base de données. Lorsque l'entrepôt de données de référence stocke les données de colonne, les services Informatica identifient la table comme une table de référence gérée. Lorsqu'une autre base de données stocke les données de colonne, les services Informatica identifient la table comme une table de référence non gérée.

Vous pouvez utiliser une base de données IBM Db2, Microsoft SQL Server, Microsoft Azure SQL, Oracle ou PostgreSQL comme entrepôt de données de référence.

Référencer les options de l'entrepôt de données dans le service de gestion de contenu

Le service de gestion de contenu est un service d'application qui gère la connexion à l'entrepôt de données de référence. Vous pouvez gérer et mettre à jour les propriétés du service de gestion de contenu dans l'outil Administrator tool.

Règles et directives pour la configuration de l'entrepôt de données de référence

Tenez compte des règles et directives suivantes lorsque vous configurez l'entrepôt de données de référence :

- Vérifiez la base de données et le schéma que le service de gestion de contenu utilisera pour les données de référence avant de créer une table de référence gérée.
- Vous pouvez éventuellement spécifier un schéma pour un entrepôt de données de référence IBM Db2, Microsoft SQL Server, Microsoft Azure SQL ou Oracle.

Si vous ne spécifiez pas un schéma de table de référence dans le service de gestion de contenu, le service utilise le schéma spécifié par la connexion à la base de données. Si vous ne spécifiez pas un schéma

dans le service de gestion de contenu ou dans la connexion à la base de données, le service utilise le schéma de base de données par défaut.

- Vous devez spécifier le schéma sur un entrepôt de données de référence PostgreSQL.
- Utilisez un pilote JDBC pour vous connecter à un entrepôt de données de référence PostgreSQL. Pour plus d'informations sur la sélection du pilote JDBC pour un entrepôt de données de référence PostgreSQL, consultez le *Guide d'installation de PowerCenter et de Data Quality*.

Structure des tables de référence

La plupart des tables de référence contiennent au moins deux colonnes. Une colonne contient les versions correctes ou requises des valeurs de données. Les autres colonnes contiennent des versions différentes des valeurs, y compris des versions alternatives qui peuvent s'afficher dans les données source.

La colonne qui contient les valeurs correctes ou requises est appelée la colonne valide. Lorsqu'une transformation lit une table de référence dans un mappage, elle cherche les valeurs dans les colonnes non valides. Lorsque la transformation trouve une valeur non valide, elle renvoie la valeur correspondante de la colonne valide. Vous pouvez également configurer une transformation pour renvoyer une seule valeur commune à la place des valeurs valides.

La colonne valide peut contenir des données en bonne et due forme, par exemple des codes postaux. Elle peut contenir des données relatives à un projet, comme par exemple le nombre d'unités de conservation de stock (SKU) propres à une organisation. Vous pouvez également créer une colonne valide à partir de données incorrectes, comme des valeurs qui contiennent des erreurs de données connues que vous voulez rechercher.

Par exemple, vous créez une table de référence qui contient une liste de numéros SKU valides dans une organisation de vente au détail. Vous ajoutez la table de référence à une transformation Libellé et créez un mappage avec la transformation. Vous exécutez le mappage avec une table de base de données de produits. Lorsque le mappage est exécuté, la transformation Libellé crée une colonne qui identifie les enregistrements de produit ne contenant pas de nombres de SKU valides.

Tables de référence et transformation Analyseur

Créez une table de référence avec une colonne unique pour utiliser les données de la table dans une opération d'analyse basée sur un modèle. Vous configurez la transformation Analyseur pour effectuer une analyse basée sur un modèle et vous importez les données de référence dans la configuration de la transformation.

Privilèges de l'entrepôt de données de référence

Le service de gestion de contenu utilise des privilèges pour restreindre les actions de l'utilisateur sur les tables de référence. Utilisez les options de sécurité de l'outil Administrator pour vérifier ou mettre à jour les privilèges du service.

Pour utiliser des tables de référence, vous devez posséder les privilèges suivants dans le service de gestion de contenu :

- Créer des tables de référence
- Éditer les données de la table de référence
- Éditer les métadonnées de la table de référence

Pour éditer les données d'une table de référence non gérée, vérifiez également que vous avez configuré l'objet de table de référence pour autoriser les modifications.

Remarque: Si vous éditez les métadonnées d'une table de référence non gérée dans une application de base de données, utilisez l'outil Analyst pour synchroniser le référentiel modèle avec la table. Vous devez

synchroniser le référentiel modèle et la table avant d'utiliser la table de référence non gérée dans Developer tool.

Paramètres et tables de référence

Les paramètres vous permettent d'identifier les tables de référence du référentiel modèle. Dans l'outil Developer tool, vous pouvez créer un paramètre permettant d'identifier la table de référence. Vous pouvez également ajouter l'emplacement de la table de référence à un fichier de paramètres.

Lors de la création d'un paramètre dans l'outil Developer tool, vous l'ajoutez à une transformation dans un mappage. Lorsque vous ajoutez l'emplacement de la table de référence à un fichier de paramètres, vous spécifiez le fichier à l'invite de commande lors de l'exécution d'un mappage. Dans les deux cas, le service d'intégration de données lit la table de référence identifiée par le paramètre lors de l'exécution du mappage.

Vous pouvez ajouter un paramètre permettant d'identifier une table de référence aux transformations suivantes :

- Transformation Convertisseur de casse
- Transformation Libellé
- Transformation Analyseur en mode d'analyse de jeton
- Transformation Normalisation

Remarque: La commande *infacmd ms runMapping* vous permet d'exécuter un mappage à l'invite de la commande.

Objets de données de référence et contrôle de version

Si le référentiel modèle qui stocke les objets de données de référence intègre une application dotée d'un contrôle de version, vous pouvez appliquer ce contrôle de version aux objets. Vous pouvez appliquer le contrôle de version aux tables de référence et aux ensembles de contenu.

Vous pouvez archiver et extraire les objets de données de référence depuis un référentiel modèle qui prend en charge le contrôle de version. Vous pouvez annuler une extraction, récupérer une version antérieure d'un objet et restaurer un objet vers une version antérieure. Lorsque les objets de données de référence ne sont pas sous soumis au contrôle de version, le référentiel modèle verrouille tout objet de données de référence que vous modifiez. Les autres utilisateurs ne peuvent pas modifier un objet verrouillé sur lequel vous travaillez. Lorsque vous fermez l'objet, le référentiel modèle libère le verrou et les autres utilisateurs peuvent modifier l'objet.

Remarque: Le contrôle de version s'applique aux métadonnées que le référentiel modèle stocke pour un objet de table de référence non gérée. Le contrôle de version ne s'applique pas aux données d'une table de référence non gérée. Vous ne pouvez pas afficher ou restaurer les données de référence depuis une version antérieure d'une table de référence non gérée.

CHAPITRE 2

Tables de référence dans l'outil Analyst

Ce chapitre comprend les rubriques suivantes :

- [Présentation des tables de référence de l'outil Analyst, 14](#)
- [Propriétés des tables de référence, 15](#)
- [Création d'une table de référence dans l'éditeur de table de référence, 16](#)
- [Créer une table de référence à partir des données de profil, 17](#)
- [Créer une table de référence à partir d'un fichier plat, 20](#)
- [Créer une table de référence à partir d'une table de base de données, 21](#)
- [Utilisation avec des tables de référence dans un référentiel modèle avec version, 23](#)
- [Mises à jour de la table de référence, 23](#)
- [Événements suivi d'audit, 27](#)
- [Règles et directives pour les tables de référence, 28](#)

Présentation des tables de référence de l'outil Analyst

Créez des tables de référence dans l'espace de travail Conception de l'outil Analyst.

Vous pouvez créer une table de référence à partir d'un fichier plat, à partir d'une source de données du référentiel modèle et à partir d'une table d'une autre base de données.

Vous pouvez créer une table de référence depuis une colonne de profil ou un sous-ensemble de données dans une colonne de profil. Vous pouvez également créer une table de référence depuis les modèles de colonne que vous choisissez à partir d'un profil.

Lorsque vous créez ou mettez à jour une table de référence, vous configurez les propriétés de la table et des colonnes de données qu'elle contient.

Propriétés des tables de référence

Vous pouvez afficher et mettre à jour les propriétés de table de référence dans l'outil Analyst tool. Une table de référence affiche des propriétés générales et des propriétés de colonne. Les propriétés générales incluent le nom de la table de référence, la date de création, le nom de connexion de la base de données et le nom de colonne valide. Les propriétés de colonne incluent les noms de colonne, les valeurs de précision et les valeurs d'échelle.

Vous pouvez afficher les propriétés en mode de lecture seule. Pour mettre à jour les propriétés, modifiez ou faites une extraction de la table de référence.

Propriétés générales de la table de référence

Les propriétés générales contiennent des informations sur l'objet de table de référence.

Le tableau suivant décrit les propriétés générales :

Propriété	Description
Nom	Nom de table de référence.
Description	Toutes les descriptions qu'un utilisateur a entrées pour la table de référence.
Emplacement	Emplacement de l'objet de table de référence dans le référentiel modèle.
Colonne valide	Nom de la colonne valide dans la table de référence.
Créé le	Date et heure de création du nom de table de référence.
Créé par	Nom de connexion de l'utilisateur qui a créé la table de référence.
Dernière modification	Date et heure de la mise à jour la plus récente de la table de référence.
Dernière modification par	Nom de connexion de l'utilisateur ayant effectué la mise à jour la plus récente.
Nom de la connexion	Nom de connexion de la base de données qui stocke les valeurs de données de référence.
Type	Type de table de référence. La table de référence peut être gérée ou non gérée.

Propriétés de la colonne Table de référence

Les propriétés de colonne contiennent des informations sur les métadonnées de la colonne.

Le tableau suivant décrit les propriétés de colonne :

Propriété	Description
Nom	Nom de colonne.
Type de données	Type de données pour les données de chaque colonne. Vous pouvez sélectionner l'un des types de données suivants : <ul style="list-style-type: none">- bigint- horodatage- décimal- entier- chaîne Remarque: Vous ne pouvez pas créer une table de référence gérée à partir d'une table Oracle qui utilise un type de données numérique, si celui-ci ne spécifie pas les valeurs pour la précision et pour l'échelle.
Précision	La précision de chaque colonne. La précision est le nombre maximal de chiffres ou de caractères que la colonne peut contenir. Les valeurs de précision que vous configurez dépendent du type de données.
Échelle	L'échelle de chaque colonne. L'échelle est le nombre maximal de chiffres que la colonne peut contenir à droite de la virgule. S'applique aux colonnes décimales. Les valeurs d'échelle que vous configurez dépendent du type de données.
Description	Une description facultative de chaque colonne.
Valeur Null autorisée	Indique si la colonne peut contenir des valeurs Null.
Clé	Identifie une colonne de clé primaire dans une table que vous importez à partir d'une source de base de données. Si vous mettez à jour une table de référence non gérée, vous ne pouvez pas ajouter de valeurs dupliquées dans une colonne qui provient d'une colonne de clé primaire dans la base de données source. Vous pouvez ajouter des valeurs dupliquées dans n'importe quelle colonne d'une table de référence gérée.

Création d'une table de référence dans l'éditeur de table de référence

Définissez la structure de la table et ajoutez des données à une table de référence dans l'éditeur de table de référence.

1. Cliquez sur **Nouveau > Table de référence**.
L'assistant **Nouvelle table de référence** s'affiche.
2. Sélectionnez l'option **Utiliser l'éditeur de table de référence** et cliquez sur **Suivant**.
3. Utilisez l'option **Ajouter une colonne** pour ajouter des colonnes à la table.

4. Configurez les propriétés de chaque colonne.
Les propriétés incluent le nom de la colonne, le type de données, la précision et l'échelle.
Si la colonne contient des données qu'une transformation peut renvoyer dans une recherche de données de référence, sélectionnez l'option Valide.
5. Vous pouvez éventuellement ajouter une colonne pour inclure les descriptions de bas niveau telles que les métadonnées dans la table de référence.
6. Éventuellement, saisissez une note d'audit pour la table.
La note d'audit apparaît dans le journal de suivi d'audit.
7. Cliquez sur **Suivant**.
8. Entrez un nom pour la table de référence et sélectionnez un emplacement pour l'objet de table de référence dans le référentiel modèle.
9. Cliquez sur **Terminer**.

Créer une table de référence à partir des données de profil

Vous pouvez utiliser les données de profil pour créer des tables de référence liées aux données source contenues dans le profil. Utilisez les tables de référence pour trouver les différents types d'informations dans les données source.

Vous pouvez utiliser un profil pour créer ou mettre à jour une table de référence selon les méthodes suivantes :

- Sélectionnez une colonne dans le profil et ajoutez-la à une table de référence.
- Parcourez une colonne de profil et ajoutez un sous-ensemble des données de colonne à une table de référence.
- Sélectionnez une colonne dans le profil et ajoutez les valeurs de forme pour cette colonne à une table de référence.

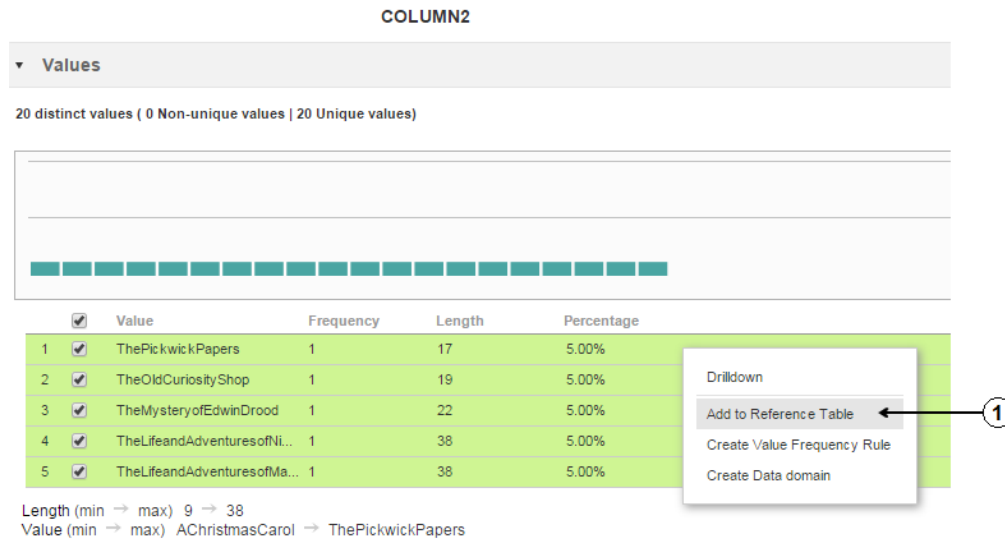
Créer une table de référence depuis les données de colonne de profil

Vous pouvez créer une table de référence depuis une ou plusieurs valeurs dans une colonne de données de profil. Sélectionnez une colonne dans un profil et sélectionnez les valeurs de colonne à ajouter à la table de référence.

1. Ouvrez l'espace de travail **Bibliothèque** dans l'outil Analyst.
2. Sélectionnez la catégorie de ressource **Profils**.
La bibliothèque affiche une liste de profils dans le référentiel modèle.
3. Ouvrez le profil qui contient la colonne à ajouter à une table de référence.
La présentation de profil répertorie les noms des colonnes de profil.
4. Vérifiez les données de colonne.
Pour afficher les données de colonne, cliquez sur le nom de colonne.

5. Dans la vue de profil détaillée, sélectionnez les valeurs de données à ajouter à la table de référence. Vous pouvez sélectionner les valeurs une par une ou vous pouvez tout sélectionner.
6. Cliquez sur le nom de la colonne avec le bouton droit de la souris et sélectionnez **Ajouter à la table de référence**.

L'image suivante montre une colonne de données dans la vue de profil détaillée :



Le chiffre 1 identifie l'option **Ajouter à la table de référence** dans l'image.

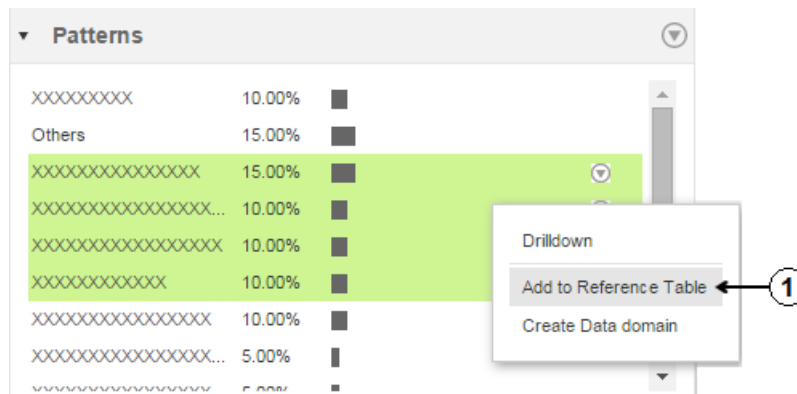
7. L'assistant **Ajouter à la table de référence** s'ouvre.
 Sélectionnez l'option permettant de **Créer une table de référence**.
Remarque: Vous pouvez également sélectionner une option permettant d'ajouter les données à une table de référence en cours.
8. Cliquez sur **Suivant**.
 Le nom de la colonne apparaît par défaut comme nom de la table de référence. Vous pouvez éventuellement mettre à jour le nom.
9. Vous pouvez éventuellement entrer une description et une valeur par défaut.
 L'outil Analyst utilise la valeur par défaut pour tout enregistrement de la table qui ne contient pas de valeur.
10. Cliquez sur **Suivant**.
11. Vérifiez les propriétés de colonne.
 Vous pouvez éventuellement choisir de créer une colonne pour les métadonnées descriptives de bas niveau.
12. Cliquez sur **Suivant**.
13. Vérifiez le nom et la description de la table de référence.
 Éventuellement, saisissez une note d'audit.
14. Sélectionnez un emplacement pour le référentiel modèle de l'objet de table de référence.
15. Cliquez sur **Terminer**.

Création d'une table de référence depuis les modèles de valeurs

Vous pouvez créer une table de référence depuis les modèles de colonnes dans une colonne de profil. Les modèles représentent la composition des valeurs de données dans un ou plusieurs champs de colonnes. Sélectionnez une colonne dans le profil et sélectionnez les modèles à ajouter à la table de référence que vous créez.

1. Ouvrez l'espace de travail **Bibliothèque** dans l'outil Analyst.
2. Sélectionnez la catégorie de ressource **Profils**.
La bibliothèque affiche une liste de profils dans le référentiel modèle.
3. Ouvrez le profil qui contient les modèles de valeur à ajouter à la table de référence.
La présentation de profil répertorie les noms des colonnes de profil.
4. Sélectionnez la colonne qui définit les données de modèle que vous voulez ajouter à la table de référence.
5. Vérifiez les modèles de données de colonne.
Pour afficher les données de colonne, cliquez sur le nom de colonne.
6. Dans la vue Profil détaillé, sélectionnez les modèles de colonne que vous voulez ajouter.
7. Avec le bouton droit de la souris, cliquez sur les modèles sélectionnés et sélectionnez **Ajouter à la table de référence**.

L'image suivante montre les modèles de données d'une colonne dans la vue Profil détaillé :



Le chiffre 1 identifie l'option **Ajouter à la table de référence** dans l'image.

8. L'assistant **Ajouter à la table de référence** s'ouvre.
Sélectionnez l'option permettant de **Créer une table de référence**.
Remarque: Vous pouvez également sélectionner une option permettant d'ajouter les données à une table de référence en cours.
9. Cliquez sur **Suivant**.
Le nom de la colonne apparaît par défaut comme nom de la table de référence. Vous pouvez éventuellement mettre à jour le nom.
10. Vous pouvez éventuellement entrer une description et une valeur par défaut.
L'outil Analyst tool utilise la valeur par défaut pour tout enregistrement de la table qui ne contient pas de valeur.
11. Cliquez sur **Suivant**.
12. Vérifiez les propriétés de colonne.

Vous pouvez éventuellement choisir de créer une colonne pour les métadonnées descriptives de bas niveau.

13. Cliquez sur **Suivant**.
14. Vérifiez le nom et la description de la table de référence.
Éventuellement, saisissez une note d'audit.
15. Sélectionnez un emplacement pour le référentiel modèle de l'objet de table de référence.
16. Cliquez sur **Terminer**.

Créer une table de référence à partir d'un fichier plat

Vous pouvez importer des données de référence depuis un fichier CSV. Utilisez l'assistant **Nouvelle table de référence** pour importer les données du fichier.

Vous devez configurer les propriétés de chaque fichier plat que vous utilisez pour créer une table de référence.

Propriétés de fichier plat de l'outil Analyst

Lorsque vous importez un fichier plat comme table de référence, vous devez configurer les propriétés pour chaque colonne de ce fichier. Les options que vous configurez déterminent la manière dont l'outil Analyst lit les données du fichier.

Le tableau suivant décrit les propriétés que vous pouvez configurer lorsque vous importez le fichier de données pour une table de référence :

Propriétés	Description
Délimiteurs	Caractères utilisés pour séparer les colonnes de données. Utilisez le champ Autre pour entrer un séparateur différent. Les délimiteurs doivent être des caractères imprimables et doivent être différents du caractère d'échappement et des guillemets s'ils sont sélectionnés. Vous ne pouvez pas sélectionner des caractères multi-octets non imprimables comme délimiteurs.
Qualificateur de texte	Guillemet qui définit les limites des chaînes de texte. Choisissez Aucun guillemet, Guillemet simple ou Guillemets doubles. Si vous sélectionnez un guillemet, l'assistant ignore les délimiteurs au sein des paires de guillemets.
Noms des colonnes	Importe les noms de colonne depuis la première ligne. Sélectionnez cette option si les noms de colonne apparaissent dans la première ligne. L'assistant utilise les données de la première ligne dans l'aperçu des noms de colonne. La valeur par défaut n'est pas activée.
Valeurs	Option pour démarrer l'importation d'une valeur à partir d'une ligne. Indique dans la prévisualisation le numéro de la ligne à partir duquel l'assistant démarre la lecture lorsqu'il importe le fichier.

Création d'une table de référence à partir d'un fichier simple

Lorsque vous créez des données de table de référence à partir d'un fichier simple, cette table utilise la structure de colonne du fichier et importe les données du fichier.

1. Cliquez sur **Nouveau > Table de référence**.
L'assistant **Nouvelle table de référence** s'affiche.
2. Sélectionnez l'option pour **Importer un fichier simple**.
3. Cliquez sur **Suivant**.
4. Cliquez sur **Choisir le fichier** pour sélectionner le fichier plat.
5. Sélectionnez une page de code qui correspond aux données dans le fichier simple.
6. Cliquez sur **Charger** pour charger les données du fichier.
7. Cliquez sur **Suivant**.
8. Configurez les propriétés du fichier simple.
Les propriétés identifient le délimiteur que le fichier utilise et si la première ligne du fichier contient des noms de colonne.
9. Pour prévisualiser les propriétés que vous avez configurées, actualisez le panneau **Aperçu**.
10. Cliquez sur **Suivant**.
11. Configurez les propriétés de chaque colonne.
Les propriétés incluent le nom de la colonne, le type de données, la précision et l'échelle.
Si la colonne contient des données qu'une transformation peut renvoyer dans une recherche de données de référence, sélectionnez l'option Valide.
12. Vous pouvez éventuellement ajouter une colonne pour inclure les descriptions de bas niveau telles que les métadonnées dans la table de référence.
13. Éventuellement, saisissez une note d'audit pour la table.
La note d'audit apparaît dans le journal de suivi d'audit.
14. Cliquez sur **Suivant**.
15. Entrez un nom pour la table de référence et sélectionnez un emplacement pour l'objet de table de référence dans le référentiel modèle.
16. Vous pouvez éventuellement entrer une description de l'objet.
17. Cliquez sur **Terminer**.

Créer une table de référence à partir d'une table de base de données

Lorsque vous créez une table de référence à partir d'une table de base de données, vous créez un objet de métadonnées dans le référentiel modèle. Vous importez éventuellement les données de la table dans l'entrepôt de données de référence.

Lorsque vous créez une table de référence gérée, vous importez les données de colonne dans l'entrepôt de données de référence. Lorsque vous créez une table de référence non gérée, vous identifiez la table de base de données qui stocke les données de colonne.

Avant de créer la table de référence, vérifiez que le domaine Informatica contient une connexion à la base de données contenant les données de référence. Si le domaine ne contient pas de connexion à la base de données, vous pouvez en définir une dans l'outil Analyst.

Pour créer une connexion à la base de données, cliquez sur **Gérer > les connexions**.

Création d'une table de référence à partir d'une table de base de données

Pour créer la table de référence, connectez-vous à une base de données et sélectionnez la table contenant les données de référence.

1. Sélectionnez **Nouveau > Table de référence**.

L'assistant **Nouvelle table de référence** s'affiche.

2. Sélectionnez l'option **Connecter à une table relationnelle**.

Pour créer une table de référence qui ne stocke pas de données dans l'entrepôt de données de référence, sélectionnez **Table non gérée**.

Pour permettre aux utilisateurs de modifier une table de référence non gérée, sélectionnez l'option **Éditable**.

Cliquez sur **Suivant**.

3. Sélectionnez la connexion de base de données dans la liste des connexions.

Cliquez sur **Suivant**.

4. Dans le panneau **Tables**, sélectionnez une table.

5. Vérifiez les propriétés de la table dans le panneau **Propriétés**.

Cliquez sur **Prévisualisation des données** pour prévisualiser les données de la table.

Cliquez sur **Suivant**.

6. Dans le panneau **Attributs des colonnes**, sélectionnez la colonne valide.

Si vous créez une table de référence gérée, vous pouvez effectuer les actions suivantes dans le panneau **Attributs des colonnes** :

- Éditer les noms des colonnes de la table de référence.
- Ajouter une colonne de métadonnées pour les descriptions de niveau de ligne.

7. Vous pouvez éventuellement ajouter une colonne pour inclure les descriptions de bas niveau telles que les métadonnées dans la table de référence.

8. Éventuellement, saisissez une note d'audit pour la table.

La note d'audit apparaît dans le journal de suivi d'audit.

9. Cliquez sur **Suivant**.

10. Entrez un nom pour la table de référence et sélectionnez un emplacement pour l'objet de table de référence dans le référentiel modèle.

11. Vous pouvez éventuellement entrer une description de la table de référence.

12. Cliquez sur **Terminer**.

Utilisation avec des tables de référence dans un référentiel modèle avec version

Ouvrez une table de référence en mode de lecture seule. Pour travailler sur la table de référence, vous devez passer en mode de modification ou vous devez extraire la table de référence depuis le référentiel modèle.

1. Sur la barre d'outils Informatica, cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.
2. Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
3. Pour modifier la version actuelle de la table de référence, cliquez sur **Modifier**.
Pour modifier la table de référence dans un référentiel modèle avec version, faites une extraction de la table de référence.
4. Lorsque vous avez terminé de travailler sur la table de référence, cliquez sur **Terminer**. L'outil Analyst tool enregistre vos modifications dans la table de référence.
Si vous avez extrait la table de référence depuis un référentiel modèle avec version, archivez l'objet. Un référentiel modèle avec version ne met pas à jour la version de la table de référence tant que vous n'avez pas archivé l'objet.

Mises à jour de la table de référence

Les données métier que contient une table de référence peuvent changer avec le temps. Examinez et mettez à jour les données et métadonnées d'une table de référence pour vérifier que la table contient des informations correctes. Mettez les tables de référence à jour dans l'outil Analyst. Vous pouvez mettre à jour les données et métadonnées d'une table de référence gérée ou non gérée.

Vous pouvez effectuer les opérations suivantes sur les données et métadonnées de table de référence :

Gérer les colonnes

Vous pouvez ajouter et supprimer des colonnes, et éditer les propriétés de colonne.

Gérer les lignes

Vous pouvez ajouter des lignes de données à une table de référence.

Éditer les valeurs de données de référence

Vous pouvez éditer une valeur de données de référence.

Remplacer les valeurs de données

Utilisez l'option **Rechercher et remplacer** pour remplacer les valeurs de données qui ne sont plus correctes ou pertinentes pour l'organisation. Vous pouvez trouver une valeur dans une colonne et la remplacer par une autre valeur. Vous pouvez remplacer toutes les valeurs d'une colonne par une valeur unique.

Exporter une table de référence

Exportez une table de référence vers un fichier de valeurs séparées par des virgules (CSV), un fichier dictionnaire ou un fichier Excel.

Activer ou désactiver les éditions sur une table non gérée

Mettez à jour une table de référence non gérée pour activer ou désactiver les éditions des données et métadonnées d'une table.

Actualiser les données de la table de référence

Rechargez les données de la table de référence dans l'outil Analyst pour afficher les dernières modifications dans les données.

Gérer les colonnes

Vous pouvez ajouter des colonnes à une table de référence et mettre à jour les propriétés de colonne. Vous pouvez également mettre à jour le statut d'édition d'une table de référence non gérée.

1. Cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.
2. Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
3. Pour modifier la version actuelle de la table de référence, cliquez sur **Modifier**.
Pour modifier la table de référence dans un référentiel modèle avec version, faites une extraction de la table de référence.
4. Ouvrez le menu **Actions** et sélectionnez **Modifier les propriétés de colonne**.
La boîte de dialogue **Modifier les propriétés de colonne** s'ouvre. Utilisez les options de cette boîte de dialogue pour effectuer les opérations suivantes :
 - Ajouter une colonne.
 - Modifier la colonne valide de la table.
 - Modifier un nom de colonne.
 - Mettre à jour le texte descriptif d'une colonne.
 - Mettre à jour le statut d'édition d'une table de référence non gérée.
 - Mettre à jour la note d'audit de la table.
5. Lorsque vous avez terminé les opérations à effectuer, cliquez sur **OK**.

Gérer les lignes

Vous pouvez ajouter, éditer ou supprimer les lignes dans une table de référence.

1. Cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.
2. Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
3. Pour modifier la version actuelle de la table de référence, cliquez sur **Modifier**.
Pour modifier la table de référence dans un référentiel modèle avec version, faites une extraction de la table de référence.
4. Éditez les lignes de données. Vous pouvez éditer les lignes de données en procédant comme suit :

- Pour ajouter une ligne, sélectionnez **Actions > Ajouter une ligne**.
Dans la boîte de dialogue **Ajouter une ligne**, entrez une valeur dans la colonne valide et dans au moins une autre colonne. Éventuellement, saisissez une note d'audit.
Cliquez sur **OK** pour ajouter la ligne.
- Pour mettre à jour une seule valeur de données, cliquez sur la valeur et mettez jour les données.
Après avoir mis à jour les données, utilisez les options de ligne pour les accepter ou les rejeter. Vous ne pouvez pas entrer une note d'audit lorsque vous entrez directement des données dans la ligne de données.
- Pour mettre à jour les valeurs de données dans une ligne, sélectionnez **Actions > Modifier la ligne**.
Dans la boîte de dialogue **Modifier la ligne**, entrez une valeur dans une ou plusieurs colonnes. Éventuellement, saisissez une note d'audit.
Cliquez sur **Appliquer** pour mettre à jour les données dans les colonnes que vous avez sélectionnées.
- Pour mettre à jour les valeurs dans plusieurs lignes, sélectionnez les lignes à modifier et sélectionnez **Actions > Modifier la ligne**.
Dans la boîte de dialogue **Modifier plusieurs lignes**, entrez une valeur dans une ou plusieurs colonnes. Éventuellement, saisissez une note d'audit.
Cliquez sur **OK** pour mettre à jour les données dans les colonnes que vous avez sélectionnées.
- Pour supprimer des lignes, sélectionnez les lignes concernées et cliquez sur **Actions > Supprimer**.
Dans la boîte de dialogue **Supprimer les lignes**, Éventuellement, saisissez une note d'audit.
Cliquez sur **OK** pour supprimer les lignes.

Remarque: Utilisez l'outil Developer pour éditer des données de ligne dans une grande table de référence. Par exemple, si une table de référence contient plus de 500 lignes, éditez-la dans l'outil Developer.

Rechercher et remplacer les valeurs

Vous pouvez rechercher et remplacer des valeurs de données dans une table de référence. Utilisez les options de recherche et remplacement lorsqu'une table contient une ou plusieurs instances d'une valeur de données que vous devez mettre à jour.

1. Cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.
2. Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
3. Pour modifier la version actuelle de la table de référence, cliquez sur **Modifier**.
Pour modifier la table de référence dans un référentiel modèle avec version, faites une extraction de la table de référence.
4. Cliquez sur **Actions > Rechercher et remplacer**.
La barre d'outils **Rechercher et remplacer** apparaît.
5. Entrez les critères de recherche dans la barre d'outils :
 - Entrez une valeur de données dans le champ **Rechercher**.
 - Sélectionnez les colonnes dans lesquelles effectuer la recherche. Par défaut, la recherche porte sur toutes les colonnes.

- Entrez une valeur de données dans le champ **Remplacer par**.
- Utilisez les options suivantes pour effectuer un remplacement individuel ou global des valeurs :
 - Utilisez les options **Suivant** et **Précédent** pour rechercher les valeurs une par une.
 - Pour remplacer une valeur, sélectionnez **Remplacer**.
 - Pour afficher toutes les instances de la valeur, sélectionnez **Tout mettre en surbrillance**.
 - Pour remplacer toutes les instances de la valeur, sélectionnez **Remplacer tout**.

Exportation des données de table de référence

Exportez les données d'une table de référence vers un fichier de valeurs séparées par des virgules, un fichier dictionnaire ou un fichier Microsoft Excel. Vous pouvez exporter les données en mode de lecture seule.

- Cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.
- Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
- Cliquez sur **Actions > Exporter les données**.
La boîte de dialogue **Exporter les données dans un fichier** s'ouvre.
Le tableau suivant décrit les options de la boîte de dialogue :

Option	Description
Nom de fichier	Nom du fichier qui doit contenir les données. L'opération d'exportation crée le fichier.
Format de fichier	Format du fichier qui doit contenir les données. Sélectionnez l'un des formats suivants : <ul style="list-style-type: none"> • csv. Fichier de valeurs séparées par des virgules. Format par défaut • xls. Fichier Microsoft Excel. • dic. Fichier dictionnaire Informatica.
Exporter noms champs comme première ligne	Option des noms des colonnes. Sélectionnez cette option pour indiquer que la première ligne du fichier contient les noms des colonnes.
Page de code	Page de code des données de référence. La page de code par défaut est UTF-8.

- Cliquez sur **OK** pour exporter le fichier.

Activer et désactiver les modifications dans une table de référence non gérée

Vous pouvez activer ou désactiver les mises à jour des valeurs et des colonnes de données dans une table de référence non gérée.

Enregistrez la table de référence avant de modifier son statut d'édition.

- Cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.

2. Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
3. Pour modifier la version actuelle de la table de référence, cliquez sur **Modifier**.
Pour modifier la table de référence dans un référentiel modèle avec version, faites une extraction de la table de référence.
4. Ouvrez le menu **Actions** et sélectionnez **Modifier les propriétés de colonne**.
La boîte de dialogue **Modifier les propriétés de colonne** s'ouvre.
5. Sélectionnez ou désélectionnez l'option **Editable**.

Remarque: Les conditions suivantes s'appliquent à une table de référence non gérée qui permet les mises à jour par l'utilisateur :

- La table de référence doit utiliser des types de données simples tels que Chaîne et Nombre.
- Ne définissez aucune contrainte sur les métadonnées de la table de référence et ne spécifiez aucune valeur par défaut pour une colonne.

Actualiser les valeurs de la table de référence

Vous devrez peut-être actualiser les valeurs que l'outil Analyst tool affiche pour la table de référence.

Pour recharger les valeurs de la table de référence, cliquez sur **Actions > Actualiser**. L'outil Analyst tool récupère les versions actuelles des valeurs de données depuis la base de données.

Événements suivi d'audit

Vous pouvez afficher un suivi d'audit des modifications que les utilisateurs ont effectuées sur une table de référence. La vue Suivi d'audit de la table de référence vous permet d'afficher les événements de suivi d'audit. Vous pouvez filtrer les événements de suivi d'audit que l'outil Analyst tool affiche.

Le tableau suivant décrit les options de filtre que vous pouvez spécifier :

Option	Description
Date	Dates de début et de fin des actions à afficher. Utilisez les options de calendrier pour définir les dates.
Type	Type d'événement de suivi d'audit. Vous pouvez afficher les types d'événements suivants : <ul style="list-style-type: none">- Données. Événements associés aux valeurs de données de la table de référence. Ils comprennent les opérations d'ajout, de suppression et de mise à jour de ligne.- Métadonnées. Événements associés aux métadonnées de la table de référence. Les événements incluent des opérations permettant de créer la table de référence, d'ajouter ou de supprimer une colonne et d'archiver la table de référence. Remarque: Vous ne pouvez pas afficher les événements de données et de métadonnées simultanément.

Option	Description
Utilisateur	Utilisateur ayant modifié la table de référence. Le filtre affiche le nom complet et le nom de connexion de l'utilisateur.
Statut	Statut des événements du journal de suivi d'audit. Le statut correspond à l'action que vous avez effectuée dans l'éditeur de table de référence. Par exemple, le statut peut indiquer qu'un utilisateur a créé la table de référence ou ajouté une ligne.

Les événements de journal de suivi d'audit comprennent également les commentaires de suivi d'audit et les valeurs de colonnes que vous avez insérées, actualisées ou supprimées.

Voir les événements de suivi d'audit

Affichez les événements de suivi d'audit afin de trouver les mises à jour que les utilisateurs ont appliquées à une table de référence. Vous pouvez afficher les événements de suivi d'audit en mode de lecture seule.

1. Cliquez sur **Ouvrir**.
La bibliothèque de ressources s'ouvre.
2. Sélectionnez la catégorie de ressource **Tables de référence**, puis sélectionnez un nom de table de référence.
La table de référence s'ouvre en mode de lecture seule.
3. Cliquez sur **Suivi d'audit**.
4. Configurez les options des filtres.
Vous pouvez filtrer par date, type ou statut de mise à jour ou par le nom de l'utilisateur qui a effectué la mise à jour.
5. Cliquez sur **Afficher**.
Les événements du journal s'affichent pour les options de filtre que vous avez spécifiées.

Règles et directives pour les tables de référence

Utilisez les règles et directives suivantes lorsque vous travaillez avec les tables de référence de l'outil Analyst :

- Quand vous créez une table de référence à partir de modèles de colonnes extrapolés dans un format, l'outil Analyst renseigne la table de référence avec des modèles de colonnes dans un format différent.
Par exemple, quand vous créez une table de référence pour la colonne modèle X(5), l'outil Analyst affiche le format suivant pour le modèle de colonne dans la table de référence : XXXXX.
- Lorsque vous importez une table de base de données Oracle, vérifiez la longueur de la colonne VARCHAR2 de la table. L'outil Analyst tool ne peut pas importer une table de base de données Oracle qui contient une colonne VARCHAR2 dont la longueur est supérieure à 1000.
- Pour lire une table de référence, vous devez exécuter les autorisations pour la connexion à la base de données qui stocke les valeurs des données de la table. Par exemple, si l'entrepôt de données de référence stocke les valeurs de données, vous devez disposer d'autorisations d'exécution sur la connexion à l'entrepôt de données de référence. Vous devez disposer d'autorisations d'exécution pour accéder en

lecture ou en écriture à la table de référence. Les autorisations de connexion à la base de données s'appliquent à toutes les données de référence dans la base de données.

- Lorsque vous exécutez un mappage avec une transformation qui spécifie une table de référence, le mappage utilise la version actuelle de la table de référence dans le référentiel modèle. Vous ne pouvez pas sélectionner une version d'historique de la table de référence lorsque vous configurez la transformation.

Si un autre utilisateur restaure la table de référence sur une version antérieure dans une session de l'outil Developer tool simultanée, les versions de la table de référence ne sont plus identiques entre les sessions. Si vous configurez et exécutez un mappage qui utilise la table de référence, il peut échouer, car la session en cours n'identifie pas la version actuelle de la table de référence. Pour vous assurer que le mappage utilise la table de référence actuelle, actualisez le référentiel modèle avant d'exécuter le mappage.

- Lorsque vous configurez une table de référence non gérée pour autoriser les modifications, vérifiez que la table de référence utilise des types de données simples tels que Chaîne et Nombre. Vérifiez également que la table de référence ne définit aucune contrainte sur les métadonnées de la table de référence ni n'utilise de valeurs par défaut pour les colonnes.
- Toutes les valeurs en microsecondes doivent être zéro dans toute valeur de données de table de références qui utilise le type de données horodatage.

CHAPITRE 3

Données de référence dans l'outil Developer

Ce chapitre comprend les rubriques suivantes :

- [Présentation des données de référence de l'outil Developer, 30](#)
- [Données de référence et transformations, 31](#)
- [Utilisation d'objets de données de référence dans un référentiel modèle avec version, 31](#)
- [Tables de référence, 32](#)
- [Ensembles de contenu, 37](#)

Présentation des données de référence de l'outil Developer

Vous pouvez créer, mettre à jour et afficher les propriétés de configuration pour les objets de données de référence dans l'outil Developer.

Utilisez l'outil Developer pour créer et mettre à jour les types d'objets suivants :

Tables de référence

Une table de référence contient la version standard et les versions alternatives d'un ensemble de valeurs de données. Vous pouvez ajouter une table de référence à une transformation dans l'outil Developer afin de vérifier que les valeurs de données sources sont précises et correctement mises en forme.

Ensembles de contenu

Un ensemble de contenu est un objet du référentiel modèle qui indique les valeurs de données de référence dans le référentiel ou dans un fichier. Un ensemble de contenu contient différents types de données de référence que vous pouvez utiliser pour effectuer des opérations de recherche dans les transformations de qualité des données.

Vous pouvez également travailler avec des fichiers de données de référence d'adresse et des fichiers de population d'identité dans l'outil Developer. Vous sélectionnez les fichiers de données de référence d'adresse lorsque vous configurez une transformation Outil de validation des adresses. Vous sélectionnez les fichiers de population d'identité lorsque vous configurez une transformation Correspondance pour l'analyse de correspondance d'identité.

Données de référence et transformations

Plusieurs transformations lisent des données de référence pour effectuer des tâches de qualité des données.

Les transformations suivantes peuvent lire les données de référence :

- Outil de validation des adresses. Lit les données de référence d'adresses afin de vérifier l'exactitude des adresses.
- Convertisseur de casse. Lit des tables de données de référence pour identifier les chaînes dont la casse doit être modifiée.
- Classeur Lit des données d'ensemble de contenu pour identifier le type d'information d'une chaîne.
- Comparaison. Lit des données de la population d'identité lors de l'analyse de doublons.
- Libellé. Lit des données d'ensemble de contenu pour identifier et libeller des chaînes.
- Correspondance. Lit des données de la population d'identité lors de l'analyse de doublons.
- Analyseur. Lit des données d'ensemble de contenu pour analyser des chaînes en se basant sur les informations qu'elles contiennent.
- Normalisation. Lit les tables de données de référence pour normaliser des chaînes dans un format commun.

Le programme d'installation de Data Quality Content contient des objets de données de référence Informatica que vous pouvez importer.

Utilisation d'objets de données de référence dans un référentiel modèle avec version

Si vous travaillez avec des tables de référence ou des ensembles de contenu dans un référentiel modèle avec version, le référentiel peut appliquer le contrôle de version aux objets. Pour appliquer le contrôle de version à un objet, un utilisateur archive l'objet dans le référentiel modèle.

Si une table de référence ou un ensemble de contenu n'est pas soumis au contrôle de version, vous pouvez ouvrir et mettre à jour l'objet en dehors du système de contrôle de version. Lorsque vous ouvrez l'objet, le référentiel modèle le verrouille de sorte qu'aucun autre utilisateur ne peut y travailler.

Si une table de référence ou un ensemble de contenu est soumis au contrôle de version, ouvrez l'objet en mode de lecture seule. Pour travailler sur l'objet, faites-en une extraction depuis le référentiel modèle. Vous pouvez également extraire l'objet puis l'ouvrir. Archivez l'objet pour en créer une version contenant vos dernières modifications.

Extraction d'objets de données de référence

Pour travailler sur une table de référence ou un ensemble de contenu qu'un utilisateur a archivé dans le référentiel modèle, faites une extraction de l'objet depuis le référentiel.

1. Dans Object Explorer, accédez à une table de référence ou à un ensemble de contenu.
2. Cliquez sur le nom de l'objet avec le bouton droit de la souris et cliquez sur **Ouvrir**.
L'objet s'ouvre en mode de lecture seule.
3. Cliquez sur le nom de l'objet avec le bouton droit de la souris, puis cliquez sur **Extraire**.
Vous pouvez modifier l'objet.

Archivage des objets de données de référence

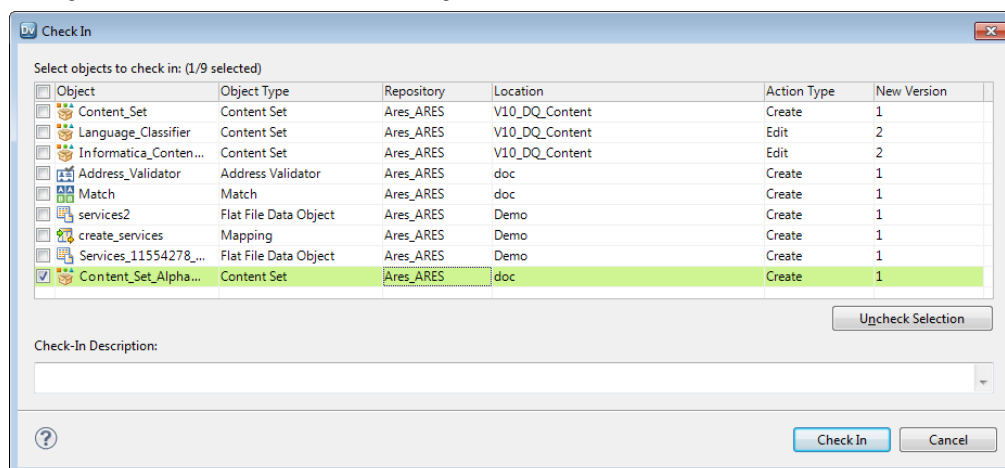
Lorsque vous avez terminé de travailler sur une table de référence ou un ensemble de contenu que vous avez extrait du référentiel modèle, archivez l'objet.

Pour afficher la liste des objets extraits actuellement, ouvrez l'onglet **Objets extraits** sous l'éditeur de table de référence.

1. Enregistrez toute modification apportée à la table de référence ou à l'ensemble de contenu.
2. Dans Object Explorer, accédez à la table de référence ou à l'ensemble de contenu.
3. Cliquez sur le nom de l'objet avec le bouton droit de la souris, puis cliquez sur **Archiver**.

La boîte de dialogue **Archiver** s'ouvre.

L'image suivante montre la boîte de dialogue :



4. Sélectionnez un ou plusieurs objets à archiver dans le référentiel.

Remarque: Vous pouvez archiver un objet qui n'est pas ouvert dans la session en cours. Vous pouvez archiver n'importe quel objet à l'état Extraît.

5. Vous pouvez éventuellement entrer une description de l'opération.
6. Cliquez sur **Archiver**.

L'opération d'archivage met à jour le numéro de version de l'objet. Si vous archivez l'objet pour la première fois, le référentiel modèle crée la version un (1) de l'objet.

Tables de référence

Vous pouvez ajouter une table de référence à une transformation dans Developer tool. Vous configurez la transformation pour trouver des valeurs de la table de référence dans les données d'entrée et pour écrire en sortie les valeurs valides correspondantes dans la table de référence.

Pour créer une table de référence dans Developer tool, utilisez l'une des méthodes suivantes :

- Créer une table de référence vide et entrer les valeurs de données.
- Créer une table de référence à partir des données d'un fichier plat.
- Créer une table de référence à partir des données d'un synonyme, d'une vue ou d'une table de base de données.

Propriétés des données de la table de référence

Vous pouvez afficher les propriétés des données et métadonnées de la table de référence dans l'outil Developer. L'outil Developer affiche les propriétés lorsque vous ouvrez la table de référence depuis le référentiel modèle.

Une table de référence affiche des propriétés générales et des propriétés de colonne. Vous pouvez afficher les propriétés de table de référence dans l'outil Developer. Vous pouvez afficher et modifier les propriétés de table de référence dans l'outil Analyst.

Le tableau suivant décrit les propriétés générales d'une table de référence :

Propriété	Description
Nom	Nom de la table de référence.
Description	Description facultative de la table de référence.

Le tableau suivant décrit les propriétés de colonne d'une table de référence :

Propriété	Description
Valide	Identifie la colonne contenant les données de référence valides.
Nom	Nom de chaque colonne.
Type de données	Type de données de chaque colonne.
Précision	Précision de chaque colonne.
Échelle	Échelle de chaque colonne.
Description	Description du contenu de la colonne. Vous pouvez éventuellement ajouter une description lorsque vous créez la table de référence.
Inclure une colonne pour les descriptions de bas niveau	Indique que la table de référence contient une colonne pour les descriptions des données de colonne.
Valeur par défaut	Valeur par défaut des champs de la colonne. Vous pouvez éventuellement ajouter une valeur par défaut lorsque vous créez la table de référence.
Nom de connexion	Nom de la connexion à la base de données contenant les valeurs de données de la table de référence.

Création d'un objet de table de référence

Sélectionnez cette option lorsque vous voulez créer une table de référence vide et ajouter des valeurs manuellement.

1. Sélectionnez **Fichier > Nouveau > Table de référence** dans le menu de l'outil Developer.
2. Dans l'assistant Nouvelle table, sélectionnez **Table de référence vide**.
3. Saisissez le nom de la table.
4. Sélectionnez un projet dans lequel stocker les métadonnées de la table.

Cliquez sur **Parcourir** lorsque vous êtes sur le champ Emplacement. La boîte de dialogue **Sélectionner un emplacement** s'affiche à l'écran et affiche les projets du référentiel. Sélectionnez le projet dont vous avez besoin.

Cliquez sur **Suivant**.

5. Ajoutez une ou plusieurs colonnes à la table. Cliquez sur l'option **Nouveau** pour créer une colonne.

Le tableau suivant décrit les propriétés de colonne que vous pouvez définir :

Propriété	Valeur par défaut
Valide	Effacé. Sélectionnez la colonne qui contient les valeurs valides.
Nom	colonne
Type de données	chaîne
Précision	10
Échelle	0
Description	Vide. Entrez éventuellement une description de la colonne.

Le tableau suivant répertorie les propriétés que vous pouvez définir pour la table de référence :

Propriété	Valeur par défaut
Vous pouvez inclure une colonne pour les descriptions de niveau de ligne.	Effacé
Description	Vide
Valeur par défaut	Vide
Note d'audit	Vide

6. Cliquez sur **Terminer**.

Création d'une table de référence à partir d'un fichier plat

Vous pouvez créer une table de référence à partir des données stockées dans un fichier plat.

1. Sélectionnez **Fichier > Nouveau > Table de référence** dans le menu de l'outil Developer.
2. Dans l'assistant Nouvelle table, sélectionnez **Table de référence à partir d'un fichier plat**.
3. Parcourez les répertoires pour accéder au fichier à utiliser comme source de données pour la table.
4. Saisissez le nom de la table.
5. Sélectionnez un projet dans lequel stocker les métadonnées de la table.

Cliquez sur **Parcourir** lorsque vous êtes sur le champ Emplacement. La boîte de dialogue **Sélectionner un emplacement** s'affiche à l'écran et affiche les projets du référentiel. Sélectionnez le projet dont vous avez besoin.

Cliquez sur **Suivant**.

6. Définissez UTF-8 comme page de code.
7. Précisez les délimiteurs utilisés par le fichier plat.

8. Si le fichier plat contient des noms de colonnes, sélectionnez l'option d'importation des noms de colonnes à partir de la première ligne du fichier.
9. Le tableau suivant décrit les propriétés facultatives de la table :

Propriété	Valeur par défaut
Qualificateur de texte	Pas de guillemets
Débuter l'importation à la ligne	Ligne 1
Séparateur de lignes	\012 LF (\n)
Traiter les délimiteurs consécutifs comme uniques	Effacé
Caractère d'échappement	Vide
Conserver le caractère d'échappement dans les données	Effacé
Nombre maximum de lignes à consulter	500

Cliquez sur **Suivant**.

10. Sélectionnez la colonne qui contient les valeurs valides.
11. Le tableau suivant décrit les propriétés facultatives :

Propriété	Valeur par défaut
Vous pouvez inclure une colonne pour les descriptions de niveau de ligne.	Effacé
Note d'audit	Vide
Valeur par défaut	Vide
Nombre maximum de lignes à consulter	500

Cliquez sur **Terminer**.

La table de référence s'ouvre dans l'espace de travail de l'outil Developer.

Créer une table de référence à partir d'une source relationnelle

Vous pouvez créer une table de référence à partir d'une table relationnelle, d'un synonyme relationnel ou d'une vue relationnelle.

Lorsque vous créez une table de référence gérée, vous importez les données de colonne dans l'entrepôt de données de référence. Lorsque vous créez une table de référence non gérée, vous identifiez la table de base de données qui stocke les données de colonne.

Avant de créer la table de référence, vérifiez que le domaine Informatica contient une connexion à la base de données contenant les données de référence.

Vous pouvez configurer une connexion de base de données dans l'Explorateur de connexion. Si Developer tool n'affiche pas l'Explorateur de connexion, sélectionnez **Fenêtre > Afficher l'aperçu > Explorateur de connexion** dans le menu de Developer tool.

Création d'une table de référence à partir d'une source relationnelle

Pour créer la table de référence, connectez-vous à une base de données et sélectionnez la table contenant les données de référence.

1. Sélectionnez **Fichier > Nouveau > Table de référence** dans le menu de Developer tool.
2. Dans l'assistant Création de table, sélectionnez **Table de référence à partir d'une source relationnelle**. Cliquez sur **Suivant**.
3. Sélectionnez la connexion à une base de données.
Cliquez sur **Parcourir** lorsque vous êtes sur le champ Connexion. La boîte de dialogue **Sélectionner une connexion** s'ouvre et affiche les connexions de base de données disponibles.
Cliquez sur **OK** lorsque vous avez sélectionné une connexion.
4. Sélectionnez une ressource de base de données.
Cliquez sur **Parcourir** lorsque vous êtes sur le champ Ressource. La boîte de dialogue **Sélectionner une ressource** s'ouvre et affiche les ressources sur la connexion de base de données. Explorez la base de données et sélectionnez un synonyme, une vue ou une table de base de données.
Vous pouvez aussi prévisualiser les informations d'entité sur la ressource.
5. Saisissez le nom de la table.
6. Sélectionnez un emplacement pour l'objet de table de référence.
Cliquez sur **Parcourir** lorsque vous êtes sur le champ Emplacement. La boîte de dialogue **Sélectionner un emplacement** s'affiche et présente les projets du référentiel.
Sélectionnez un emplacement et cliquez sur **Suivant**.
7. Pour créer une table de référence qui ne stocke pas de données dans l'entrepôt de données de référence, sélectionnez **Table non gérée**.
Pour permettre aux utilisateurs de modifier une table de référence non gérée, sélectionnez l'option **Éditable**.
Cliquez sur **Suivant**.
8. Sélectionnez la colonne qui contient les valeurs valides.
9. Le tableau suivant décrit les propriétés facultatives que vous pouvez spécifier :

Propriété	Valeur par défaut
Vous pouvez inclure une colonne pour les descriptions de niveau de ligne.	Effacé
Description	Effacé
Valeur par défaut	Vide
Note d'audit	Vide
Nombre maximum de lignes à consulter	500

10. Cliquez sur **Terminer**.

Ensembles de contenu

Un ensemble de contenu est un objet de référentiel modèle qui stocke les données ou métadonnées pour d'autres objets de données de référence. Un ensemble de contenu peut inclure des jeux de caractères, des ensembles de modèles, les jeux de jetons, des expressions régulières, des modèles probabilistes et des modèles de classeur. Utilisez un jeu de contenu pour définir et organiser les objets de données de référence qui sont associés à un seul projet, type d'information ou objectif métier.

L'outil Developer inclut les jeux de caractères définis par le système et des jeux de jetons qui ne s'affichent pas dans le référentiel modèle. Pour afficher et utiliser les objets définis par le système, configurez une stratégie dans la transformation Libellé, la transformation Analyseur ou la transformation Normalisation.

Ensembles de caractères

Un jeu de caractères contient des expressions qui identifient certains caractères et certaines plages de caractères. Vous pouvez utiliser des jeux de caractères dans les transformations Libellé qui utilisent le mode de libellé de caractère.

Les plages de caractère spécifient une plage séquentielle de codes de caractère. Par exemple, la plage de caractère « [A-C] » correspond aux caractères majuscules « A, » « B, » et « C ». Cette plage de caractère ne correspond pas aux caractères minuscules « a, » « b, » ou « c ».

Utilisez les jeux de caractères pour identifier un caractère spécifique ou une plage de caractères comme faisant partie des opérations de libellé. Par exemple, vous pouvez libeller tous les numéros d'une colonne qui contient des numéros de téléphone. Après avoir libellé les nombres, vous pouvez identifier les modèles avec une transformation Analyseur et écrire les modèles problématiques dans des ports de sortie séparés.

Propriétés de jeu de caractères

Configurez les propriétés qui déterminent les opérations de libellé de caractères pour un jeu de caractères.

Le tableau suivant décrit les propriétés d'un jeu de caractères défini par l'utilisateur :

Propriété	Description
Étiquette	Définit l'étiquette appliquée par une transformation Libellé aux données qui correspondent au jeu de caractères.
Mode standard	Active une vue d'édition simple qui inclut des champs pour la plage de début et la plage de fin.
Plage de début	Spécifie le premier caractère dans une plage de caractères.
Plage de fin	Spécifie le dernier caractère dans une plage de caractères. Pour une plage contenant un caractère unique, ne remplissez pas ce champ.
Mode avancé	Active une vue d'édition avancée dans laquelle vous pouvez saisir manuellement des plages de caractères à l'aide de caractères de plages et de caractères délimiteurs.
Caractère de plage	Modifie de manière temporaire le symbole représentant une plage de caractères. Le caractère de plage rétablit le caractère par défaut lorsque vous fermez le jeu de caractères.
Caractère délimiteur	Modifie de manière temporaire le symbole qui sépare les plages de caractères. Le caractère délimiteur rétablit le caractère par défaut lorsque vous fermez le jeu de caractères.

Modèles de classeur

Un modèle de classeur analyse les chaînes d'entrée et détermine les types d'information que les chaînes sont le plus susceptibles de contenir. Vous pouvez utiliser un modèle de classeur dans une transformation Classeur.

Un modèle de classeur contient des lignes de données de référence et des valeurs de libellés. Les lignes représentent les données d'entrée du port que vous pouvez connecter à la transformation Classeur. Les valeurs de libellés décrivent les types d'informations contenus par les lignes de données. Lors de la configuration d'un modèle de classeur, attribuez un libellé à chaque ligne de données de référence du modèle.

Pour lier les lignes de données de référence aux libellés d'un modèle de classeur, vous devez compiler le modèle. Le processus de compilation génère une série d'associations logiques entre les lignes de données et les valeurs de libellés. Lors de l'exécution d'un mappage qui lit le modèle, le service d'intégration de données applique la logique du modèle aux données d'entrée de la transformation Classeur. Le service d'intégration de données renvoie les libellés qui décrivent le plus précisément les informations de chaque champ de données d'entrée.

Vous pouvez créer un modèle de classeur dans l'outil Developer. Le référentiel modèle stocke l'objet de modèle de classeur. L'outil Developer écrit les lignes de données, les libellés et les données de compilation dans un fichier de la structure de répertoires Informatica.

Ensembles de modèles

Un ensemble de modèles contient des expressions qui identifient les modèles de données dans la sortie d'une opération de libellé de jeton. Vous pouvez utiliser les ensembles de modèles pour analyser le port de sortie des données à jeton et écrire des chaînes correspondantes sur un ou plusieurs ports de sortie. Utilisez les ensembles de modèles dans les transformations Analyseur qui utilisent le mode d'analyse de modèles.

Par exemple, vous pouvez configurer une transformation Analyseur pour utiliser les ensembles de modèles qui identifient les noms et les initiales. Cette transformation utilise les ensembles de modèles pour analyser la sortie d'une transformation Libellé en mode libellé de jeton. Vous pouvez configurer la transformation Analyseur pour écrire les noms et les initiales dans la sortie pour séparer les ports.

Propriétés de l'ensemble de modèles

Configurez les propriétés qui déterminent les modèles dans un ensemble de modèles.

Le tableau suivant décrit la propriété pour un ensemble de modèles défini par l'utilisateur :

Propriété	Description
Patron	Définit les modèles que l'analyseur de modèles recherche. Vous pouvez entrer plusieurs modèles pour un ensemble de modèles. Vous pouvez entrer des modèles construits à partir d'une combinaison de caractères génériques, de caractères et de chaînes.

Modèles probabilistes

Un modèle probabiliste analyse les valeurs de données d'entrée et détermine les types d'informations que les valeurs sont le plus susceptibles de contenir. Utilisez un modèle probabiliste dans une transformation Libellé et une transformation Analyseur.

Un modèle probabiliste contient des valeurs de données de référence et des valeurs de libellés. Les valeurs de données de référence représentent les données d'un port d'entrée que vous connectez à la

transformation. Les valeurs de libellés décrivent les types d'informations que les valeurs de données de référence contiennent. Vous attribuez un libellé à chaque valeur de données de référence du modèle.

Pour lier les valeurs de données de référence aux libellés d'un modèle probabiliste, vous devez compiler le modèle. Le processus de compilation génère une série d'associations logiques entre les valeurs de données et les libellés. Lorsque vous exécutez un mappage qui lit le modèle, le service d'intégration de données applique la logique du modèle aux données d'entrée de la transformation. Le service d'intégration de données renvoie le libellé qui décrit les plus précisément les valeurs de données d'entrée.

Créez un modèle probabiliste dans l'outil Developer. Le référentiel modèle stocke l'objet de modèle probabiliste. L'outil Developer écrit les valeurs de données, les libellés et les données de compilation dans un fichier dans la structure de répertoires Informatica.

Expressions régulières

Dans le contexte d'ensembles de contenu, une expression régulière est une expression que vous pouvez utiliser dans des opérations d'analyse et de libellé. Utilisez les expressions régulières pour identifier une ou plusieurs chaînes dans les données d'entrée. Vous pouvez utiliser des expressions régulières dans les transformations Analyseur qui utilisent le mode d'analyse de jeton. Vous pouvez également utiliser des expressions régulières dans les transformations Libellé qui utilisent le mode libellé de jeton.

Les transformations Analyseur utilisent des expressions régulières pour faire correspondre les modèles dans les données d'entrée et analyser toutes les chaînes de correspondance vers une ou plusieurs sorties. Par exemple, vous pouvez utiliser une expression régulière pour identifier toutes les adresses électroniques dans les données d'entrée et analyser chaque composant d'adresse électronique vers une sortie différente.

Les transformations Libellé utilisent des expressions régulières pour faire correspondre un modèle d'entrée et créer un libellé unique. Les expressions régulières qui ont plusieurs sorties ne génèrent pas plusieurs libellés.

Propriétés de l'expression régulière

Configurez les propriétés qui permettent de déterminer la méthode d'identification et d'écriture de chaînes de sortie d'une expression régulière.

Le tableau suivant décrit les propriétés d'une expression régulière définie par l'utilisateur :

Propriété	Description
Nombre de sorties	Définit le nombre de ports de sortie que l'expression régulière écrit.
Expression régulière	Définit un modèle que la transformation Analyseur utilise pour faire correspondre des chaînes.
Expression de test	Contient des données que vous entrez pour tester l'expression régulière. Dès que vous entrez des données dans ce champ, celui-ci met en évidence les chaînes qui correspondent à l'expression régulière.
Expression suivante	Déplace vers la chaîne suivante qui correspond à l'expression régulière et modifie la police de cette chaîne en gras.
Expression précédente	Déplace la chaîne précédente qui correspond à l'expression régulière et modifie la police de cette chaîne en gras.

Ensembles de jetons

Un jeu de jetons contient des expressions identifiant les jetons spécifiques. Vous pouvez utiliser des jeux de jetons dans les transformations Libellé qui utilisent le mode libellé de jeton. Vous pouvez également utiliser des jeux de jetons dans les transformations Analyseur qui utilisent le mode d'analyse de jeton.

Utilisez les jeux de jetons pour identifier des jetons dans le cadre des opérations de libellé et d'analyse. Par exemple, vous pouvez utiliser un jeu de jetons pour étiqueter toutes les adresses électroniques qui utilisent un format « NomCompte@NomDomaine ». Après avoir étiqueté les jetons, vous pouvez utiliser la transformation Analyseur pour écrire les adresses électroniques dans les ports de sortie que vous indiquez.

Propriétés des jeux de jetons

Configurez les propriétés qui déterminent les opérations de libellé pour un jeu de jetons.

Le tableau suivant décrit les propriétés d'un jeu de caractères défini par l'utilisateur :

Propriété	Mode de jeu de jetons	Description
Nom	N/A	Définit le nom du jeu de jetons.
Description	N/A	Décrit le jeu de jetons.
Options de jeu de jetons	N/A	Définit si le jeu de jetons utilise le mode d'expression régulière ou de caractère.
Étiquette	Expression régulière	Définit le libellé qu'une transformation Libellé applique aux données qui correspondent au jeu de jetons.
Expression régulière	Expression régulière	Définit un modèle que la transformation Libellé utilise pour établir des correspondances avec les chaînes.
Expression de test	Expression régulière	Contient des données que vous entrez pour tester l'expression régulière. Lorsque vous saisissez des données dans ce champ, il met en surbrillance les chaînes qui correspondent à l'expression régulière.
Expression suivante	Expression régulière	Déplace vers la chaîne suivante qui correspond à l'expression régulière et modifie la police de cette chaîne en gras.
Expression précédente	Expression régulière	Déplace la chaîne précédente qui correspond à l'expression régulière et modifie la police de cette chaîne en gras.

Propriété	Mode de jeu de jetons	Description
Étiquette	Caractère	Définit l'étiquette appliquée par une transformation Libellé aux données qui correspondent au jeu de caractères.
Mode standard	Caractère	Active une vue d'édition simple qui inclut des champs pour la plage de début et la plage de fin.
Plage de début	Caractère	Spécifie le premier caractère dans une plage de caractères.
Plage de fin	Caractère	Spécifie le dernier caractère dans une plage de caractères. Pour des plages à caractère unique, ne remplissez pas ce champ.
Mode avancé	Caractère	Active une vue d'édition avancée dans laquelle vous pouvez saisir manuellement des plages de caractères à l'aide de caractères de plages et de caractères délimiteurs.
Caractère de plage	Caractère	Modifie de manière temporaire le symbole représentant une plage de caractères. Le caractère de plage rétablit le caractère par défaut lorsque vous fermez le jeu de caractères.
Caractère délimiteur	Caractère	Modifie de manière temporaire le symbole qui sépare les plages de caractères. Le caractère délimiteur rétablit le caractère par défaut lorsque vous fermez le jeu de caractères.

Règles et directives pour les modèles probabilistes et les modèles de classeur

Chaque modèle probabiliste et de classeur du référentiel modèle identifie un fichier dans la structure de répertoires Informatica. Les fichiers contiennent les valeurs de données et les libellés que vous ajoutez au modèle dans l'outil Developer. Les fichiers contiennent également la logique de compilation qui définit les associations entre les valeurs de données et les libellés.

Envisagez les règles et directives suivantes lorsque vous travaillez avec des modèles probabilistes ou des modèles de classeur :

- Lorsque vous exécutez un mappage qui inclut un modèle, le service d'intégration de données applique la logique du modèle compilé aux données d'entrée de la transformation. Le service d'intégration de données ne lit pas les valeurs de données ou les libellés du modèle pendant l'exécution du mappage.

- Vous pouvez éventuellement supprimer les valeurs de données et les libellés d'un modèle probabiliste ou d'un modèle de classeur. Par exemple, vous pouvez décider de supprimer des données sensibles ou privées d'un modèle. Vous pouvez supprimer les valeurs de données individuelles et les libellés dans l'outil Developer. Vous pouvez supprimer toutes les valeurs de données et les libellés lorsque vous exportez un modèle depuis le référentiel modèle.

Remarque: Si vous supprimez toutes les valeurs de données et les libellés d'un modèle, vous ne pouvez pas compiler le modèle.

- Lorsque vous supprimez une ou plusieurs valeurs de données ou des libellés d'un modèle, la logique du modèle compilé ne représente plus les données actuelles du fichier modèle. Pour synchroniser la logique du modèle, les valeurs de données et les libellés, compilez à nouveau le modèle. Ne compilez pas le modèle si vous souhaitez conserver la logique du modèle actuelle.
- Pour protéger les données dans un modèle de classeur ou un modèle probabiliste, sauvegardez le fichier de modèle dans la structure de répertoires Informatica. Sauvegardez le fichier avant de supprimer toutes les valeurs de données et les libellés d'un modèle.
- Recherchez les fichiers de modèle dans la machine hôte du service de gestion du contenu.

Par défaut, les fichiers de modèle probabiliste se trouvent à l'emplacement suivant et portent l'extension de nom de fichier suivante :

```
<Informatica_Installation_Directory>/tomcat/bin/ner/<filename>.ner
```

Par défaut, les fichiers de modèle de classeur se trouvent à l'emplacement suivant et portent l'extension de nom de fichier suivante :

```
<Informatica_Installation_Directory>/tomcat/bin/classifier/<filename>.classifier
```

- Si vous procédez à une mise à niveau de l'installation Informatica, il est possible que vous deviez compiler les modèles probabilistes et les modèles de classeur avant de les utiliser dans un mappage. Si un modèle ne contient pas de données, remplacez le fichier actuel dans la structure de répertoires Informatica par le fichier de sauvegarde qui contient les données.

Gestion des libellés dans les modèles de classeur et les modèles probabilistes

La boîte de dialogue **Gérer les libellés** vous permet de vérifier les libellés d'un modèle probabiliste ou d'un modèle de classeur et de les mettre à jour.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Ouvrez la boîte de dialogue **Gérer les libellés**.

La boîte de dialogue répertorie les libellés du modèle.

LIENS CONNEXES :

- ["Gestion des libellés du modèle de classeur" à la page 49](#)
- ["Configuration d'un modèle probabiliste" à la page 64](#)

Création d'un ensemble de contenu

Créez un jeu de contenu pour gérer des objets de données de référence qui font référence à un seul projet, type d'information ou objectif métier.

1. Dans la vue **Explorateur d'objets**, sélectionnez un projet ou un dossier dans lequel stocker le jeu de contenu.
2. Cliquez sur **Fichier > Nouveau > Ensemble de contenu**.

3. Entrez un nom pour l'ensemble de contenu.
4. Vous pouvez éventuellement sélectionner **Parcourir** pour changer l'emplacement du référentiel modèle pour l'ensemble de contenu.
5. Cliquez sur **Terminer**.

Création d'un objet de données de référence dans un ensemble de contenu

Vous pouvez créer un jeu de caractères, un ensemble de formes, un jeu de jetons, une expression régulière, un modèle probabiliste et un modèle de classeur dans un ensemble de contenu.

1. Ouvrez un ensemble de contenu dans l'éditeur et sélectionnez la vue **Contenu**.
2. Sélectionnez un type d'objet de données de référence.
3. Cliquez sur **Ajouter**.
4. Entrez un nom pour l'objet de données de référence.
Entrez éventuellement une description de l'objet.
5. Configurez les propriétés de l'objet de données de référence.
6. Cliquez sur **Terminer**.

Astuce: Vous pouvez copier des objets de données de référence d'un ensemble de contenu vers un autre. Utilisez les options **Copier dans** et **Coller depuis** pour créer une copie d'un objet dans un ensemble de contenu. Utilisez la touche **CTRL** pour sélectionner plusieurs objets de l'ensemble de contenu.

CHAPITRE 4

Modèles de classeur

Ce chapitre comprend les rubriques suivantes :

- [Présentation des modèles de classeur, 44](#)
- [Structure du modèle de classeur, 45](#)
- [Scores de classeur, 45](#)
- [Exemple de transformation Classeur, 45](#)
- [Options de modèle de classeur, 46](#)
- [Données de référence du modèle de classeur, 48](#)
- [Données de libellé du modèle de classeur, 48](#)
- [Configuration d'un modèle de classeur, 50](#)
- [Opérations de filtre et de recherche, 54](#)
- [Opérations de copier/coller, 55](#)

Présentation des modèles de classeur

Un modèle de classeur est un objet de données de référence dans un ensemble de contenu. Utilisez un modèle de classeur pour analyser les chaînes de texte longues qui contiennent plusieurs valeurs. Un modèle de classeur identifie le type d'informations le plus courant dans chaque chaîne.

Ajoutez un modèle de classeur à une transformation Classeur. La transformation recherche les valeurs communes des données du modèle de classeur et des données de chaque ligne d'entrée. La transformation utilise ces valeurs communes pour classer le type d'informations que représente chaque ligne.

Vous utilisez un modèle de classeur lorsque les données d'entrée présentent les caractéristiques suivantes :

- Les données d'entrée contiennent du texte. Les modèles de classeur appliquent des processus en langage naturel aux données textuelles pour identifier les types d'informations contenus dans le texte. Les processus en langage naturel détectent les mots pertinents contenus dans la chaîne d'entrée. Les processus en langage naturel ignorent les mots non pertinents.
- Les chaînes de données d'entrée contiennent plusieurs valeurs. Par exemple, vous pouvez créer une colonne de données qui inclut le contenu d'un e-mail dans chaque champ.

La transformation Classeur lit les types de données Chaîne. La transformation n'impose aucune limite quant à la longueur des chaînes d'entrée.

Vous devez compiler les modèles de classeur dans l'outil Developer. Lorsque vous compilez un modèle, vous créez des associations entre des valeurs de données similaires dans le modèle. La transformation Classeur utilise les données compilées pour rechercher des informations dans les données d'entrée.

Structure du modèle de classeur

Un modèle de classeur contient des valeurs de données de référence et des valeurs de libellés. Les valeurs de données de référence représentent les données à classer. Les valeurs de libellés spécifient les types d'informations qu'une transformation Classeur peut identifier dans les données.

Un modèle de classeur contient également des données de compilation. Lorsqu'une transformation Classeur compare les données d'entrée aux données du modèle, elle renvoie les valeurs de libellés qui décrivent chaque ligne des données d'entrée. Lors de la compilation d'un modèle de classeur, vous créez ou mettez à jour les données de compilation. Lorsqu'une transformation Classeur compare les données d'entrée aux données du modèle, la transformation renvoie les valeurs de libellés qui décrivent chaque ligne des données d'entrée.

L'outil Developer écrit les valeurs de données de référence, les valeurs de libellés et les données de compilation dans un fichier de la structure de répertoire d'Informatica. L'objet de modèle de classeur du référentiel modèle stocke le nom de fichier. Lors de l'enregistrement d'un modèle de classeur, écrivez dans le fichier les valeurs de données de référence actuelles et les valeurs de libellés. Lors de la compilation du modèle, mettez à jour les données de compilation dans le fichier. Vous pouvez lire le nom de fichier depuis les propriétés de modèle de l'outil Developer.

Scores de classeur

Une transformation Classeur compare chaque ligne de données d'entrée à chaque ligne de données de référence dans un modèle de classeur. La transformation calcule un score pour chaque comparaison. Les scores représentent les degrés de similarité entre la ligne d'entrée et les lignes de données de référence.

Lorsque vous exécutez un mappage qui contient une transformation Classeur, le mappage renvoie le libellé qui identifie la ligne des données de référence dont le score est le plus élevé. Le score est compris entre 0 et 1. Un score élevé indique une correspondance élevée entre les données d'entrée et les données du modèle.

Vérifiez les scores du classeur pour vous assurer que la sortie du libellé décrit précisément chaque ligne de données d'entrée. Vous pouvez également vérifier les scores pour vous assurer que le modèle de classeur est approprié pour les données d'entrée. Si la sortie de transformation contient un important pourcentage de scores bas, le modèle de classeur est peut-être inapproprié. Pour améliorer les comparaisons, compilez à nouveau le modèle. Si le modèle compilé n'améliore pas les scores, remplacez le modèle dans la transformation.

Exemple de transformation Classeur

Vous pouvez utiliser un modèle de classeur et une transformation Classeur pour classer les e-mails en fonction du texte qu'ils contiennent.

Par exemple, vous êtes gestionnaire de données dans le centre de support client d'un fabricant de logiciel. Vous êtes chargé de vérifier les courriels envoyés par les clients au centre de support. L'organisation compte des clients dans de nombreux pays et le centre de support reçoit des courriels dans de nombreuses langues. Vous voulez trier les courriels par langue, de façon à pouvoir envoyer chaque courriel au service le plus à même de répondre au client.

Pour ce faire, procédez comme suit :

1. Écrivez les courriels dans un seul fichier ou une seule table de base de données.
2. Créez un objet de données dans le référentiel modèle capable de lire le fichier ou la table de base de données.
3. Créez des objets de données dans le référentiel modèle pour chaque langue utilisée dans un message.
4. Créez un modèle de classeur contenant un exemple de texte pour chaque langue.

Remarque: Vous pouvez utiliser des échantillons de données provenant des courriels en tant que données source du modèle.

5. Ajoutez le modèle de classeur à une transformation Classeur réutilisable.
6. Configurez un mappage afin d'appliquer la transformation Classeur aux données de message.

Pour ce faire, procédez comme suit :

- Ajoutez la transformation Classeur et les objets de données au mappage.
- Connectez le port d'entrée d'une transformation Classeur à l'objet de données source.
- Connectez les ports de sortie de la transformation Classeur aux objets de données cibles.

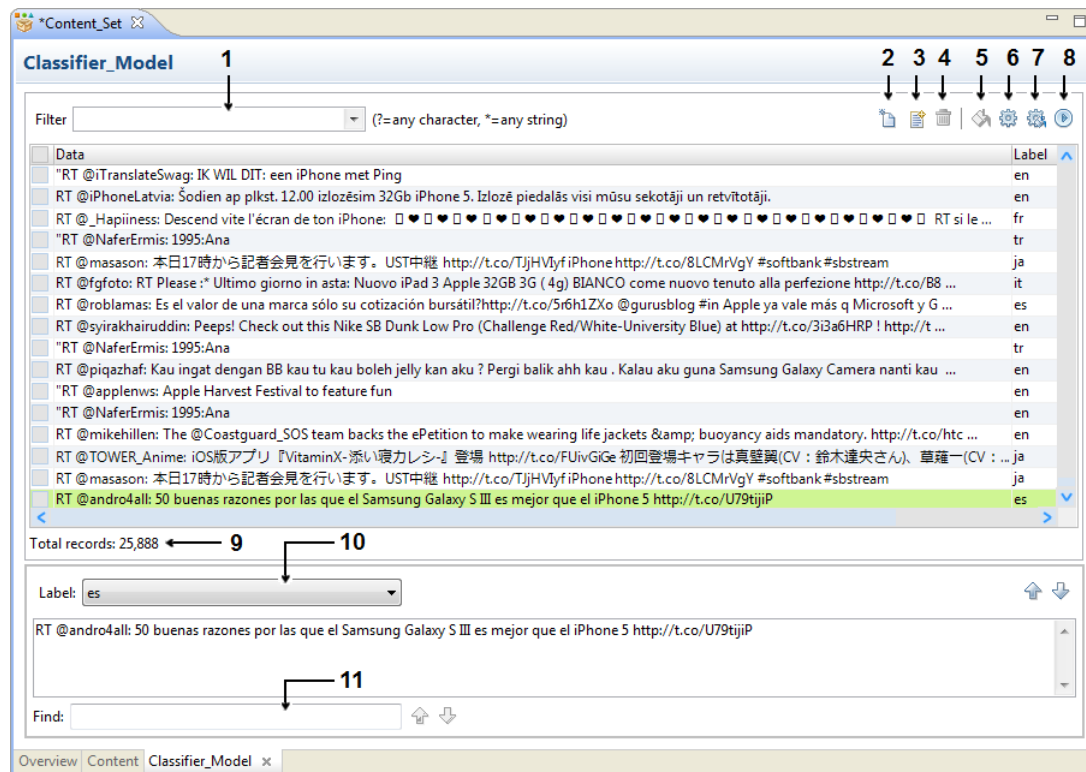
Lorsque vous exécutez le mappage, la transformation Classeur analyse les e-mails et écrit leur texte dans la cible de données appropriée. Vous pouvez partager les données cibles avec les membres de l'équipe de chaque service.

Options de modèle de classeur

L'outil Developer affiche les données de modèle de classeur dans un éditeur disposant d'un volet supérieur et d'un volet inférieur. Le volet supérieur affiche toutes les lignes de données de référence et tous les libellés que vous attribuez aux données. Le volet inférieur affiche le contenu de la ligne que vous sélectionnez.

Le volet supérieur vous permet de vérifier les lignes de données de référence et d'identifier toutes les lignes qui n'utilisent pas de libellé. Le volet inférieur vous permet de vérifier le contenu d'une ligne et d'attribuer un libellé à cette ligne. Le volet supérieur affiche environ 100 caractères de données sur chaque ligne. Le volet inférieur affiche toutes les données de la ligne que vous sélectionnez.

L'image suivante illustre l'éditeur du modèle de classeur :



L'éditeur comprend les options suivantes :

1. Champ Filtre
Permet de filtrer la liste des lignes de données de référence en fonction des valeurs de données ou du libellé que vous spécifiez.
2. Ajouter une ligne
Insère une ligne de données de référence vide.
3. Adjonction de données
Importe des données dans le référentiel modèle à partir d'un objet de données.
4. Supprimer
Supprime les lignes de données de référence que vous sélectionnez. Utilisez les cases à cocher pour sélectionner les lignes.
5. Attribuer un libellé
Attribue un libellé à une ou plusieurs lignes de données de référence que vous sélectionnez. Utilisez les cases à cocher pour sélectionner les lignes.
6. Modifier les propriétés
Affiche les propriétés du modèle de classeur.
7. Gérer les libellés
Ouvre la boîte de dialogue **Gérer les libellés** qui vous permet d'ajouter ou de supprimer des valeurs de libellés dans le modèle de classeur.
8. Compiler
Compile le modèle de classeur.

9. Nombre total d'enregistrements

Indique le nombre de lignes de données de référence dans le modèle de classeur.

10. Champ Libellé

Affiche une valeur de libellé que vous pouvez appliquer à la ligne de données de référence actuelle.

11. Champ Rechercher

Permet de trouver la valeur de données que vous spécifiez dans la ligne de données de référence en cours.

Données de référence du modèle de classeur

Un modèle de classeur contient une colonne de données de référence qui peut inclure des phrases, des paragraphes ou des pages de texte. Les données de référence représentent les différents types d'entrée de texte qu'une transformation Classeur peut lire dans un mappage. Lorsque vous créez un modèle, vérifiez que les données de référence comprennent les types de textes que vous vous attendez à rencontrer lorsque vous exécutez le mappage.

Vous pouvez utiliser les données source de mappage pour créer un modèle de classeur. Sélectionnez un échantillon de données source et copiez-le dans le modèle.

Tenez compte des règles et directives suivantes lorsque vous utilisez des données de référence de modèle de classeur :

- Un champ de données de référence peut être de n'importe quelle longueur. Vous pouvez entrer des pages de texte dans chaque champ de données.
- Vous devez importer des données de référence à partir d'un objet de données.
- Vous ne pouvez pas éditer des valeurs de données de référence. Cependant, vous pouvez supprimer une ligne de données.
- Lorsque vous compilez un modèle de classeur, le processus de compilation ignore toutes les valeurs de nombre dans les données de référence.

Données de libellé du modèle de classeur

Un modèle de classeur contient un ou plusieurs libellés descriptifs qui résument les types d'informations se trouvant dans les lignes de données de référence. Attribuez un libellé à chaque ligne de données de référence.

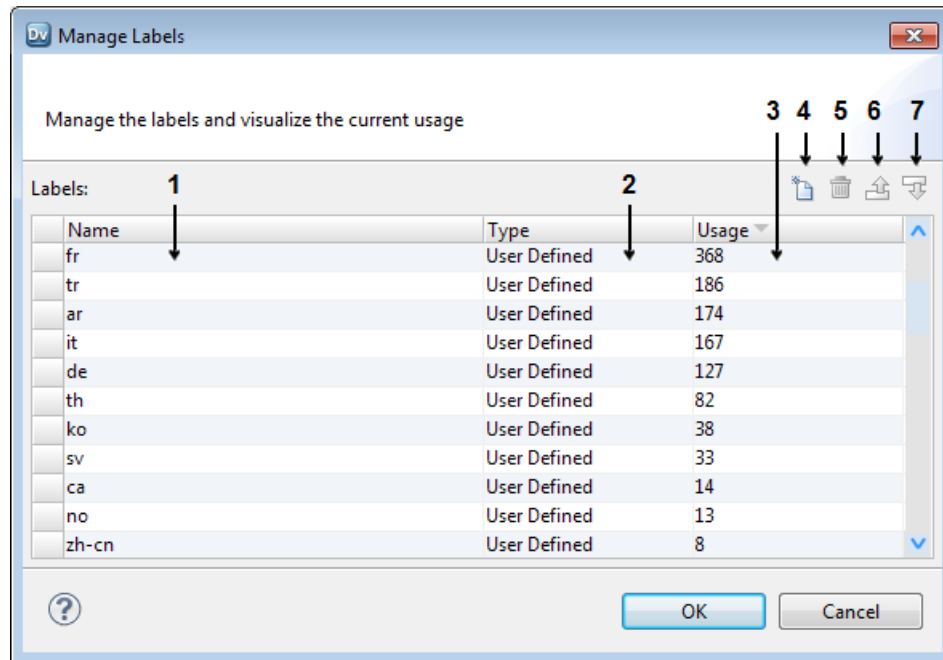
Lorsque vous ajoutez à un modèle de classeur des données issues d'une source de données, vous pouvez spécifier une colonne en tant que colonne de données de libellé. Vous pouvez également créer des libellés dans le modèle.

Les libellés sont indépendants des valeurs de données de référence qu'ils décrivent. Si vous supprimez les lignes de données de référence qui utilisent un libellé, vous ne supprimez pas le libellé du modèle. Si vous supprimez un libellé, vous ne supprimez pas les valeurs de données de référence que vous avez associées au libellé.

Gestion des libellés du modèle de classeur

La boîte de dialogue **Gérer les libellés** vous permet de vérifier et de mettre à jour les valeurs de libellés dans un modèle de classeur. Vous pouvez également trier et mettre à jour les valeurs de libellés.

L'image suivante représente la boîte de dialogue **Gérer les libellés** :



La boîte de dialogue **Gérer les libellés** contient les éléments suivants :

1. Colonne Nom.
Contient les valeurs de libellés que la transformation Classeur peut appliquer aux lignes de données d'entrée. Vous pouvez trier les libellés par nom.
2. Colonne Type.
Identifie la source des valeurs de libellés. Le modèle de classeur identifie tous les libellés comme étant des valeurs définies par l'utilisateur.
3. Colonne Utilisation.
Indique le nombre de lignes de données de référence qui utilisent chaque libellé. Vous pouvez trier les libellés en fonction du nombre de lignes.
4. Bouton Ajouter.
Ajoute un libellé au modèle de classeur. Entrez une valeur de libellé dans la colonne Nom de la ligne.
Remarque: pour mettre à jour une valeur de libellé, double-cliquez dessus et entrez celle dont vous avez besoin.
5. Bouton Supprimer.
Supprime un libellé depuis le modèle de classeur.
6. Flèche vers le haut.
Déplace le libellé d'une ligne vers le haut dans la boîte de dialogue.
7. Flèche vers le bas.
Déplace le libellé d'une ligne vers le bas dans la boîte de dialogue.

Configuration d'un modèle de classeur

Le processus permettant de configurer un modèle de classeur commence par les données à classer. Le contenu des données de référence que vous ajoutez au modèle doit indiquer les données que vous connectez à la transformation Classeur. La transformation compare les valeurs de données et les modèles de données d'entrée aux valeurs de données et aux modèles du modèle de classeur.

Pour créer un modèle de classeur à utiliser dans une transformation Classeur, procédez comme suit :

1. Identifiez les valeurs de données de référence et les valeurs de libellés à ajouter au modèle.
Vous pouvez utiliser une partie des données à classer. Créez un objet de données dans le référentiel modèle capable de lire le fragment de données.
2. Créez un ensemble de contenu et ajoutez-lui un modèle de classeur.
3. Ajoutez les valeurs de données de référence au modèle.
4. Ajoutez les valeurs de libellés au modèle.
Vous pouvez importer les données à partir de l'objet de données dans le référentiel modèle. Vous pouvez également entrer une ligne de données de référence ou un libellé.
5. Attribuez un libellé à chaque ligne de données de référence.
Vous pouvez attribuer un libellé à plusieurs lignes en une seule opération.
6. Compilez le modèle.

Après avoir compilé le modèle de classeur, vous pouvez l'utiliser dans une transformation Classeur.

Création d'un modèle de classeur

Utilisez un objet de données en tant que source de données du modèle de classeur.

Un modèle de classeur fonctionne de manière optimale lorsque vous utilisez les données d'entrée de la transformation Classeur en tant que source de données de référence du modèle.

1. Dans l'explorateur d'objets, ouvrez et créez un ensemble de contenu.
2. Sélectionnez la vue **Contenu**.
3. Sélectionnez **Modèles de classeur**, puis cliquez sur **Ajouter**.
L'assistant Modèle de classeur s'ouvre.
4. Saisissez un nom pour le modèle de classeur.
Entrez éventuellement une description textuelle du modèle.
5. Parcourez le référentiel modèle et sélectionnez l'objet de données qui contient les données à importer.
Ne sélectionnez pas d'objet de données de média social.
Cliquez sur **Suivant**.
6. Vérifiez les colonnes de l'objet de données et sélectionnez une ou plusieurs colonnes à ajouter au modèle. Vous pouvez ajouter des colonnes de données de référence et une colonne de libellé au cours de la même opération.
 - Pour importer une colonne de données comme données de référence, sélectionnez le nom de la colonne et cliquez sur **Données**.
Vous pouvez sélectionner plusieurs colonnes de données. L'outil Developer fusionne le contenu des colonnes que vous sélectionnez en une seule colonne.
 - Pour importer une colonne de données en tant que valeurs de libellés, sélectionnez le nom de la colonne et cliquez sur **Libellé**.

Lors de l'importation des données de référence et des valeurs de libellés, l'outil Developer attribue le libellé de chaque ligne à la chaîne de données de référence sur la même ligne. Vous pouvez prévisualiser les données avant de sélectionner les colonnes. Vous pouvez modifier les attributions de libellé après avoir créé le modèle.

Cliquez sur **Suivant**.

7. Sélectionnez le nombre de lignes à importer à partir de la source de données.

Par défaut, l'outil Developer importe toutes les lignes depuis la source de données. Si vous saisissez un nombre, le modèle compte les lignes depuis le début de l'ensemble de données.

8. Cliquez sur **Terminer** et enregistrez le modèle.

Après avoir créé le modèle, vérifiez les attributions de libellé et compilez le modèle.

Ajout de données à un modèle de classeur à partir d'une source de données

Vous pouvez importer dans un modèle de classeur plusieurs lignes de valeurs de données de référence ou des valeurs de libellés en une seule opération.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.

2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.

3. Cliquez sur **Adjonction de données**.

L'assistant Modèle de classeur s'ouvre.

4. Parcourez le référentiel modèle et sélectionnez l'objet de données qui contient les données à importer.

Ne sélectionnez pas d'objet de données de média social.

Cliquez sur **Suivant**.

5. Vérifiez les colonnes de l'objet de données et sélectionnez une ou plusieurs colonnes à ajouter au modèle. Vous pouvez ajouter des colonnes de données de référence et une colonne de libellé au cours de la même opération.

- Pour importer une colonne de données comme données de référence, sélectionnez le nom de la colonne et cliquez sur **Données**.

Vous pouvez sélectionner plusieurs colonnes de données. L'outil Developer fusionne le contenu des colonnes que vous sélectionnez en une seule colonne.

- Pour importer une colonne de données en tant que valeurs de libellés, sélectionnez le nom de la colonne et cliquez sur **Libellé**.

Lors de l'importation des données de référence et des valeurs de libellés, l'outil Developer attribue le libellé de chaque ligne à la chaîne de données de référence sur la même ligne. Vous pouvez prévisualiser les données avant de sélectionner les colonnes. Vous pouvez modifier les attributions de libellé après avoir créé le modèle.

Cliquez sur **Suivant**.

6. Sélectionnez le nombre de lignes à importer à partir de la source de données.

Par défaut, l'outil Developer importe toutes les lignes depuis la source de données. Si vous saisissez un nombre, le modèle compte les lignes depuis le début de l'ensemble de données.

7. Cliquez sur **Terminer** et enregistrez le modèle.

Ajout d'une ligne de données de référence à un modèle de classeur

Vous pouvez ajouter une ligne de données de référence à un modèle de classeur.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Cliquez sur **Ajouter une ligne**.

L'outil Developer ajoute une ligne en dessous de la ligne actuelle dans les données de référence.

4. Entrez les valeurs de données de référence sur la ligne.

Vous pouvez utiliser les raccourcis Windows pour coller les données sur la ligne.

Ajout d'un libellé à un modèle de classeur

Vous pouvez ajouter un libellé à un modèle de classeur.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Ouvrez la boîte de dialogue **Gérer les libellés**.

La boîte de dialogue répertorie les libellés du modèle.

4. Cliquez sur **Nouveau**.

L'outil Developer ajoute une ligne en bas de la liste des libellés.

5. Double-cliquez sur la valeur par défaut de la colonne Nom et entrez un nom de libellé.

6. Cliquez sur **OK**.

Après avoir créé le libellé, vous pouvez l'attribuer à une ou plusieurs lignes des données de référence. La colonne Utilisation de la boîte de dialogue **Gérer les libellés** indique le nombre de lignes qui utilisent le libellé.

Attribution d'un libellé à des lignes de données de référence

Vous pouvez attribuer un libellé à une ou plusieurs lignes de données de référence en une seule opération.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Sélectionnez une ou plusieurs lignes de données de référence. Utilisez les cases à cocher pour sélectionner les lignes.

Remarque: Vous pouvez utiliser l'option de filtre pour afficher toutes les lignes contenant la valeur de données spécifiée. Cochez la case Tout sélectionner pour sélectionner toutes les lignes contenant cette valeur.

4. Cliquez sur **Attribuer un libellé**.

L'outil Developer affiche la liste des libellés dans le modèle de classeur.

5. Sélectionnez une valeur de libellé, puis cliquez sur **Attribuer**.

L'outil Developer met à jour les lignes de données de référence que vous avez sélectionnées avec la valeur de libellé.

Vous pouvez éventuellement compiler le modèle de façon à ajouter les noms des libellés à la logique du modèle de classeur.

Identification de valeurs de libellés inutilisées

La boîte de dialogue **Gérer les libellés** vous permet de trouver une valeur de libellé inutilisée dans le modèle de classeur. La boîte de dialogue **Gérer les libellés** affiche les données d'utilisation des valeurs de libellés dans le modèle de classeur. Les données d'utilisation vous permettent de vérifier le nombre de lignes de données de référence qui utilisent une valeur de libellé et de trouver des valeurs de libellés inutilisées.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Ouvrez la boîte de dialogue **Gérer les libellés**. La boîte de dialogue répertorie les libellés du modèle de classeur.
4. Vérifiez les données de la colonne Utilisation de chaque libellé.

La colonne Utilisation répertorie le nombre de lignes de données de référence qui utilisent le libellé. Si une valeur de libellé est inutilisée, la colonne Utilisation affiche une valeur de zéro.

Suppression de ligne d'un modèle de classeur

Vous pouvez supprimer une ou plusieurs lignes de données de référence d'un modèle de classeur en une seule opération.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Sélectionnez une ou plusieurs lignes de données de référence. Utilisez les cases à cocher pour sélectionner les lignes.
4. Cliquez sur **Supprimer**.

L'outil Developer supprime les lignes que vous avez sélectionnées dans le modèle de classeur.

Pour annuler l'opération, appuyez sur les touches Ctrl + Z.

Suppression d'un libellé d'un modèle de classeur

Utilisez la boîte de dialogue **Gérer les libellés** pour supprimer un libellé d'un modèle de classeur.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Ouvrez la boîte de dialogue **Gérer les libellés**.
4. Cliquez sur **Supprimer**.
5. Cliquez sur **Oui** pour confirmer l'opération.

L'outil Developer supprime le libellé du modèle. L'outil Developer ne supprime pas de ligne de données de référence qui utilise le libellé.

6. Cliquez sur **OK** pour fermer la boîte de dialogue.

Pour annuler l'opération, appuyez sur les touches Ctrl + Z.

Compilation d'un modèle de classeur

À chaque fois que vous éditez une valeur de libellé ou la valeur de données de référence dans un modèle de classeur, vous devez compiler le modèle. Lorsque vous compilez le modèle, vous mettez à jour les données de compilation dans le modèle.

- Pour mettre à jour les données de compilation, ouvrez le modèle dans l'outil Developer et cliquez sur **Compiler**.

Opérations de filtre et de recherche

Les options de filtre vous permettent d'afficher ou de masquer les lignes de données de référence qui répondent aux critères spécifiés. Lors de l'application d'un filtre, vous pouvez effectuer des actions supplémentaires sur les lignes de données que le modèle de classeur affiche. Par exemple, vous pouvez appliquer une valeur de libellé à toutes les lignes de données.

L'option de filtre vous permet d'effectuer les tâches suivantes :

- Rechercher les lignes de données de référence qui contiennent une valeur que vous entrez.
- Rechercher les lignes de données de référence qui utilisent un libellé que vous sélectionnez.
- Rechercher les lignes de données de référence qui n'utilisent pas de libellé.

Vous pouvez également rechercher une valeur de données dans une ligne de données de référence.

Utilisation d'une valeur de données afin de filtrer les lignes de données de référence

Le filtre vous permet de vérifier qu'une ou plusieurs lignes de données de référence contiennent les valeurs de données que vous attendez.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Entrez une valeur dans le champ Filtre.

Vous pouvez inclure des caractères génériques.

L'outil Developer affiche les lignes de données de référence qui contiennent le texte du filtre.

Utilisation d'une valeur de libellé de façon à filtrer les lignes de données de référence

Le filtre vous permet d'afficher ou de masquer les lignes de données de référence qui utilisent le libellé sélectionné.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Sélectionnez une valeur de libellé depuis le menu de filtre.

L'outil Developer affiche les lignes de données de référence qui utilisent la valeur de libellé.

Remarque: pour rechercher une ligne de données de référence qui n'utilise pas de libellé, sélectionnez l'option **Aucun libellé** dans le menu de filtre.

Recherche d'une valeur dans une ligne de données de référence

Le champ Rechercher vous permet de rechercher une valeur de données dans une ligne que vous sélectionnez.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Sélectionnez une ligne de données de référence.
4. Entrez une valeur de données dans le champ Rechercher.

Le modèle met en surbrillance les instances de la valeur dans la ligne de données de référence.

5. Utilisez les flèches vers le haut ou vers le bas pour rechercher des instances supplémentaires de la valeur dans la ligne.

Opérations de copier/coller

Vous pouvez copier un modèle de classeur d'un ensemble de contenu vers un autre dans un référentiel modèle. Copiez un modèle de classeur pour partager des ressources avec d'autres utilisateurs de l'outil Developer.

Vous pouvez copier un modèle vers un autre ensemble de contenu ou importer un modèle dans l'ensemble de contenu actuel. Vous pouvez importer plusieurs modèles de plusieurs ensembles de contenu dans le référentiel en une seule fois.

Lorsque vous copiez un modèle, le service de gestion du contenu crée une copie du fichier de données du modèle sur la machine du service. Chaque modèle utilise un fichier de données différent.

Copie d'un modèle de classeur dans un autre ensemble de contenu

Vous pouvez copier un modèle de classeur d'un ensemble de contenu vers un autre dans un référentiel modèle. Lorsque vous copiez un modèle de classeur, vous spécifiez l'objet de modèle et les ensembles de contenu source et de destination.

1. Ouvrez l'ensemble de contenu qui contient le modèle de classeur.
2. Sélectionnez un modèle de classeur et cliquez sur **Copier dans**.
3. Parcourez le référentiel modèle et sélectionnez un ensemble de contenu.

Vous pouvez copier le modèle de classeur vers un ensemble de contenu dans le projet actuel ou dans un autre projet.

4. Cliquez sur **OK**.

L'outil Developer copie le modèle de classeur dans l'ensemble de contenu sélectionné.

Importation d'un modèle de classeur depuis un autre ensemble de contenu

Vous pouvez importer un modèle de classeur d'un ensemble de contenu vers un autre dans un référentiel modèle. Lorsque vous importez un modèle de classeur, vous spécifiez un ou plusieurs objets de modèle et les ensembles de contenu source et de destination.

1. Ouvrez l'ensemble de contenu qui doit contenir le modèle de classeur.
2. Sélectionnez un modèle de classeur et cliquez sur **Coller depuis**.
3. Parcourez le référentiel modèle et sélectionnez un modèle de classeur.

Vous pouvez coller le modèle de classeur d'un ensemble de contenu dans le projet actuel ou un autre projet.

4. Cliquez sur **OK**.

L'outil Developer colle le modèle de classeur dans l'ensemble de contenu actuel.

CHAPITRE 5

Modèles probabilistes

Ce chapitre comprend les rubriques suivantes :

- [Présentation des modèles probabilistes, 56](#)
- [Structure du modèle probabiliste, 57](#)
- [Exemple de transformation Libellé, 58](#)
- [Exemple de transformation Analyseur, 59](#)
- [Options de modèle probabiliste, 59](#)
- [Données de référence du modèle probabiliste, 62](#)
- [Données de libellé du modèle probabiliste, 62](#)
- [Propriétés du modèle probabiliste, 64](#)
- [Configuration d'un modèle probabiliste, 64](#)
- [Opérations de copier/coller, 71](#)

Présentation des modèles probabilistes

Un modèle probabiliste est un objet de données de référence que vous créez dans un ensemble de contenu. Utilisez un modèle probabiliste pour analyser une chaîne de données qui contient plusieurs valeurs de données. Un modèle probabiliste identifie les types d'informations présents dans chaque valeur de la chaîne. Vous pouvez ajouter un modèle probabiliste à une transformation Libellé et une transformation Analyseur.

Utilisez un modèle probabiliste dans une transformation Libellé pour attribuer un libellé descriptif à chaque valeur d'une chaîne d'entrée. La transformation Libellé écrit les libellés sur un seul port de sortie. Utilisez un modèle probabiliste dans une transformation Analyseur pour écrire chaque valeur d'une chaîne d'entrée sur un port qui représente les informations que contient la valeur. La transformation Analyseur crée un port de sortie pour chaque type d'information.

Vous concevez et compilez un modèle probabiliste dans l'outil Developer. Lorsque vous définissez un modèle probabiliste, vous ajoutez une série de lignes de données au modèle et vous attribuez un libellé à chaque valeur de chaque ligne. Lorsque vous compilez un modèle probabiliste, l'outil Developer crée des associations entre les valeurs de données et les libellés que vous avez ajoutés. La transformation Libellé et la transformation Analyseur utilisent des processus en langage naturel pour comparer les données du modèle probabiliste et les données du port d'entrée.

Les processus en langage naturel utilisent les techniques suivantes pour identifier les types d'informations qui figurent dans les valeurs de données :

- Les processus en langage naturel peuvent reconnaître les valeurs de données similaires et leur attribuer le même libellé.
- Les processus en langage naturel peuvent comparer une valeur de données aux valeurs adjacentes dans la chaîne. Les processus en langage naturel analysent la séquence de valeurs pour comprendre l'utilisation de chaque chaîne et vérifier les types d'informations que les chaînes représentent.

Structure du modèle probabiliste

Un modèle probabiliste contient des lignes de valeurs de données de référence et des valeurs de libellés. Les valeurs de données de référence représentent les différentes valeurs qui peuvent s'afficher dans les données d'entrée de la transformation. Les valeurs de libellés identifient les types d'informations auxquels vous vous attendez dans les données d'entrée.

Un modèle probabiliste contient également des données de compilation. Les transformations Libellé et Analyseur utilisent les données de compilation afin de comparer les similarités entre les données de référence du modèle et les données d'entrée de la transformation. Lors de la compilation d'un modèle probabiliste, vous créez ou mettez à jour les données de compilation.

Une ligne de données peut contenir une ou plusieurs valeurs. Chaque ligne de données peut présenter une structure différente. Vous pouvez attribuer le même libellé à différentes valeurs dans une ligne de données. Par ailleurs, vous pouvez attribuer un libellé différent à des valeurs identiques qui figurent sur différentes positions d'une ligne. Le service d'intégration de données prend en considération les positions relatives des valeurs dans la chaîne d'entrée lorsque le mappage s'exécute. Affectez chaque libellé à une valeur de données au moins avant de compiler le modèle probabiliste.

L'outil Developer écrit les valeurs de données de référence, les valeurs de libellés et les données de compilation dans un fichier de la structure de répertoire d'Informatica. L'objet de modèle probabiliste du référentiel modèle stocke le nom de fichier. Lors de l'enregistrement d'un modèle probabiliste, vous devez écrire dans le fichier les valeurs de données de référence actuelles et les valeurs de libellés. Lors de la compilation du modèle, mettez à jour les données de compilation dans le fichier. Vous pouvez lire le nom de fichier depuis les propriétés de modèle de l'outil Developer.

Remarque: Pour optimiser les fonctionnalités du modèle probabiliste, vérifiez que chaque ligne de données contient plusieurs valeurs de données de référence. L'ordre des valeurs de chaque ligne doit correspondre autant que possible à l'ordre dans lequel les valeurs s'affichent dans les données d'entrée de la transformation. Si les lignes de données contiennent des valeurs de données de référence, les transformations Libellé et Analyseur ne peuvent pas appliquer les processus en langage naturel pendant l'analyse probabiliste.

Exemple de transformation Libellé

La base de données client d'une compagnie d'assurance contient plusieurs erreurs d'entrée de données. Vous êtes gestionnaire des données dans cette compagnie d'assurance. Vous configurez un mappage avec une transformation Libellé pour déterminer les différents types de données que contient chaque colonne.

Le tableau suivant décrit des exemples de données provenant de la base de données client :

ID ligne	Champ 1	Champ 2	Champ 3
1	19132954	AIM SECURITIES	PETRIE TAYBRO
2	10110169	JASE TRAPANI	BANK OF NEW YORK
3	10111786	WANGER ASSET MANAGEMENT, LLP	JAN SEEDORF
4	10112299	FELIX LEVINGER	HARVARD MAGAZINE
5	10112036	DESCHÊNES & FILS LTÉE (QUEBEC)	RICHARD TREMBLAY
6	BERGER ASSOCIATES	10111101	DAREEN HULSMAN
7	19131385	EAGLE FINANCIAL GROUP INC	PATRICK MCKINNIE
8	LAKENYA PASKETT	WHITEHALL FINANCIAL GROUP	15954710

Lorsque vous exécutez le mappage, la transformation Libellé compare les données d'entrée aux données de référence du modèle probabiliste. La transformation Libellé sélectionne un libellé pour les données de chaque port d'entrée. La transformation écrit les libellés sur un port de sortie. Chaque ligne de sortie contient un ensemble de libellés qui définit la structure de données sur la ligne d'entrée correspondante.

Le tableau suivant décrit les libellés que la transformation Libellé ajoute au port de sortie :

ID ligne	Libellés de sortie
1	numéro du contact de l'organisation
2	contact de l'organisation, numéro
3	numéro du contact de l'organisation
4	contact de l'organisation, numéro
5	numéro du contact de l'organisation
6	organisation, numéro du contact
7	organisation, numéro du contact
8	numéro du contact, organisation

Exemple de transformation Analyseur

Un supermarché stocke des descriptions de produits dans une colonne d'une table de base de données. Ces descriptions contiennent plusieurs valeurs de données qui représentent différents types d'informations. Vous êtes gestionnaire des données dans ce supermarché. Vous voulez créer des colonnes pour les différents types d'informations figurant dans les descriptions des produits.

Vous configurez un mappage avec une transformation Analyseur pour organiser les valeurs de données dans les champs appropriés.

Le fragment de donnée suivant contient la description du produit « Jus d'orange » :

Jus d'orange Sunnydream sans sucre 350 ml

Le tableau suivant décrit les données de sortie que la transformation Analyseur crée à partir des données d'entrée :

Nom du produit	Type de produit	Détails du produit	Taille du produit
Sunnydream	Jus d'orange	Sans sucre	350 ml

Options de modèle probabiliste

Lors de la modification d'un modèle probabiliste, vous pouvez travailler dans les vues Données ou Libellé.

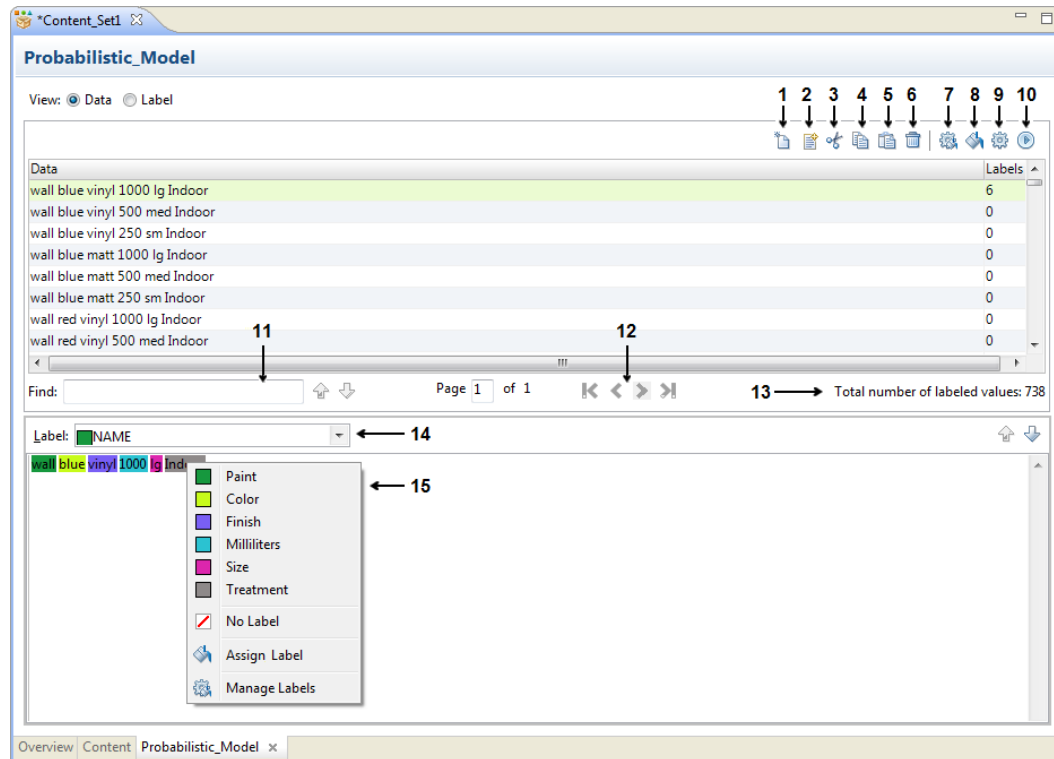
La vue Données vous permet d'ajouter des lignes de données de référence au modèle et d'attribuer des libellés aux valeurs de données de chaque ligne. La vue Libellé vous permet de vérifier les informations détaillées sur l'utilisation des valeurs de libellés dans le modèle. Vous pouvez ajouter des libellés au modèle probabiliste dans les vues Données et Libellé.

Vue Données du modèle probabiliste

La vue Données affiche les lignes de données de référence du modèle probabiliste et le nombre de valeurs de libellés que vous attribuez à chaque ligne. La vue Données affiche également le nombre total de libellés que vous avez attribués aux valeurs dans le modèle actuel.

Lorsque vous sélectionnez une ligne de données de référence, les valeurs de la ligne s'affichent dans un éditeur sous le champ Rechercher. Pour attribuer un libellé à une valeur de données de référence d'une ligne, dans l'éditeur, cliquez sur la valeur avec le bouton droit de la souris et sélectionnez une valeur de libellé.

L'image suivante illustre les options du modèle probabiliste que vous pouvez utiliser lorsque vous sélectionnez la vue Données :



La vue Données comprend les options suivantes :

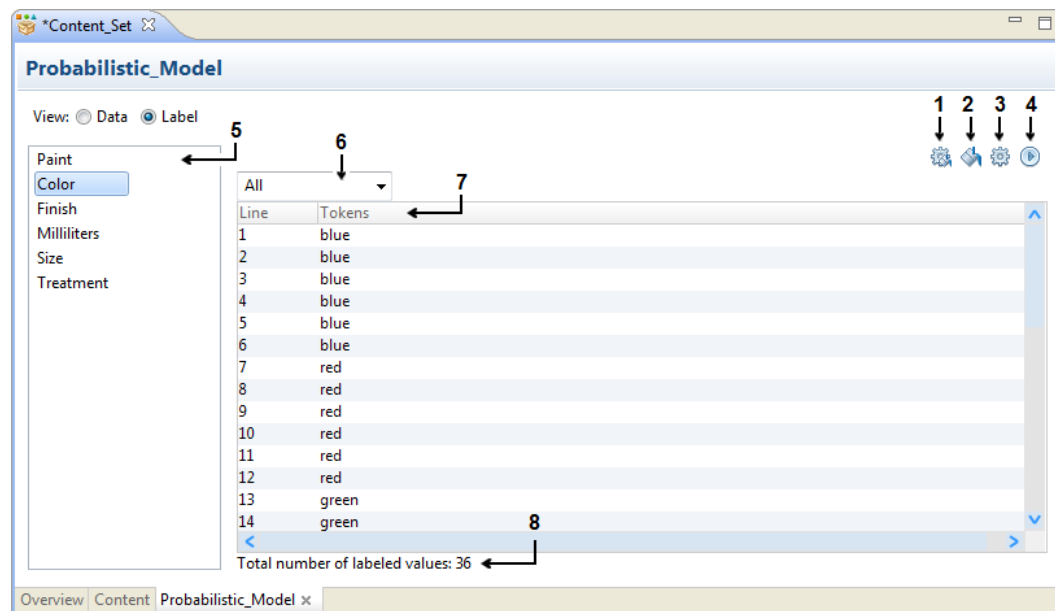
1. Ajouter une ligne
Insère une ligne de données vide.
2. Ajouter des données
Importe des données dans le référentiel modèle à partir d'un objet de données.
3. Couper
Supprime une ligne de données du modèle probabiliste et l'ajoute au presse-papiers.
4. Copier
Copie une ligne de données dans le presse-papiers.
5. Coller
Colle une ligne de données du presse-papiers dans le modèle probabiliste.
6. Supprimer
Supprime une ligne de données du modèle probabiliste.
7. Gérer les libellés
Ouvre la boîte de dialogue **Gérer les libellés** qui vous permet d'ajouter ou de supprimer des valeurs de libellés du modèle probabiliste.
8. Attribuer un libellé
Attribue un libellé à une ou plusieurs valeurs de données de référence que vous sélectionnez. Cette option vous permet d'attribuer un libellé à toutes les instances d'une valeur de données de référence dans le modèle.

9. Modifier les propriétés
Affiche les propriétés du modèle probabiliste.
10. Compiler
Compile le modèle probabiliste.
11. Champ Rechercher
Trouve des lignes dans le modèle qui contiennent la valeur de données de référence que vous entrez. Les flèches vers le haut et vers le bas vous permettent d'accéder aux lignes qui contiennent la valeur.
12. Flèches Précédent et Suivant
Permettent le déplacement vers les lignes de valeurs de données précédentes et suivantes dans le modèle.
13. Nombre total de valeurs avec libellés
Indique le nombre de valeurs de données de référence qui utilisent un libellé.
14. Champ Libellé
Affiche une valeur de libellé que vous pouvez appliquer à la valeur de données de référence sélectionnée.
15. Menu Libellé
Affiche une liste d'options que vous pouvez utiliser afin d'attribuer un libellé à une ou plusieurs valeurs de données de référence. Pour ouvrir le menu, dans l'éditeur de données de référence, cliquez sur une valeur de données de référence avec le bouton droit de la souris.

Vue Libellé du modèle probabiliste

La vue Libellé répertorie les libellés que vous définissez dans le modèle probabiliste. Lors de la sélection d'un libellé, la vue Libellé affiche les valeurs de données que vous avez attribuées au libellé dans chaque ligne.

L'image suivante illustre les options du modèle probabiliste que vous pouvez utiliser lorsque vous sélectionnez la vue Données :



La vue Libellé comprend les options suivantes :

1. Gérer les libellés
Ouvre la boîte de dialogue **Gérer les libellés** qui vous permet d'ajouter ou de supprimer des valeurs de libellés du modèle probabiliste.
2. Attribuer un libellé
Attribue un libellé à une ou plusieurs valeurs de données de référence que vous sélectionnez.
Vous pouvez attribuer un libellé à une ou à plusieurs valeurs de données en une seule opération.
3. Modifier les propriétés
Affiche les propriétés du modèle probabiliste.
4. Compiler
Compile le modèle probabiliste.
5. Liste des valeurs de libellés
Répertorie les libellés que vous pouvez attribuer aux valeurs de données de référence dans le modèle.
6. Filtre d'attribution
Filtre la liste des valeurs de données de référence qui utilisent le libellé sélectionné. Les options de filtre affichent ou masquent les valeurs de données de référence en fonction de la méthode que vous avez utilisée pour attribuer le libellé aux valeurs de données.
Lorsque vous appliquez un filtre, le nombre total de valeurs libellées de la vue Libellé reflète le nombre de valeurs répondant à la condition de filtre.
7. Colonne Valeur de données de référence
Répertorie les valeurs de données de référence qui utilisent le libellé en cours.
8. Nombre total de valeurs avec libellés
Indique le nombre de valeurs de données de référence qui utilisent le libellé en cours.

Données de référence du modèle probabiliste

Les valeurs de données de référence d'un modèle probabiliste représentent les types de données d'entrée que vous pouvez connecter à une transformation dans un mappage.

Vous pouvez ajouter, éditer et supprimer des lignes de données de référence dans l'outil Developer. Vous pouvez coller des données depuis le presse-papiers ou en importer depuis une source de données. Une fois que vous avez ajouté les valeurs de données de référence, attribuez un libellé à chacune d'elle dans chaque ligne.

Données de libellé du modèle probabiliste

Les valeurs de libellé dans un modèle probabiliste représentent les types d'informations que les valeurs de données de référence peuvent contenir. Lorsque vous ajoutez des lignes de données de référence à un

modèle, attribuez un libellé à chaque valeur de données de chaque ligne. Les libellés que vous ajoutez au modèle s'affichent dans la vue Libellé et dans les options de menu de la vue Données.

Vous pouvez attribuer tous les libellés du modèle à toutes les valeurs de données de référence. Si la même valeur a des significations différentes dans différentes lignes de données de référence, vous pouvez attribuer un libellé distinct à chaque valeur de chaque ligne.

La plage des valeurs de libellés peut correspondre à la plage des ports d'entrée que les transformations Libellé ou Analyseur lisent lors de l'analyse probabiliste. Le modèle probabiliste doit contenir au moins une valeur de libellé que la transformation peut appliquer aux valeurs de données de chaque port d'entrée.

Par exemple, un entrepôt peut stocker des données d'inventaire dans un fichier de données séparées par des virgules qui définit huit colonnes. Concevez un mappage permettant d'analyser les données d'inventaire dans une table de base de données. Créez un modèle probabiliste avec une valeur de libellé pour chaque colonne de données. Lors de l'exécution du mappage, la transformation Analyseur écrit chaque valeur des données d'entrée dans la colonne correcte de la table cible.

Le tableau suivant montre les colonnes des données d'inventaire ainsi que les valeurs de libellés que vous pouvez créer dans un modèle probabiliste :

Nom de colonne d'inventaire	Nom du libellé
Product_Name	Product_Name
Quantité	Quantité
Emplacement	Emplacement
Code-barres	Code-barres
UGS	Stock_Keeping_Unit
Arrival_Date	Arrival_Date
Cost_Price	Cost_Price

Remarque: vous pouvez utiliser les noms des colonnes d'entrée ou d'autres noms. Il n'est pas nécessaire que les noms correspondent.

Libellé de dépassement

Lorsqu'une transformation ne peut pas appliquer de libellé à une valeur de données d'entrée, elle traite la valeur de données en tant que données de dépassement. La transformation Libellé applique un libellé de dépassement à toutes les valeurs de données qu'elle ne peut pas identifier. La transformation Analyseur écrit toutes les valeurs de données qu'elle ne peut pas identifier dans un port de dépassement.

Le tableau suivant montre comment une transformation Analyseur peut utiliser un port de dépassement pour analyser les éléments de données d'adresse qu'un modèle probabiliste ne reconnaît pas :

Données d'entrée	Port Street_Name	Port Street_Descriptor	Port de dépassement
Park Place	Park	Place	Pas de données de dépassement
Park Avenue	Park	Avenue	Pas de données de dépassement

Données d'entrée	Port Street_Name	Port Street_Descriptor	Port de dépassement
Madison Avenue	Madison	Avenue	Pas de données de dépassement
Central Park	Central	Park	Pas de données de dépassement
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

La transformation Analyseur écrit également des valeurs dans un port de dépassement lorsque le nombre de valeurs d'entrée est supérieur au nombre de libellés figurant dans le modèle. Avant d'utiliser un modèle probabiliste dans une transformation, vérifiez les données d'entrée et assurez-vous que le modèle contient le nombre approprié de valeurs de libellés.

Propriétés du modèle probabiliste

Vous pouvez vérifier les propriétés générales et les propriétés avancées d'un modèle probabiliste.

Pour ouvrir l'éditeur de propriétés, sélectionnez l'option **Modifier les propriétés** dans la vue Données ou Libellé.

Les propriétés générales affichent le nom du modèle probabiliste, une éventuelle description et le nom du fichier de données du modèle. Les propriétés avancées affichent les propriétés de calcul que l'outil Developer utilise pour compiler le modèle probabiliste.

L'élément de base de la compilation de modèles probabilistes est le *n-gramme*. Un *n-gramme* est une série de lettres qui suit ou précède d'autres lettres pour former un mot. Lorsqu'un mappage s'exécute, la transformation Libellé ou Analyseur crée plusieurs *n-grammes* pour chaque valeur de la colonne de données de référence du modèle probabiliste. La transformation compare les valeurs des données d'entrée aux valeurs de données de référence et aux *n-grammes*. Les propriétés avancées d'un modèle probabiliste déterminent la manière dont le modèle probabiliste gère les *n-grammes* et d'autres fonctionnalités de modèle.

Remarque: Les valeurs par défaut dans les propriétés avancées représentent les paramètres souhaités pour l'analyse probabiliste et la compilation du modèle probabiliste. Si vous modifiez une propriété avancée, vous risquez d'affecter négativement la précision de l'analyse probabiliste. Ne modifiez pas les propriétés avancées sauf si vous comprenez les conséquences des modifications que vous effectuez.

LIENS CONNEXES :

- ["Règles et directives pour les modèles probabilistes et les modèles de classeur" à la page 41](#)

Configuration d'un modèle probabiliste

Les étapes permettant de configurer un modèle probabiliste commencent par le type d'analyse que vous voulez effectuer. Un modèle probabiliste dans une transformation Libellé vous permet d'identifier les types d'informations de chaque valeur dans une chaîne d'entrée. Alors qu'un modèle probabiliste dans une

transformation Analyseur vous permet d'analyser les valeurs de données d'une chaîne d'entrée sur différents ports de sortie.

Vous pouvez utiliser le même modèle probabiliste pour libeller ou analyser des données. Lors de l'utilisation du modèle dans une transformation Libellé, celle-ci crée un seul port de sortie pour chaque port d'entrée que vous sélectionnez. Lors de l'utilisation du modèle dans une transformation Analyseur, celle-ci crée un port de sortie pour chaque type de données d'entrée qu'elle identifie.

Pour créer un modèle probabiliste, procédez comme suit :

1. Identifiez les valeurs de données de référence et les valeurs de libellés à ajouter au modèle.

Vous pouvez utiliser un fragment des données que vous souhaitez analyser. Créez un objet de données dans le référentiel modèle capable de lire le fragment de données.

2. Créez un ensemble de contenu et ajoutez-y un modèle probabiliste.

3. Ajoutez les valeurs de données de référence au modèle.

4. Ajoutez les valeurs de libellés au modèle.

Vous pouvez importer les données à partir de l'objet de données dans le référentiel modèle. Vous pouvez également entrer une ligne de données de référence ou un libellé.

Pour utiliser le modèle probabiliste de façon à analyser les données, assurez-vous qu'il contient une valeur de libellé pour chaque port de sortie que la transformation doit créer.

5. Attribuez un libellé à chaque valeur de données de référence de chaque ligne.

Vous pouvez attribuer un libellé à plusieurs valeurs de données de référence en une seule opération.

6. Compilez le modèle.

Après avoir compilé le modèle probabiliste, vous pouvez l'utiliser dans une transformation.

Création d'un modèle probabiliste vide

Vous pouvez créer un objet de modèle probabiliste qui ne contient pas de données de référence ni de données de libellés. Créez un modèle vide et ajoutez-lui des données ou importez-en.

1. Dans l'explorateur d'objets, ouvrez et créez un ensemble de contenu.

2. Sélectionnez la vue Contenu.

3. Sélectionnez **Modèles probabilistes**, puis cliquez sur **Ajouter**.

L'assistant Modèle probabiliste s'ouvre.

4. Sélectionnez l'option **Modèle probabiliste**.

Cliquez sur **Suivant**.

5. Saisissez un nom pour le modèle probabiliste.

Entrez éventuellement une description textuelle du modèle.

6. Cliquez sur **Terminer**.

Création d'un modèle probabiliste à partir d'un objet de données

Vous pouvez utiliser un objet de données en tant que source de données d'un modèle probabiliste.

Un modèle probabiliste fonctionne de manière optimale lorsque vous utilisez les données d'entrée d'une transformation Libellé ou Analyseur en tant que source des données de référence du modèle.

1. Dans l'explorateur d'objets, ouvrez et créez un ensemble de contenu.

2. Sélectionnez la vue **Contenu**.

3. Sélectionnez **Modèles probabilistes**, puis cliquez sur **Ajouter**.
L'assistant Modèle probabiliste s'ouvre.
4. Sélectionnez l'option **Modèle probabiliste issu d'objets de données**.
Cliquez sur **Suivant**.
5. Saisissez un nom pour le modèle probabiliste.
Entrez éventuellement une description textuelle du modèle.
6. Parcourez le référentiel modèle et sélectionnez l'objet de données qui contient les données à importer.
Ne sélectionnez pas d'objet de données de média social.
Cliquez sur **Suivant**.
7. Vérifiez les colonnes de l'objet de données et sélectionnez une ou plusieurs colonnes à ajouter au modèle. Vous pouvez ajouter des colonnes de données de référence et une colonne de libellé au cours de la même opération.
 - Pour importer une colonne de données comme données de référence, sélectionnez le nom de la colonne et cliquez sur **Données**.
Vous pouvez sélectionner plusieurs colonnes de données. L'outil Developer fusionne le contenu des colonnes que vous sélectionnez en une seule colonne.
 - Pour importer une colonne de données en tant que valeurs de libellés, sélectionnez le nom de la colonne et cliquez sur **Libellé**.

Lors de l'importation des données de référence et des valeurs de libellés, l'outil Developer attribue le libellé de chaque ligne à la chaîne de données de référence sur la même ligne. Vous pouvez prévisualiser les données avant de sélectionner les colonnes. Vous pouvez modifier les attributions de libellé après avoir créé le modèle.

Cliquez sur **Suivant**.
8. Sélectionnez le nombre de lignes à importer à partir de la source de données.
Par défaut, l'outil Developer importe toutes les lignes depuis la source de données. Si vous saisissez un nombre, le modèle compte les lignes depuis le début de l'ensemble de données.
9. Spécifiez les délimiteurs des valeurs de données que vous importez.
Vous pouvez spécifier différents délimiteurs pour les valeurs de données de référence et les valeurs de libellés. Le délimiteur par défaut est un caractère d'espacement.
10. Cliquez sur **Terminer** et enregistrez le modèle.
Après avoir créé le modèle probabiliste, vérifiez les attributions de libellé et compilez le modèle.

Ajout de données à un modèle probabiliste à partir d'une source de données

Vous pouvez importer dans un modèle probabiliste plusieurs lignes de valeurs de données de référence et des valeurs de libellés en une seule opération.

1. Ouvrez l'ensemble de contenu qui contient le modèle probabiliste.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Cliquez sur **Adjonction de données**.
L'assistant Modèle probabiliste s'ouvre.
4. Parcourez le référentiel modèle et sélectionnez l'objet de données qui contient les données à importer.
Ne sélectionnez pas d'objet de données de média social.

Cliquez sur **Suivant**.

5. Vérifiez les colonnes de l'objet de données et sélectionnez une ou plusieurs colonnes à ajouter au modèle. Vous pouvez ajouter des colonnes de données de référence et une colonne de libellé au cours de la même opération.

- Pour importer une colonne de données comme données de référence, sélectionnez le nom de la colonne et cliquez sur **Données**.

Vous pouvez sélectionner plusieurs colonnes de données. L'outil Developer fusionne le contenu des colonnes que vous sélectionnez en une seule colonne.

- Pour importer une colonne de données en tant que valeurs de libellés, sélectionnez le nom de la colonne et cliquez sur **Libellé**.

Lors de l'importation des données de référence et des valeurs de libellés, l'outil Developer attribue le libellé de chaque ligne à la chaîne de données de référence sur la même ligne. Vous pouvez prévisualiser les données avant de sélectionner les colonnes. Vous pouvez modifier les attributions de libellé après avoir créé le modèle.

Cliquez sur **Suivant**.

6. Sélectionnez le nombre de lignes à importer à partir de la source de données.

Par défaut, l'outil Developer importe toutes les lignes depuis la source de données. Si vous saisissez un nombre, le modèle compte les lignes depuis le début de l'ensemble de données.

7. Spécifiez les délimiteurs des valeurs de données que vous importez.

Vous pouvez spécifier différents délimiteurs pour les valeurs de données de référence et les valeurs de libellés. Le délimiteur par défaut est un caractère d'espacement.

8. Cliquez sur **Terminer** et enregistrez le modèle.

Ajout d'une ligne de données de référence à un modèle probabiliste

Utilisez la vue Données pour ajouter une ligne vide à un modèle probabiliste.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
Sélectionnez le nom du modèle et cliquez sur **Modifier**.
2. Sélectionnez la vue Données.
3. Pour ajouter une ligne vide au modèle, cliquez sur **Nouveau**.
4. Sélectionnez la ligne que vous avez ajoutée et entrez-y une ou plusieurs valeurs de données de référence.
5. Enregistrez le modèle probabiliste.

Après avoir enregistré le modèle, attribuez un libellé à chaque valeur de la ligne. Vous pouvez éventuellement compiler le modèle.

Ajout d'un libellé à un modèle probabiliste

Vous pouvez ajouter un seul libellé à un modèle probabiliste. Ajoutez un libellé pour chaque type d'information que les valeurs de données du modèle représentent. Si vous utilisez le modèle probabiliste dans une transformation Analyseur, ajoutez un libellé pour chaque port de sortie qui sera créé par la transformation.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.

3. Dans la vue Données ou Libellé, cliquez sur **Gérer les libellés**.
La boîte de dialogue **Gérer les libellés** s'affiche.
 4. Dans la boîte de dialogue **Gérer les libellés**, cliquez sur **Nouveau**.
Un libellé s'affiche dans la première ligne vide de la boîte de dialogue.
 5. Modifiez le nom du libellé. Éventuellement, mettez à jour la couleur du libellé.
 6. Cliquez sur **OK** pour ajouter le libellé au modèle.
 7. Enregistrez le modèle probabiliste.
- Après avoir ajouté le libellé, vous devez l'attribuer à une valeur de données au moins.

Attribution d'un libellé à une valeur de données de référence

Vous pouvez attribuer un libellé à une valeur de données dans une ligne de données de référence.

Si une même valeur de données s'affiche à différents emplacements de la ligne ou sur plusieurs lignes, vous pouvez lui attribuer différents libellés.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
 2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
 3. Sélectionnez la vue Données.
 4. Recherchez une valeur de données sans libellé ou dont le libellé est incorrect. Les valeurs de données qui utilisent un libellé disposent d'un code de couleur.
 5. Sélectionnez la ligne de données qui contient la valeur de données.
La ligne s'affiche dans l'éditeur.
 6. Cliquez avec le bouton droit de la souris sur une valeur de données dans l'éditeur et sélectionnez un libellé dans le menu contextuel.
L'outil Developer attribue le libellé à la valeur de données.
 7. Enregistrez le modèle probabiliste.
- Après avoir effectué l'enregistrement, vous pouvez éventuellement compiler le modèle.

Attribution d'un libellé à plusieurs valeurs de données

Vous pouvez attribuer un libellé à plusieurs valeurs de données de référence en une seule opération.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Cliquez sur **Attribuer un libellé**.
La boîte de dialogue **Attribuer un libellé à plusieurs valeurs** s'ouvre.
4. Entrez un ou plusieurs caractères dans le champ Rechercher.
Ce champ accepte les caractères génériques.
5. Vous pouvez éventuellement sélectionner d'autres critères de recherche.
Vous pouvez sélectionner ou désactiver les options suivantes :
 - Respecter la casse.
Indique que l'opération de recherche est sensible à la casse. N'utilisez pas les caractères génériques pour cette option.

- Rechercher la chaîne complète. Indique que l'opération recherche une correspondance exacte entre les caractères de la valeur de données de référence et les caractères que vous entrez. N'utilisez pas les caractères génériques pour cette option.
 - Ignorer les valeurs libellées.
Indique que l'opération de recherche ignore toutes les valeurs de données de référence qui utilisent un libellé.
6. Sélectionnez un libellé à attribuer aux valeurs de données de référence qui correspondent aux critères de recherche.
- Vous pouvez également sélectionner l'option **Aucun libellé**. Cette option vous permet de supprimer le libellé des valeurs de données de référence qui incluent les caractères que vous entrez.
7. Cliquez sur **Démarrer**.
- L'outil Developer tool attribue le libellé à toutes les valeurs de données de référence qui correspondent aux critères de recherche que vous définissez.
- Remarque:** Pour afficher les valeurs de données de référence que vous avez libellées en une seule opération, utilisez le filtre **Attribué en bloc** dans la vue Libellé.

Suppression de ligne d'un modèle probabiliste

Vous pouvez supprimer une ou plusieurs lignes de données de référence d'un modèle probabiliste en une seule opération.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
 2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
 3. Dans la vue Données, sélectionnez une ou plusieurs lignes de données de référence.
 4. Cliquez sur **Supprimer**.
L'outil Developer supprime les lignes que vous avez sélectionnées dans le modèle de classeur.
- Pour annuler l'opération, appuyez sur les touches Ctrl + Z.

Suppression d'un libellé d'un modèle probabiliste

Lorsque vous supprimez une valeur de libellé d'un modèle, toute valeur de données de référence utilisant le libellé est conservée dans le modèle. Attribuez une autre valeur de libellé à chaque valeur de données de référence.

1. Ouvrez l'ensemble de contenu qui contient le modèle.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Dans la vue Données ou Libellé, cliquez sur **Gérer les libellés**.
4. Dans la boîte de dialogue **Gérer les libellés**, sélectionnez une valeur de libellé.
5. Cliquez sur **Supprimer**.
6. Cliquez sur **OK** pour supprimer le libellé.
7. Enregistrez le modèle probabiliste.

Remarque: Un libellé est un élément structurel dans un modèle probabiliste. Si vous ajoutez ou supprimez un libellé après avoir ajouté le modèle à une transformation, vous invalidez l'opération qui utilise le modèle. Pour utiliser le modèle que vous avez mis à jour, supprimez et recréez l'opération de transformation.

Compilation du modèle probabiliste

Lors de la mise à jour des données ou des attributions de libellé dans un modèle probabiliste, vous pouvez compiler le modèle. Compilez le modèle afin de mettre à jour la logique du modèle avec les associations entre les valeurs de données de référence actuelles et les valeurs de libellés actuelles.

Avant de compiler le modèle probabiliste, vérifiez que chaque valeur de libellé identifie au moins une valeur de données de référence.

- Pour compiler le modèle, ouvrez-le dans l'outil Developer et cliquez sur **Compiler**.

Recherche de lignes de données dans un modèle probabiliste

La vue Données vous permet de trouver les lignes de données de référence qui contiennent la valeur que vous entrez.

1. Ouvrez l'ensemble de contenu qui contient le modèle probabiliste.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Sélectionnez la vue Données.
4. Entrez un ou plusieurs caractères dans le champ **Rechercher**.
La vue Données affiche la première ligne dans le modèle contenant la valeur que vous avez saisie.
5. Utilisez les flèches vers le haut ou vers le bas pour accéder à d'autres lignes qui contiennent la valeur.

Filtrage des valeurs de données de référence par l'attribution d'un libellé

Utilisez la vue Libellé pour trouver les valeurs de données de référence qui utilisent un libellé spécifié. Filtrer les résultats en fonction de la méthode que vous avez utilisée pour attribuer le libellé.

1. Ouvrez l'ensemble de contenu qui contient le modèle probabiliste.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Dans la vue Libellé, sélectionnez une valeur de libellé.
Le modèle probabiliste affiche une liste des valeurs de données de référence qui utilisent le libellé. Le modèle affiche également le nombre de valeurs de données qui utilisent le libellé.
4. Appliquer un filtre à la liste des valeurs de données de référence qui utilisent le libellé.
Sélectionnez l'un des filtres suivants :
 - Tous. Affiche les valeurs de données de référence qui utilisent le libellé. L'option par défaut est Tout.
 - Attribué par l'utilisateur. Affiche toutes les valeurs de données de référence que vous avez sélectionnées individuellement lorsque vous avez attribué le libellé.
 - Attribué en bloc. Affiche les valeurs de données de référence auxquelles vous avez attribué un libellé comme partie d'une opération d'attribution en bloc.

Le modèle probabiliste affiche les valeurs de données de référence qui répondent à la condition de filtre.

Recherche de valeurs de libellés inutilisées

La vue Libellé vous permet de trouver une valeur de libellé que vous n'avez pas attribuée à une valeur de données de référence. Vous devez attribuer chaque libellé à au moins une valeur de données de référence.

1. Ouvrez l'ensemble de contenu qui contient le modèle probabiliste.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.

3. Dans la vue Libellé, sélectionnez une valeur de libellé.

Le modèle probabiliste affiche une liste des valeurs de données de référence qui utilisent le libellé. Le modèle affiche également le nombre total de valeurs de données qui utilisent le libellé.

Si le nombre total de valeurs de données est de zéro, vous n'avez attribué le libellé à aucune valeur de données de référence du modèle probabiliste.

Opérations de copier/coller

Vous pouvez copier un modèle probabiliste d'un ensemble de contenu vers un autre dans un référentiel modèle. Copiez un modèle probabiliste pour partager les ressources avec d'autres utilisateurs de l'outil Developer.

Vous pouvez copier un modèle vers un autre ensemble de contenu ou importer un modèle dans l'ensemble de contenu actuel. Vous pouvez importer plusieurs modèles de plusieurs ensembles de contenu dans le référentiel en une seule fois.

Lorsque vous copiez un modèle, le service de gestion du contenu crée une copie du fichier de données du modèle sur la machine hôte des services Informatica. Chaque modèle utilise un fichier de données différent.

Copie d'un modèle probabiliste dans un autre ensemble de contenu

Vous pouvez copier un modèle probabiliste d'un ensemble de contenu vers un autre dans un référentiel modèle. Lorsque vous copiez un modèle probabiliste, vous spécifiez l'objet de modèle et les ensembles de contenu source et de destination.

1. Ouvrez l'ensemble de contenu qui contient le modèle probabiliste.
2. Sélectionnez un modèle probabiliste et cliquez sur **Copier dans**.
3. Parcourez le référentiel modèle et sélectionnez un ensemble de contenu.

Vous pouvez copier le modèle probabiliste vers un ensemble de contenu dans le projet en cours ou un autre projet.

4. Cliquez sur **OK**.

L'outil Developer copie le modèle probabiliste dans l'ensemble de contenu sélectionné.

Importation d'un modèle probabiliste depuis un autre ensemble de contenu

Vous pouvez importer un modèle probabiliste d'un ensemble de contenu vers un autre dans un référentiel modèle. Lorsque vous importez un modèle probabiliste, vous spécifiez un ou plusieurs objets de modèle et les ensembles de contenu source et de destination.

1. Ouvrez l'ensemble de contenu devant contenir le modèle probabiliste.
2. Sélectionnez un modèle probabiliste et cliquez sur **Coller depuis**.
3. Parcourez le référentiel modèle et sélectionnez un modèle probabiliste.

Vous pouvez coller le modèle probabiliste d'un ensemble de contenu dans le projet actuel ou un autre projet.

4. Cliquez sur **OK**.

L'outil Developer colle le modèle probabiliste dans l'ensemble de contenu actuel.

Copie des lignes de données de référence dans le presse-papiers

Vous pouvez copier une ou plusieurs lignes de données de référence d'un modèle probabiliste dans le presse-papiers. Vous pouvez coller les lignes dans un autre modèle probabiliste.

1. Ouvrez l'ensemble de contenu qui contient le modèle probabiliste.
2. Sélectionnez le nom du modèle et cliquez sur **Modifier**.
3. Dans la vue Données, sélectionnez une ou plusieurs lignes de données de référence.
4. Utilisez les touches CTRL + C pour copier les lignes dans le presse-papiers.

L'opération copie les données de référence et les valeurs de libellés que vous avez attribuées aux données de référence.

Vous pouvez utiliser les touches CTRL + V pour coller les lignes dans un éditeur de texte ou dans la vue Données d'un autre modèle probabiliste.

ANNEXE A

Données de référence et Informatica Data Engineering Integration

- [Données de référence et Informatica Data Engineering Integration, 73](#)

Données de référence et Informatica Data Engineering Integration

Informatica Data Engineering Integration combine un domaine Informatica et les applications client Informatica avec un cluster Hadoop. Vous pouvez transférer des mappages à partir de l'outil Developer tool vers la grappe et exécuter les mappages sur les nœuds de la grappe.

L'opération de refoulement copie les données de la table de référence, les données de jeu de contenu et les données de population d'identité que le mappage lit dans le cluster. Une fois le mappage exécuté, la grappe supprime les données de référence copiées par l'opération de refoulement avec le mappage.

L'opération de refoulement ne copie pas les données de référence de validation des adresses. Si vous transférez un mappage qui effectue une validation des adresses, vous devez installer les fichiers de données de référence de validation des adresses sur chaque nœud qui exécute le mappage. La grappe ne supprime pas les fichiers de données de référence de validation des adresses une fois que le mappage de validation des adresses est exécuté.

Données de référence pour la validation des adresses

Lorsque vous exécutez un mappage de validation des adresses dans un environnement Hadoop, les fichiers de données de référence d'adresse doivent résider sur chaque DataNode sur lequel le mappage est exécuté. Informatica Data Engineering Integration s'installe avec un script shell que vous pouvez utiliser pour installer les fichiers sur les DataNodes.

Utilisez le script shell pour installer les fichiers de données de référence d'adresse sur les DataNodes en une seule opération. Le script lit un fichier qui contient les noms ou les adresses IP des nœuds. Le script copie les fichiers de données de référence d'adresse sur chaque nœud que le fichier identifie.

Le nom du script est `copyRefDataToComputeNodes.sh`.

Recherchez le script dans le répertoire suivant de l'installation de Data Engineering Integration :

<répertoire d'installation Informatica>/tools/dq/av

Le tableau suivant décrit les options utilisées par le script :

Option	Description
-n	Fichier qui contient la liste des noms ou des adresses IP des DataNodes dans la grappe Hadoop. Entrez chaque nom de nœud ou adresse IP sur une ligne distincte dans le fichier. Par défaut, le script lit le fichier depuis le répertoire \$BASEDIR/HadoopDataNodes , où \$BASEDIR est l'emplacement du script shell.
-p	Invite permettant de confirmer que vous souhaitez installer les fichiers de données de référence d'adresse. Par défaut, le script affiche une invite vous permettant de confirmer votre souhait de copier les fichiers à partir du répertoire source vers les répertoires cibles sur les DataNodes. Si vous exécutez le script shell sur une planification, vous pouvez désactiver l'invite. La valeur de l'option par défaut est Y. Pour désactiver l'invite, définissez la valeur sur N.
-s	Répertoire source des fichiers de données de référence d'adresse que le script copie sur les nœuds. Par défaut, le script lit les fichiers à partir du répertoire /reference_data sur la machine locale. Remarque: Les fichiers de données de référence d'adresse utilisent l'extension de nom de fichier .MD. Le répertoire source doit contenir les fichiers de données de référence d'adresse et aucun autre fichier.
-t	Répertoire de chaque nœud sur lequel le script copie les fichiers de données de référence d'adresse. Par défaut, le script copie les fichiers dans le répertoire /reference_data de chaque nœud.
-u	Nom d'utilisateur de l'utilisateur qui exécute le script. L'utilisateur doit disposer d'un accès shell sécurisé sans mot de passe pour les nœuds.

Installation des fichiers de données de référence d'adresse

Pour installer des fichiers de données de référence d'adresse sur le DataNodes d'une grappe Hadoop, exécutez le script shell `copyRefDataToComputeNodes.sh`. Ou, définissez une tâche permettant d'exécuter le script shell dans une application de planificateur de tâches selon des intervalles de temps que vous spécifiez.

Avant d'exécuter le script ou de définir la tâche, vérifiez les valeurs d'option spécifiées pour le script. Vous pouvez accepter les valeurs par défaut ou les mettre à jour.

Installation des fichiers de données de référence d'adresse lors de l'invite de commande

Pour installer les fichiers lors de l'invite de commande, procédez comme suit :

1. Lors de l'invite de commande, ouvrez le répertoire suivant :
<répertoire d'installation Informatica>/tools/dq/av
2. Exécutez `copyRefDataToComputeNodes.sh`.
Vous pouvez éventuellement entrer une ou plusieurs valeurs pour les options de script. Si vous n'entrez pas de valeur pour une option, le script s'exécute avec la valeur de l'option par défaut.
Par défaut, le script vous invite à confirmer l'installation des fichiers. Pour installer les fichiers, entrez Y.

Installation des fichiers de données de référence d'adresse à l'aide d'une tâche planifiée

Vous pouvez définir une tâche afin d'exécuter le script shell selon des intervalles de temps que vous spécifiez. Ajoutez la tâche à une application de planificateur de tâche. Si vous définissez une tâche pour installer les fichiers, vous devez désactiver l'invite pour confirmer l'installation.

Pour désactiver l'invite, définissez l'option suivante dans le script shell :

-p n

INDEX

C

contrôle de version
ensembles de contenu [13](#), [31](#)
tables de référence [13](#)
tables de référence dans l'outil Analyst tool [23](#)
tables de référence dans l'outil Developer tool [31](#)
créer une table de référence depuis les données de colonne de profil
tables de référence [17](#)
Créer une table de référence depuis les modèles de colonnes
tables de référence [19](#)
créer une table de référence manuellement
tables de référence [16](#)

D

Data Engineering Integration
conditions requises pour les données de référence [73](#)
installation de données de référence d'adresse [74](#)
script d'installation des données de référence d'adresse [73](#)

E

ensembles de caractères [37](#)
ensembles de contenu
contrôle de version [13](#), [23](#), [31](#)
ensembles de caractères [37](#)
ensembles de jetons [40](#)
ensembles de modèles [38](#)
expressions régulières [39](#)
modèles de classeur [38](#)
modèles probabilistes [38](#)
ensembles de jetons [40](#)
ensembles de modèles [38](#)
environnement Hadoop
conditions requises pour les données de référence [73](#)
installation de données de référence d'adresse [74](#)
script d'installation des données de référence d'adresse [73](#)
exporter une table de référence
tables de référence [26](#)
expressions régulières [39](#)

G

gérer les colonnes
tables de référence [24](#)
gérer les lignes
tables de référence [24](#)

I

importer une table de référence
tables de référence [21](#)

M

modèles de classeur
dans les ensembles de contenu [38](#)
règles et directives [41](#)
modèles probabilistes
dans les ensembles de contenu [38](#)
règles et directives [41](#)

O

Outil Analyst
rechercher et remplacer des valeurs de données de référence [25](#)

P

privilèges
service de gestion de contenu [12](#)

S

service de gestion de contenu
privilèges de la table de référence [12](#)
Service de gestion de contenu
entrepôt de données de référence [11](#)

T

tables de référence
Actualisation dans l'outil Analyst [27](#)
contrôle de version [13](#), [23](#), [31](#)
créer une table de référence depuis les colonnes de profil [17](#)
Créer une table de référence depuis les modèles de colonnes [19](#)
créer une table de référence manuellement [16](#)
dans une analyse basée sur un modèle [12](#)
entrepôt de données de référence [11](#)
exporter une table de référence [26](#)
gérées et non gérées [11](#)
gérer les colonnes [24](#)
gérer les lignes [24](#)
importer une table de référence [21](#)
Présentation de Developer tool [32](#)
Présentation de l'outil Analyst [14](#)
privilèges [12](#)
propriétés dans l'outil Analyst [15](#)
propriétés dans l'outil Developer [33](#)

tables de référence (*a continué*)

rechercher et remplacer des valeurs dans l'outil Analyst [25](#)

Service de gestion de contenu [11](#)

tables de référence gérées [11](#)

tables de référence non gérées [11](#)

voir les tables de suivi d'audit [28](#)

tables de référence gérées [11](#)

tables de référence non gérées

activer et désactiver les modifications [26](#)

définition [11](#)

tables de référence non gérées (*a continué*)

synchronisation avec le référentiel modèle [12](#)

V

voir les événements de table d'audit

tables de référence [28](#)