

Guide de l'utilisateur IBM SPSS Modeler
Entity Analytics 15



Remarque : Avant d'utiliser ces informations et le produit qu'elles concernent, lisez les informations générales sous Remarques sur p. 66.

Cette version s'applique à et à toutes les publications et modifications ultérieures jusqu'à mention contraire dans les nouvelles versions.

Les captures d'écran des produits Adobe sont reproduites avec l'autorisation de Adobe Systems Incorporated.

Les captures d'écran des produits Microsoft sont reproduites avec l'autorisation de Microsoft Corporation.

Matériel sous licence - Propriété d'IBM

Droits limités pour les utilisateurs au sein d'administrations américaines : utilisation, copie ou divulgation soumise au GSA ADP Schedule Contract avec IBM Corp.

Préface

IBM® SPSS® Modeler est le puissant utilitaire de Data mining de IBM Corp.. SPSS Modeler aide les entreprises et les organismes à améliorer leurs relations avec les clients et les citoyens grâce à une compréhension approfondie des données. A l'aide des connaissances plus précises obtenues par le biais de SPSS Modeler, les entreprises et les organismes peuvent conserver les clients rentables, identifier les opportunités de vente croisée, attirer de nouveaux clients, détecter les éventuelles fraudes, réduire les risques et améliorer les services gouvernementaux.

L'interface visuelle de SPSS Modeler met à contribution les compétences professionnelles de l'utilisateur, ce qui permet d'obtenir des modèles prédictifs plus efficaces et de trouver des solutions plus rapidement. SPSS Modeler dispose de nombreuses techniques de modélisation, telles que les algorithmes de prévision, de classification, de segmentation et de détection d'association. Une fois les modèles créés, l'utilisateur peut utiliser IBM® SPSS® Modeler Solution Publisher pour les remettre aux responsables, où qu'ils se trouvent dans l'entreprise, ou pour les transférer vers une base de données.

A propos de IBM Business Analytics

Le logiciel IBM Business Analytics fournit des informations complètes, cohérentes et précises que les preneurs de décision utilisent avec confiance pour améliorer la performance du marché. Un portefeuille étendu d'outils de [business intelligence](#), d'[analyses prédictives](#), de [performance financière et de gestion de stratégie](#), et des [applications analytiques](#) offre des connaissances claires, immédiates et applicables pour améliorer l'efficacité actuelle ainsi que la capacité de prévoir les résultats futurs. Combinées avec de riches solutions industrielles, des pratiques éprouvées et des services professionnels, les organisations de toutes tailles peuvent atteindre la productivité la plus élevée, automatiser des décisions en toute tranquillité et fournir de meilleurs résultats.

Dans le cadre de ce portefeuille, le logiciel IBM SPSS Predictive Analytics aide les organisations à prévoir des événements futurs et à agir en conséquence pour mener à de meilleurs résultats. Des clients dans le domaine commercial, gouvernemental et académique à travers le monde font confiance à la technologie IBM SPSS et considèrent qu'elle représente un avantage compétitif pour attirer, retenir et ajouter des clients, tout en réduisant la fraude et en atténuant les risques. En incorporant le logiciel IBM SPSS dans leur opérations quotidiennes, les organisations deviennent des entreprises prédictives – capables de diriger et d'automatiser les décisions pour atteindre les buts qu'ils se sont fixés et obtenir des avantages compétitifs sensibles. Pour informations supplémentaires ou pour joindre un revendeur, visitez le site <http://www.ibm.com/spss>.

Assistance technique

L'assistance technique est à la disposition des clients pour la maintenance des produits. Les clients peuvent contacter l'assistance technique pour obtenir de l'aide concernant l'utilisation des produits IBM Corp. ou l'installation dans l'un des environnements matériels pris en charge. Pour joindre l'assistance technique, consultez le site Web de IBM Corp. à l'adresse <http://www.ibm.com/support>. Lorsque vous contactez l'assistance technique, n'oubliez pas de préparer vos identifiants, le nom de votre société et votre contrat d'assistance.

Contenu

1	Analyses d'entités	1
	A propos des analyses d'entités	1
	Analyses d'entités et analyses prédictives	3
2	Analyses d'entités avec IBM SPSS Modeler	4
	Utilisation des analyses d'entités avec IBM SPSS Modeler	4
	Étape 1 : lire les données source dans SPSS Modeler	5
	Étape 2 : créer le référentiel	6
	Étape 3 : Connecter SPSS Modeler au référentiel	7
	Étape 4 : mapper les champs d'entrée sur les fonctionnalités du référentiel	7
	Étape 5 : exporter les données vers le référentiel et résoudre les conflits	8
	Étape 6 : analyse des identités résolues	11
	Étape 7 : résoudre de nouvelles observations en fonctionnalité du référentiel	11
	Étape 8 : générer des alertes	13
3	Tâches des analyses d'entités	14
	À propos des tâches	14
	Définition d'un référentiel d'entité (nœud Export EA)	14
	Le référentiel d'entité	14
	Connexion à une source de données	15
	Création d'un référentiel	16
	Mappage de champs d'entrées sur les fonctionnalités (nœud Export EA)	19
	Affichage des mappages de champs (nœud Export EA)	22
	Configuration d'un référentiel d'entité	22
	Affichage des mappages de sources de données	24
	Gestion des fonctionnalités du référentiel	24
	Ajout ou modification d'une fonctionnalité	26
	Gestion des types d'entités	28
	Définition du seuil de mise en correspondance des entités	31
	Réutilisation d'une configuration de référentiel	31
	Enregistrement de vos modifications de configuration	32
	Fermeture de la fenêtre de configuration	32
	Analyse des identités résolues (nœud source Entity Analytics (EA))	32
	Sélection d'une source de données	32
	Nouveau nom des champs de données	33

Définition des informations de type pour les champs de données	34
Ajout de nœuds au flux	34
Comparaison de nouvelles observations au contenu du référentiel (nœud Mise en flux EA)	34
Mappage de champs d'entrées sur les fonctionnalités (nœud Mise en flux EA)	36
Affichage des mappages de champs (nœud Mise en flux EA)	37
Affichage des sources de données (nœud Mise en flux EA)	38
Sortie du nœud Mise en flux EA	40
Utiliser IBM SPSS Modeler Entity Analytics avec les autres produits IBM SPSS	40
Tâches administratives	41
Configuration des tâches de port	41
Gestion des identifiants de l'administrateur pour la base de données du référentiel	42
Déplacement du répertoire de stockage du référentiel	43
Définition des options flux pour les champs date/heure et horodatage	43
Exécution de IBM SPSS Modeler Entity Analytics avec le client SPSS Modeler et SPSS Modeler Server installés sur le même système Windows	44
Purge d'un référentiel d'entité	44
Suppression d'un référentiel d'entité	44
Suppression d'un référentiel quand aucune connexion n'est possible	45

4 Fonctionnement des analyses d'entités 47

A propos de cet exemple	47
Le modèle d'origine	47
Ajout d'analyses d'entités	51
Ajout des données source au référentiel	52
Lecture des identités résolues	53
Comparaison de la sortie des analyses d'entités au modèle d'origine	61
Récapitulatif	65

Annexe

A Remarques 66

Index 69

Analyses d'entités

A propos des analyses d'entités

IBM® SPSS® Modeler Entity Analytics ajoute une dimension entièrement nouvelle aux analyses prédictives IBM® SPSS® Modeler. Alors que les analyses prédictives essaient de prévoir les comportements futurs à partir de données passées, les analyses d'entités se concentrent sur l'amélioration de la cohérence des données actuelles en résolvant les conflits d'identités dans les enregistrements eux-mêmes. Une identité peut être celle d'un individu, d'une organisation, d'un objet ou d'une autre entité pour laquelle une ambiguïté peut exister. La résolution d'identité peut être vitale dans de nombreux domaines, y compris la gestion de la relation client, la détection de la fraude, le blanchiment d'argent et la sécurité nationale et internationale.

Imaginons que vous disposez des enregistrements clients suivants provenant de deux sources différentes et que vous ne savez pas s'ils font référence à la même personne ou non.

Source 1

Enreg. n° 70001
Nom : Jon Smith
Adresse : 123 Main Street
Numéro fiscal : 555-00-1111
Permis de conduire : 0001133107
Carte de crédit : 10229127

Source 2

Enreg. n° : 9103
Nom : JONATHAN Smith
Date de naissance : 17/06/1934
Téléphone : 555-1212
Carte de crédit : 10229128
E-mail : jls@mail.com
Adresse IP : 9.50.18.77

Il n'y a pas de correspondances exactes dans les données entre les deux enregistrements. Cependant, si nous introduisons une troisième source, nous trouvons des attributs communs.

Source 3

Enreg. n° : 6251
Nom : Jon Smith
Téléphone : 555-1212
Permis de conduire : 0001133107
Carte de crédit : 10229132

Le numéro de permis de conduire relie les enregistrements de la Source 1 et de la Source 3 alors que le numéro de téléphone relie les Sources 2 et 3. Par conséquent, nous pouvons raisonnablement assumer que les trois sources font référence à la même personne.

Mais que se passe-t-il si la distinction n'est pas si simple ? Il est possible que nous ne disposions que d'un petit nombre de données nous permettant de nous faire une opinion. Examinez les deux enregistrements suivants.

Source 4

Enreg. n° : S45286
Nom : John T Smith Jr
Adresse : 456 Main Street
Téléphone : 703-555-2000
Date de naissance : 03/12/1984

Enreg. n° : S45287
Nom : John T Smith
Adresse : 456 Main Street
Téléphone : 703-555-2000
Permis de conduire : 009900991

Apparemment, il ne s'agit pas du même M. Smith que dans les enregistrements précédents. Les différences sont assez pertinentes pour que nous puissions éliminer cette possibilité. Cependant, il reste un problème. Deux enregistrements différents, de la même source de données, semblent être associés à la même personne. S'agit-il d'enregistrements en double ? Nous ne pouvons en être sûrs sans trouver un autre enregistrement du même genre qui nous donnera d'autres informations, peut-être d'une autre source.

Source 5

Enreg. n° : 769582-2
Nom : John T Smith Sr
Adresse : 456 Main Street
Téléphone : 703-555-2000
Permis de conduire : 009900991
Date de naissance : 25/06/1959

Cela résout le problème. Les deux enregistrements de la Source 4 ne sont pas des doublons mais sont en fait un père et un fils portant le même nom, vivant à la même adresse et utilisant le même numéro de téléphone. Sur un système manuel, il aurait fallu des semaines de recherche avant de découvrir l'enregistrement pouvant résoudre les identités. Avec un système d'analyse d'entités automatisé, le temps de résolution est réduit de manière considérable.

Analyses d'entités et analyses prédictives

Si toutes vos données étaient composées d'une seule source d'enregistrements complets et sans ambiguïtés, il serait relativement simple pour IBM® SPSS® Modeler de résoudre les conflits d'identités. En utilisant uniquement des analyses prédictives, vous pourriez lire vos données dans SPSS Modeler, effectuer votre traitement et obtenir des résultats fiables.

Dans le monde réel, cependant, la situation est généralement très différente. Les données sont rarement complètes, souvent ambiguës et éparpillées dans de nombreuses sources de données qui enregistrent de nombreux attributs différents avec peu de champs en double. Une partie de la valeur des analyses d'entités tient dans la collecte de données provenant de différentes sources pour les rassembler dans une seule zone de stockage centrale, appelée **référentiel**. Le système d'analyses d'entités examine ensuite attentivement les données pour résoudre les conflits, ajoutant un identifiant unique aux enregistrements qui proviennent de la même personne ou de la même organisation.

Le tableau suivant illustre les différences entre les deux types d'analyses.

Table 1-1

Différences entre analyses prédictives et analyses d'entités

Caractéristique	Analyse prédictive	Analyses d'entités
Formation	Basé sur des ensembles et des intervalles numériques peu importants	Basé sur de grands ensembles, par exemple des noms, des adresses
Données utilisées pour l'apprentissage	Sous-ensemble (partition d'apprentissage)	Toutes les données utilisées
Généralisation	L'algorithme est généralisé dans les données d'apprentissage pour former un modèle précis	Données conservées dans des structures adaptées à la mise en correspondance des entités et à la détection des relations
Détection des fraudes	Enregistrements signalés comme potentiellement frauduleux s'ils ont les caractéristiques habituelles des demandes frauduleuses	Enregistrements signalés comme potentiellement frauduleux s'ils sont associés à des enregistrements frauduleux ou s'ils proviennent des mêmes individus mais avec des identités différentes

Analyses d'entités avec IBM SPSS Modeler

Utilisation des analyses d'entités avec IBM SPSS Modeler

Vous vous rendez compte que vous pourriez avoir des problèmes d'identité avec vos données. En quoi IBM® SPSS® Modeler Entity Analytics peut-il vous aider à résoudre ce problème ? La procédure suivante est une suggestion. Il est possible que vous deviez l'adapter à vos besoins.

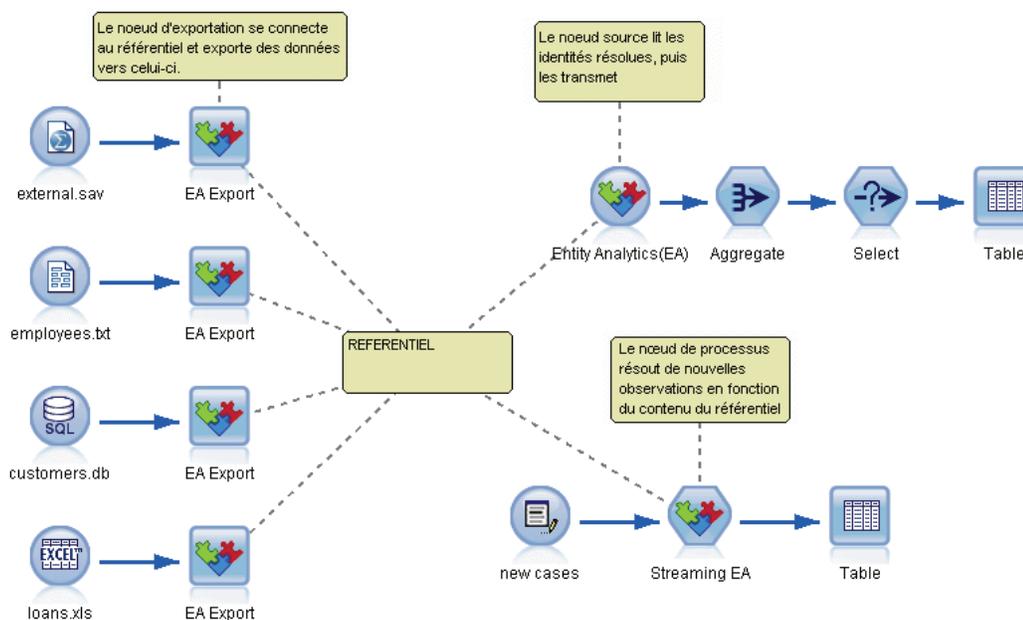
- lire les données source dans IBM® SPSS® Modeler
- Créer un référentiel prêt à héberger les données
- Connecter SPSS Modeler au référentiel
- Mapper les champs de données sur les fonctionnalités du référentiel
- Exporter les données dans le référentiel et résoudre les identités
- Analyser les identités résolues
- résoudre de nouvelles observations en fonctionnalité du référentiel
- Générer les alertes nécessaires (par lots ou en temps réel)

À ce moment-là, il est nécessaire que vous sachiez comment SPSS Modeler fonctionne. SPSS Modeler est un outil particulièrement convivial basé sur la représentation graphique d'un flux de données traversant un certain nombre de nœuds. Chaque nœud représente une étape spécifique du flux de travail.

SPSS Modeler propose de nombreux nœuds qui couvrent toutes les fonctionnalités de Data Mining standard. IBM SPSS Modeler Entity Analytics ajoute des nœuds spécifiques aux analyses d'entités. Il s'agit du nœud Export EA, du nœud source de données d'Entity Analytics (EA) et du nœud de processus Mise en flux EA.

La figure suivante illustre ce processus.

Figure 2-1
Le processus d'analyses d'entités



Étape 1 : lire les données source dans SPSS Modeler

Votre première tâche est de lire vos données dans SPSS Modeler au moyen d'un nœud source, signalé dans SPSS Modeler par une icône arrondie.

Figure 2-2
nœud source IBM SPSS Modeler



Les données peuvent être dans n'importe quel format pris en charge par SPSS Modeler, comme les fichiers texte, les tables de base de données, les feuilles de calcul, les fichiers XML, etc., mais chaque format nécessite un nœud source SPSS Modeler correspondant. Dans l'illustration, il s'agit d'un nœud source de base de données.

Chaque fichier source de données doit contenir un champ qui définit chaque enregistrement de manière unique. Si une source de données ne contient pas ce genre de champ, vous pouvez facilement en ajouter un dans SPSS Modeler. Pour plus d'informations, reportez-vous à la section [Ajout d'un identifiant d'enregistrement unique](#) dans le chapitre 3 sur p. 15.

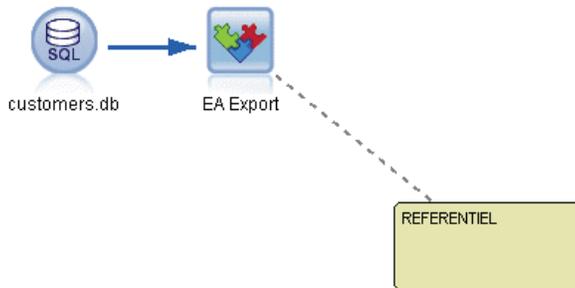
Pour plus d'informations, reportez-vous à la section [Connexion à une source de données](#) dans le chapitre 3 sur p. 15.

Étape 2 : créer le référentiel

L'élément central de vos analyses d'entités est le référentiel qui est la zone de stockage centrale dans laquelle vous rassemblez tous vos enregistrements de données.

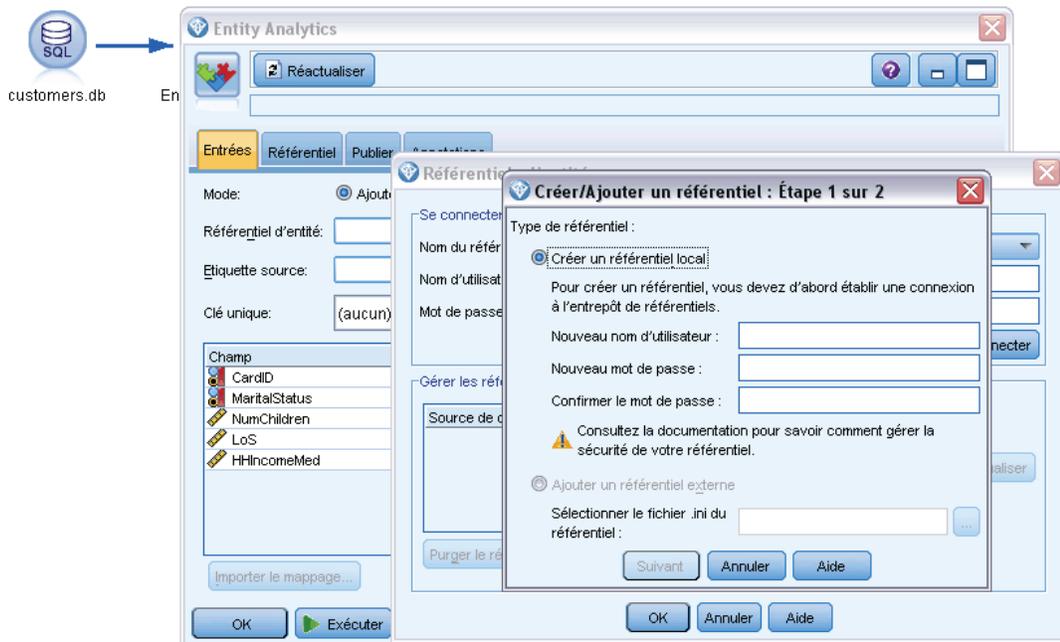
Pour créer un référentiel, commencez par connecter la source de données à un nœud Export EA, représenté par l'icône carrée.

Figure 2-3
Connexion à un référentiel



Depuis le nœud Export, vous pouvez créer un nouveau référentiel (ou sélectionner un référentiel existant) prêt à recevoir les données exportées.

Figure 2-4
Création d'un référentiel



Le processus de création d'un référentiel est décrit ultérieurement. Pour plus d'informations, reportez-vous à la section [Définition d'un référentiel d'entité \(nœud Export EA\)](#) dans le chapitre 3 sur p. 14.

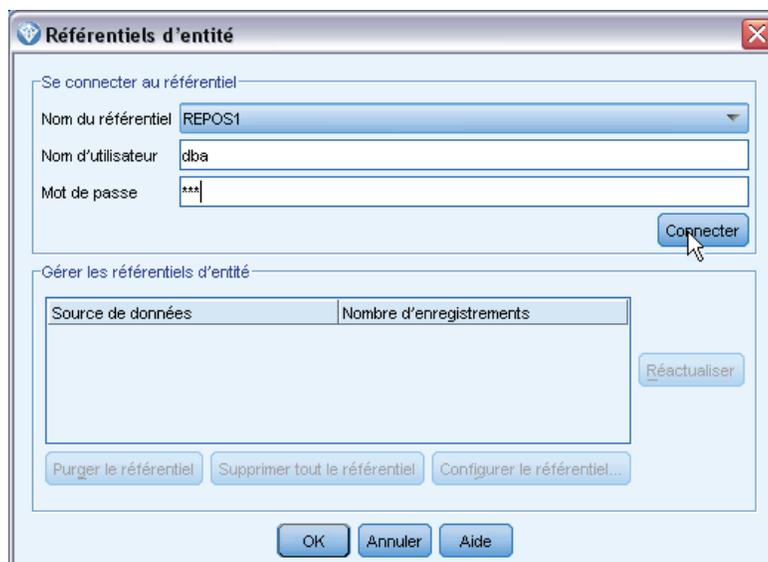
Lorsque vous avez configuré un référentiel, vous pouvez conserver son contenu de plusieurs façons. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) dans le chapitre 3 sur p. 22.

Étape 3 : Connecter SPSS Modeler au référentiel

Après avoir créé le référentiel, connectez-le au flux SPSS Modeler.

Figure 2-5

Connexion au référentiel



Pour plus d'informations, reportez-vous à la section [Options du référentiel d'entité](#) dans le chapitre 3 sur p. 18.

Étape 4 : mapper les champs d'entrée sur les fonctionnalités du référentiel

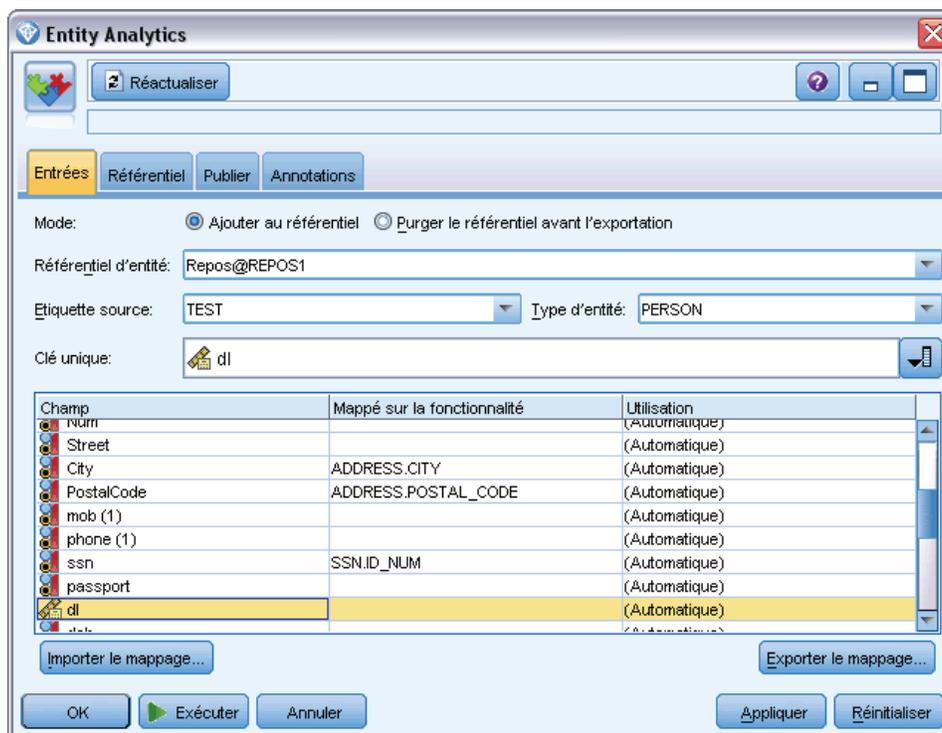
Les sources de données peuvent contenir de nombreux types d'informations sur les entités. Certains types d'informations sont communs à la majorité des sources de données d'entités, alors que d'autres peuvent être spécifiques à une source de données particulière. Dans un référentiel d'entités, ces différents types d'informations s'appellent des **fonctionnalités**. Le référentiel propose de nombreuses fonctionnalités standard et vous pouvez également créer vos propres fonctionnalités.

Une fonction de référentiel est un type d'informations individuel pouvant être utilisé avec une source de données d'entités. Certaines fonctions (par exemple, Prénom, Nom, Date de naissance, etc.) peuvent être utilisées avec de nombreuses sources de données différentes alors que d'autres fonctions seront spécifiques à une source de données particulière. Une fonction est généralement

l'équivalent d'un champ dans un enregistrement de données ou d'une colonne dans une table de base de données.

Lorsque vous avez créé un référentiel et que vous y êtes connecté, désignez un champ de vos données d'entrée comme champ **clé unique** qui sera ensuite utilisé dans les analyses. Mappez également les champs de données d'entrée sur leurs fonctionnalités correspondantes dans le référentiel. Cela permet d'éviter les doublons dans le référentiel lorsque plusieurs sources de données utilisent différents noms pour les champs contenant le même type d'informations. Le nœud Export EA propose une table de mappage dans laquelle vous pouvez créer des mappages.

Figure 2-6
Mappages des champs d'entrée sur les fonctionnalités

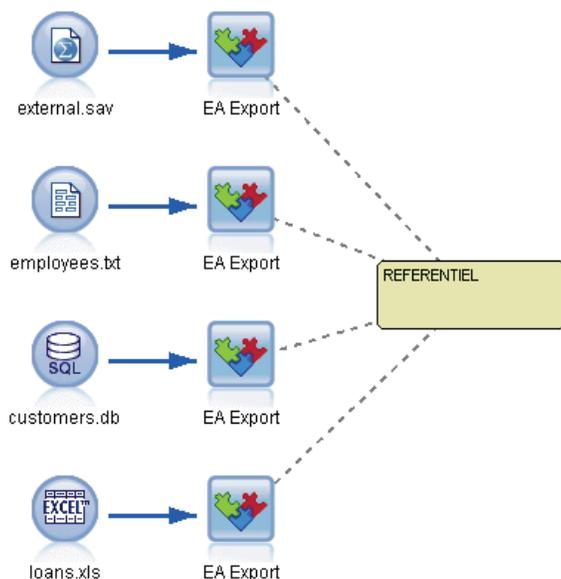


Pour plus d'informations, reportez-vous à la section [Mappage de champs d'entrées sur les fonctionnalités \(nœud Export EA\)](#) dans le chapitre 3 sur p. 19.

Étape 5 : exporter les données vers le référentiel et résoudre les conflits

Chaque nœud source de données nécessite son propre nœud Export EA, par conséquent, si vos données sont éparpillées entre plusieurs sources différentes, votre flux pourrait ressembler à cela.

Figure 2-7
Exportation des données dans le référentiel à partir de plusieurs sources de données



Vous pouvez choisir de lire les enregistrements d'une, de plusieurs ou de toutes vos sources de données si vous en avez plusieurs. Le système Entity Analytics analyse les enregistrements que vous sélectionnez et ajoute un champ d'identifiant nommé \$EA_ID à chacun d'eux. Là où deux enregistrements ou plus, associés à des identités ambiguës précédentes peuvent désormais être résolus, les identifiants ajoutés à ces enregistrements sont uniques dans tout le référentiel. Le système ajoute également un champ qui affiche la source de données d'où provient l'enregistrement.

Connectez chaque nœud source de données à son propre nœud Export EA, mappez les champs d'entrée sur les fonctionnalités du référentiel puis exécutez le flux pour exporter les données de SPSS Modeler au référentiel et résoudre les conflits d'identités en une seule opération. Pour mieux comprendre comment cela fonctionne, imaginez que vous disposez des enregistrements suivants dans quatre sources de données différentes.

Données externes

Nom	Téléphone	Risque de crédit
Mike	555-1234	560
Joe	555-4567	780

Employés

Nom	Adresse	Téléphone
Michael	1234 5th Street	555-1234
Fred	543 1st Avenue	555-9876

Clients

Nom	Adresse	Epargne
Susan	1234 5th Street	\$1234
Joe	777 Oak Street	\$5

Prêts

Nom	Adresse	Téléphone	Prêt
Sue	1234 5th Street	555-1234	\$10,000
Joseph	777 Oak Street	555-4567	\$50,000

Comme nous l'avons vu, vous exportez chaque source de données tour à tour dans le référentiel. Ce faisant, le référentiel met à jour la résolution de chaque enregistrement. Dans le référentiel, chaque enregistrement est précédé d'un champ d'identifiant (appelé *SEA ID*) et d'un champ d'indicateur source (appelé *SEA-SRC*), qui affiche la source de données d'où provient l'enregistrement. Ainsi dans notre exemple, lorsque vous avez exporté les quatre sources de données, le contenu du référentiel ressemble à ceci.

Table 2-1

Exemple de contenu de référentiel après l'exportation

SEA-ID	SEA-SRC	Nom	Téléphone	Adresse	Risque de crédit	Epargne	Prêt
1	Employés	Michael	555-1234	1234 5th St			
1	External	Mike	555-1234		560		
2	Clients	Joe		777 Oak St		\$5	
2	External	Joe	555-4567		780		
2	Prêts	Joseph	555-4567	777 Oak St			\$50,000
3	Employés	Fred	555-9876	543 1st Ave			
4	Clients	Susan		1234 5th St		\$1234	
4	Prêts	Sue	555-1234	1234 5th St			\$10,000

Le système Entity Analytics a déterminé que *Mike* dans l'ensemble de données *externe* est la même personne que *Michael* dans l'ensemble de données *Employés* en fonctionnalité d'un numéro de téléphone commun et lui a attribué l'ID 1.

Le cas de *Joe* dans l'ensemble de données *externe* est un peu plus compliqué. Est-il la même personne que le *Joe* dans *Clients* ? Il est impossible de répondre à partir de deux sources de données uniquement mais nous disposons d'une troisième source, *Prêts*, qui contient un *Joseph*. Nous avons maintenant une correspondance : le numéro de téléphone de Joseph est le même que celui de Joe dans l'ensemble de données *externe*. En fonctionnalité de cette information, le système détermine qu'ils sont la même personne et leur donne l'identifiant 2.

Il n'existe qu'un seul enregistrement pour *Fred* qui reçoit l'ID 3. *Susan* de *Clients* est identifiée comme la même personne que *Sue* de *Prêts* car elles ont la même adresse et elle reçoit l'ID 4.

Remarque : Voici un exemple de correspondance positive pour les besoins de l'illustration. Vous auriez pu choisir un ensemble de règles plus pessimiste, de sorte qu'un nom unique et un numéro de téléphone ou une adresse ne constituent pas en eux-même une correspondance exacte, et affecter aux deux enregistrements le même identifiant.

Étape 6 : analyse des identités résolues

Après avoir résolu les conflits d'identité dans le référentiel, vous pouvez maintenant analyser et traiter les résultats. Par exemple, si vous suspectez une activité frauduleuse avec l'existence d'enregistrements en double pour la même identité, vous pouvez produire un rapport contenant les doublons.

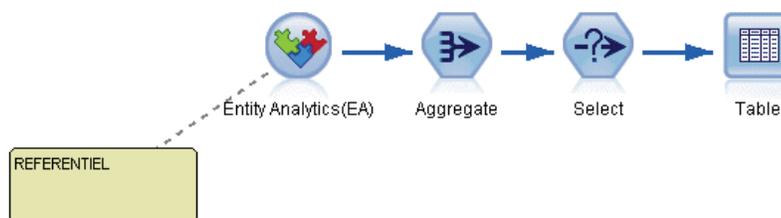
Commencez par créer un nœud source Entity Analytics (EA) puis reliez-le au référentiel.

La sortie du nœud source est composée des champs suivants.

- Le champ d'identifiant ajouté par le système (*\$EA ID* dans l'exemple de l'étape 5)
- Le champ d'indicateur source ajouté par le système (*\$EA-SRC* dans l'exemple de l'étape 5)
- Le champ clé unique que vous avez désigné à l'étape 4

Pour consulter la sortie dans SPSS Modeler, vous pouvez relier un nœud de sortie SPSS Modeler tel qu'un nœud Table ou un nœud Rapport et exécuter cette partie du flux. Si vous souhaitez résumer la sortie, qui peut être de taille importante, vous pouvez inclure des nœuds d'opérations d'enregistrement, tels que les nœuds Agréger ou Sélectionner, dans l'illustration suivante.

Figure 2-8
Consultation de la sortie



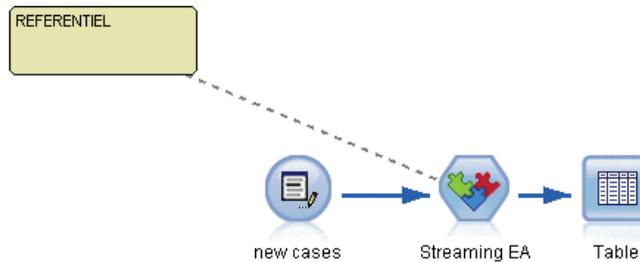
Le nœud source Entity Analytics(EA) est entièrement décrit plus tard. Pour plus d'informations, reportez-vous à la section [Analyse des identités résolues \(nœud source Entity Analytics \(EA\)\)](#) dans le chapitre 3 sur p. 32.

Étape 7 : résoudre de nouvelles observations en fonctionnalité du référentiel

Vous avez maintenant résolu les identités de tous les enregistrements dans toutes vos sources de données. Que se passe-t-il si vous souhaitez comparer un ensemble de nouveaux enregistrements pour vérifier leur relation aux données que vous possédez déjà, afin de réaliser un meilleur scoring ? C'est là que le nœud Mise en flux EA entre en jeu.

D'abord, ajoutez un nouveau nœud source de données SPSS Modeler pour lire vos nouvelles données dans le flux. Ensuite, connectez ce nœud source à un nœud Mise en flux EA. Pour consulter la sortie, ajoutez un nœud Table comme précédemment, pour que cette partie du flux ressemble à ceci.

Figure 2-9
Résolution de nouvelles observations



Lorsque vous exécutez cette partie du flux, le nœud Mise en flux EA lit chaque nouvel enregistrement et le compare au contenu du référentiel. S'il trouve des enregistrements correspondants dans le référentiel, le nœud Mise en flux EA produit tous les enregistrements correspondants avec le nouvel enregistrement auxquels il ajoute les champs ID et d'indicateur source. Si aucune correspondance n'est trouvée, le nœud de processus génère uniquement le nouvel enregistrement avec les champs ID et d'indicateur source en plus.

Pour illustrer cette situation, imaginons que le référentiel est constitué de contenu qui a été produit par le nœud source Entity Analytics (EA). Consultez [Table 2-1](#) sur p. 10.

Maintenant, nous recevons les nouveaux enregistrements suivants. Sont-ils associés à quelqu'un que nous connaissons déjà ?

Table 2-2
Nouveaux enregistrements à évaluer

Nom	Adresse	Téléphone	Prêt
Suzan	1234 5th Street	555-1234	\$100,000
Mark	888 9th Ave	555-9999	\$60,000

En comparant les nouvelles données au contenu du référentiel existant, le nœud Mise en flux EA met en correspondance le premier nouvel enregistrement avec la personne possédant l'identifiant 4 dans les enregistrements existants. Cependant, aucune correspondance n'est trouvée pour le deuxième nouvel enregistrement qui reçoit un nouvel identifiant unique 5.

Le nœud Mise en flux EA ajoute les champs d'identifiant et d'indicateur source et génère les nouveaux enregistrements avec leurs enregistrements correspondants. Par conséquent, la sortie ressemblera à ceci.

Table 2-3
Sortie du nœud Mise en flux EA

SEA-ID	SEA-SRC	Nom	Téléphone	Adresse	Risque de crédit	Epargne	Prêt
4	Client	Susan		1234 5th St		\$1234	
4	Prêt	Sue	555-1234	1234 5th St			\$10,000
4	Nouveau prêt	Suzan	555-1234	1234 5th Street			\$100,000
5	Nouveau prêt	Mark	555-9999	888 9th Ave			\$60,000

Cette sortie peut ensuite être agrégée par identifiant d'analyses d'entités et transmise à d'autres nœuds en aval pour un traitement ultérieur.

Le nœud Mise en flux EA est décrit entièrement plus tard.

Étape 8 : générer des alertes

Encore une fois, une activité potentiellement suspecte peut apparaître. Dans ce cas, la personne avec l'identifiant 4 a déjà fait l'objet d'un prêt de 10 000 \$ et demande maintenant un autre prêt de 10 fois cette somme en utilisant un nom légèrement différent. Bien entendu, cela peut être tout à fait acceptable et n'avoir pas été fait dans un but frauduleux. Toutefois, si une telle activité est suspecte selon vos règles de marché, il peut être nécessaire d'examiner ces données plus en détails.

Vous pouvez, par exemple, joindre et exécuter un nœud Table SPSS Modeler ou un nœud Rapport, imprimer le contenu de sa fenêtre de sortie et le faire lire à quelqu'un pour que des alertes soient générées manuellement. Vous pouvez également transmettre la sortie du nœud Mise en flux EA vers un modèle d'évaluation des risques que vous avez créé précédemment dans IBM® SPSS® Modeler afin de produire un ensemble de scores qui reflète mieux vos règles de marché. Une autre possibilité est d'exporter la sortie vers une base de données ou un autre support pour un traitement ultérieur. Avec SPSS Modeler, vous disposez d'un large choix d'actions correspondant à vos besoins spécifiques.

Tâches des analyses d'entités

À propos des tâches

Cette section décrit les tâches d'analyses d'entités suivantes.

- Définition d'un référentiel d'entité
- Configuration d'un référentiel d'entité
- analyse des identités résolues
- Résolution de nouvelles observations en fonctionnalité du référentiel d'entité
- Purge d'un référentiel d'entité
- Suppression d'un référentiel d'entité
- Administration des analyses d'entités

Définition d'un référentiel d'entité (nœud Export EA)

Le processus de définition d'un référentiel d'entité est composé des tâches suivantes.

1. Se connecter à une source de données. Pour plus d'informations, reportez-vous à la section [Connexion à une source de données](#) sur p. 15.
2. Créer le référentiel. Pour plus d'informations, reportez-vous à la section [Création d'un référentiel](#) sur p. 16.
3. Mapper les champs d'entrées dans la source de données sur les fonctionnalités du référentiel. Pour plus d'informations, reportez-vous à la section [Mappage de champs d'entrées sur les fonctionnalités \(nœud Export EA\)](#) sur p. 19.

Lorsque vous avez défini les mappages, vous pouvez les afficher pour la source de données actuelle ou pour toutes les sources de données connues du référentiel. Pour plus d'informations, reportez-vous à la section [Affichage des mappages de champs \(nœud Export EA\)](#) sur p. 22.

Le référentiel d'entité

Le référentiel fournit une zone de stockage centrale, qui tient lieu de cache de données pour toutes les informations sur les entités. Parce que le référentiel s'exécute en temps réel, il n'a qu'un seul état. Par conséquent, le concept de version n'existe pas dans un référentiel d'entité. Le référentiel contient l'état actuel de toutes les données d'entrées, par conséquent, il est possible qu'il atteigne une taille importante.

Vous pouvez gérer le contenu du référentiel au moyen d'une interface graphique simple. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.

Remarque : La version de IBM® SPSS® Modeler Entity Analytics fournie avec IBM® SPSS® Modeler Premium prend en charge un seul référentiel hébergé sur le produit IBM solidDB qui est livré avec IBM SPSS Modeler Entity Analytics. Dans cette version, vous devez supprimer le référentiel existant avant de pouvoir en créer un nouveau. Une mise à niveau vers IBM SPSS Modeler Entity Analytics faisant l'objet d'une licence distincte (appelée IBM SPSS Modeler Entity Analytics détachée) est disponible et vous permet de faire coexister plusieurs référentiels sur le même système ; chaque référentiel pouvant contenir jusqu'à 10 millions de lignes et utiliser plus de quatre cœurs de processeur. Contactez votre représentant local de l'assistance technique d'IBM pour plus d'informations.

Connexion à une source de données

Commencez par lire vos données source dans SPSS Modeler à l'aide d'un nœud source.

Pour se connecter à une source de données

1. Dans l'onglet Sources de la palette de nœuds au bas de la fenêtre principale de SPSS Modeler, faites un double clic sur une icône correspondant au type des données source. Cette action ajoute un nœud source à l'écran.
2. Sur l'écran, faites un double clic sur l'icône pour ouvrir la boîte de dialogue.
3. Dans le champ Fichier, saisissez l'emplacement et le nom du fichier de données source.
4. Remplissez le reste de la boîte de dialogue selon les besoins (cliquez sur le bouton Aide pour plus d'informations), puis cliquez sur OK.
5. Si le fichier de données source ne contient pas de champ qui identifie chaque enregistrement de manière unique, ajoutez-en un à l'aide du nœud Calculer. Pour plus d'informations, reportez-vous à la section [Ajout d'un identifiant d'enregistrement unique](#) sur p. 15.

Ajout d'un identifiant d'enregistrement unique

Chaque fichier source de données qui est une entrée du référentiel d'entités doit contenir un champ qui identifie chaque enregistrement de manière unique. Si un fichier source de données ne contient pas ce champ, vous pouvez en ajouter un à l'aide du nœud Calculer SPSS Modeler.

Pour ajouter un identifiant d'enregistrement unique à un fichier source de données

1. Sur l'écran, cliquez sur le nœud source que vous avez ajouté lors de la tâche précédente.
2. Dans l'onglet Options de champ de la palette de nœuds, faites un double clic sur l'icône Calculer pour joindre un nœud Calculer au nœud source.
3. Sur l'écran, faites un double clic sur le nœud Calculer pour ouvrir sa boîte de dialogue.
4. Dans le champ Calculer, remplacez le nom par défaut par un nom significatif (comme ID) pour le champ d'identifiant que vous ajoutez.
5. Vérifiez que le champ Calculer comme est défini sur Formule

6. Définissez le Type de champ sur Continu.
7. Dans la zone de texte Formule, saisissez @INDEX et cliquez sur OK.

Création d'un référentiel

Vous devez créer un référentiel pour stocker les données d'entrée.

Pour créer un référentiel

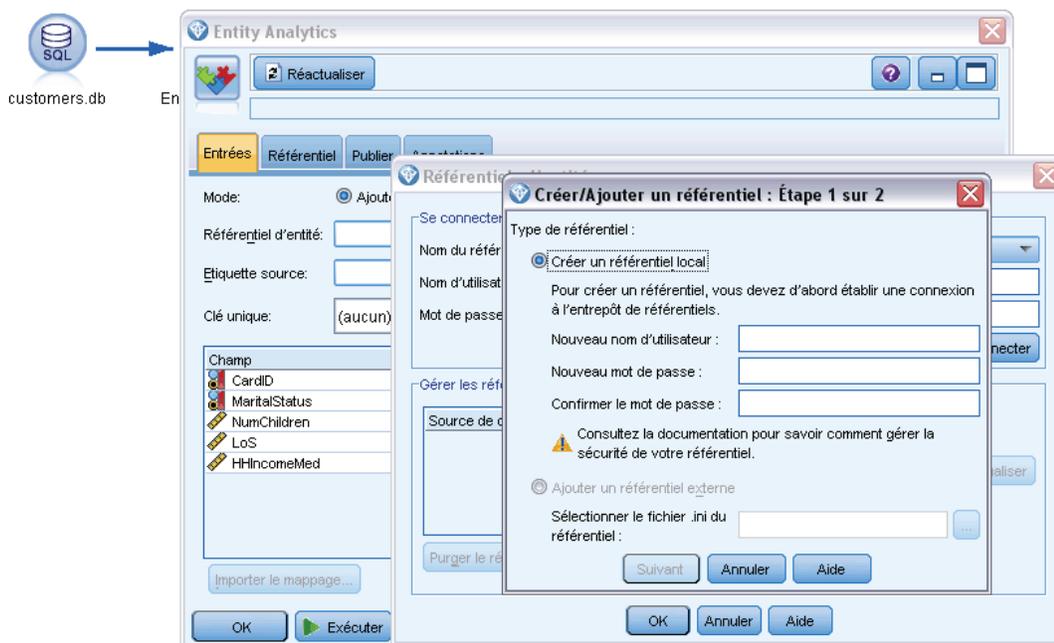
1. Dans l'onglet Exporter de la palette de nœuds SPSS Modeler, placez un nœud Export EA dans l'espace de travail du flux.

Remarque : Si vous créez un référentiel pour la première fois, utilisez un nœud Export EA et connectez-le au nœud source SPSS Modeler qui contient les données à ajouter au référentiel (ou au nœud Calculer, si vous en avez ajouté un pour obtenir un champ d'identifiant unique). Pour connecter les nœuds, effectuez les actions suivantes :

- Faites un clic droit sur le nœud source SPSS Modeler.
 - Choisissez Connecter.
 - Cliquez sur le nœud Export EA.
2. Faites un double clic sur le nœud Export EA pour ouvrir sa boîte de dialogue.
 3. Cliquez sur la liste Référentiel d'entité.
 4. Cliquez sur <Parcourir...> pour afficher la boîte de dialogue Référentiels d'entités.
 5. Dans la boîte de dialogue Référentiels d'entités, cliquez sur le champ Nom du référentiel.
 6. Choisissez <Créer un nouveau référentiel...> pour afficher l'assistant Créer un référentiel.

Créer l'assistant Référentiel

Figure 3-1
Créer l'assistant Référentiel



Étape 1

Au cours de cette étape, vous choisissez ou non de créer un référentiel local, à l'aide du produit IBM solidDB fourni avec IBM SPSS Modeler Entity Analytics, ou d'utiliser une base de données externe comme référentiel.

Créer le référentiel local. Indiquez le nom d'utilisateur et le mot de passe de l'administrateur de la base de données IBM solidDB qui doit héberger le référentiel que vous êtes en train de créer. Confirmez le mot de passe et cliquez sur Suivant.

Remarque : si vous avez ensuite besoin de modifier les identifiants de l'administrateur, vous pouvez le faire à l'aide de l'éditeur de ligne de commande de la base de données. Pour plus d'informations, reportez-vous à la section [Gestion des identifiants de l'administrateur pour la base de données du référentiel](#) sur p. 42.

Ajouter un référentiel externe. Utilisez cette option si vous souhaitez vous servir d'une base de données externe pour héberger le référentiel. Indiquez l'emplacement du fichier .ini de la base de données dans le champ Sélectionner le fichier .ini du référentiel et cliquez sur Suivant.

Étape 2

Nouveau nom de référentiel. Saisissez un nom unique pour le nouveau référentiel.

Importer la configuration de. (Pour le référentiel local uniquement) Si vous souhaitez baser la configuration de ce référentiel sur un référentiel existant, choisissez le référentiel ici, sinon choisissez Par défaut. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.

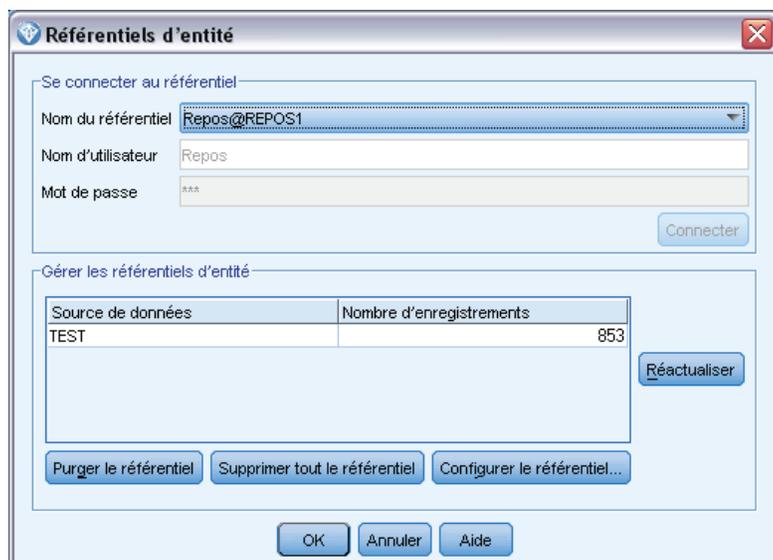
Si vous choisissez un référentiel existant, saisissez les informations de connexion si elles sont différentes de celles saisies sur l'écran précédent.

Cliquez sur OK pour créer le nouveau référentiel et afficher la boîte de dialogue Instances de résolution d'entités dans laquelle vous pourrez vous connecter au référentiel.

Options du référentiel d'entité

La boîte de dialogue Référentiels d'entités contient un certain nombre d'options permettant de créer, de se connecter à, de configurer et de gérer un référentiel d'entités.

Figure 3-2
Options du référentiel d'entité



Se connecter au référentiel Utilisez ces options pour créer un nouveau référentiel d'entités ou pour vous connecter à un référentiel existant.

- **Nom du référentiel.** Affiche le référentiel d'entités actuel s'il existe. Pour choisir un autre référentiel s'il en existe plusieurs, sélectionnez-en un dans la liste.
Pour créer un nouveau référentiel, sélectionnez <Créer un nouveau référentiel...>. Cette action lance un assistant qui vous guidera dans le processus de création.
- **Nom d'utilisateur.** Saisissez votre nom d'utilisateur pour le référentiel sélectionné.
- **Mot de passe.** le mot de passe correspondant à ce nom d'utilisateur.
- **Connecter.** Cliquez pour vous connecter au référentiel actuel.

Gérer les référentiels d'entité. Le tableau répertorie les sources de données chargées dans le référentiel actuel (celui auquel vous êtes connecté) et affiche le nombre d'enregistrements dans chaque source de données.

- **Actualiser.** Met à jour la source de données et les informations sur sa taille dans le tableau, par exemple si vous avez ajouté une nouvelle source de données ou si vous avez modifié la taille d'une source de données existante.
- **Purger le référentiel.** Supprime toutes les données source du référentiel mais conserve toutes les informations de configuration. Vous pouvez utiliser cette option si les informations de configuration sont encore utiles, mais vous souhaitez supprimer tous les enregistrements de données du référentiel. Pour plus d'informations, reportez-vous à la section [Purge d'un référentiel d'entité](#) sur p. 44.

Supprimer le référentiel en entier. Détruit entièrement le contenu du référentiel actuel et ses informations de configuration. Pour plus d'informations, reportez-vous à la section [Suppression d'un référentiel d'entité](#) sur p. 44.

Configurer le référentiel. Affiche une fenêtre dans laquelle vous pouvez configurer le référentiel actuel. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.

Mappage de champs d'entrées sur les fonctionnalités (nœud Export EA)

Le référentiel propose un nombre de fonctionnalités prédéfinies standard. Des sources de données différentes peuvent utiliser des noms de champ différents, Address1 ou Address Line 1) pour des types d'informations qui correspondent à la même fonctionnalité. Pour éviter les doublons, il est nécessaire de mapper les champs de source d'entrée sur les fonctionnalités spécifiques du référentiel. Vous n'avez pas besoin de mapper chaque champ de l'ensemble de données, seuls ceux susceptibles de correspondre à la même fonctionnalité dans d'autres ensembles de données.

Lorsqu'une source de données utilise des champs correspondant à d'autres types d'informations qui ne sont pas prédéfinis dans le référentiel, vous pouvez créer de nouvelles fonctionnalités depuis la fenêtre Configuration du référentiel. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.

Pour mapper des champs d'entrée sur les fonctionnalités

1. Reliez un nœud Export EA à un nœud source de données sur l'espace de travail du flux. Chaque nœud source de données que vous utilisez doit être relié à son propre nœud Export EA.
2. Ouvrez le nœud Export EA pour afficher l'onglet Entrées qui contient les options de mappage des champs d'entrée. Pour plus d'informations, reportez-vous à la section [Options d'entrée du référentiel pour le mappage](#) sur p. 20.
3. Dans le nœud Export EA, sélectionnez l'onglet Référentiel pour consulter les tâches de mappage de la source de données actuelle ou de toutes les sources de données si vous en utilisez plusieurs.
4. Pour enregistrer un ensemble de tâches de mappage (par exemple, à utiliser avec un autre nœud source de données), cliquez sur Exporter le mappage.

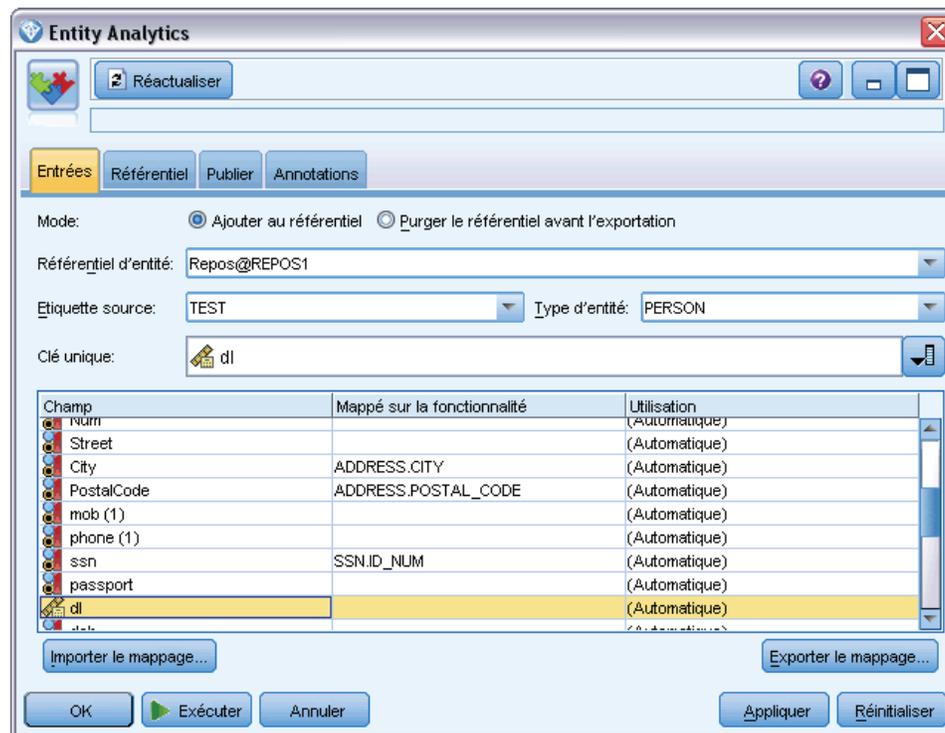
Lorsque vous avez terminé le mappage du premier nœud source de données, répétez le processus pour tous les autres nœuds source de données que vous souhaitez utiliser.

Options d'entrée du référentiel pour le mappage

L'onglet Entrées contient les options de mappage des champs source de données sur les fonctionnalités du référentiel prêtes à être exportées vers le référentiel. Configurez les tâches de mappage dans cet onglet, en option, cliquez sur l'onglet Référentiel pour afficher le mappage d'autres sources de données, puis cliquez sur Exécuter pour exporter les données dans le référentiel.

Si vous avez déjà stocké un ensemble de mappages dans un fichier XML, vous pouvez les utiliser en cliquant sur Importer le mappage.

Figure 3-3
Mappages des champs d'entrée sur les fonctionnalités



Mode. Gardez la sélection par défaut, Ajouter au référentiel, si vous souhaitez ajouter les enregistrements du fichier source au contenu existant dans le référentiel. Si vous souhaitez effacer le contenu du référentiel mais conserver les informations de configuration avant d'ajouter les enregistrements source, utilisez l'option Purger le référentiel avant l'exportation.

Référentiel d'entité. Affiche le référentiel d'entités actuel s'il existe. Pour choisir un autre référentiel s'il en existe plusieurs, sélectionnez-en un dans la liste. Pour créer un nouveau référentiel, choisissez <Parcourir...> pour afficher une boîte de dialogue à partir de laquelle vous pourrez créer le référentiel. Pour plus d'informations, reportez-vous à la section [Options du référentiel d'entité](#) sur p. 18.

Balise source. Une liste des balises indiquant les sources de données actuellement connues par le référentiel. En choisir une dans la liste ou choisir <Ajouter une nouvelle balise source...> pour créer une balise pour une nouvelle source de données.

Type d'entité. Une liste de types d'entité (c'est-à-dire des ensembles de fonctionnalités) définis dans le référentiel. En choisir un dans la liste ou choisir <Ajouter un nouveau type d'entité...> pour afficher la fenêtre de configuration du référentiel, où vous pouvez définir un nouveau type d'entité. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.

Clé unique. (obligatoire) Le champ d'entrée à utiliser pour les identifiants uniques des enregistrements de données.

Table de mappage. Dans cette table, vous pouvez mapper chaque champ d'entrée sur une fonctionnalité correspondante dans le référentiel. Si aucune fonctionnalité adaptée n'existe dans le type d'entité sélectionné, vous pouvez créer une nouvelle fonctionnalité ici.

- **Champ.** L'ensemble des champs d'entrée dans la source de données sélectionnée. Chaque champ contient une icône indiquant le niveau de mesure (c'est-à-dire le type de données) du champ.
- **Mappé sur fonctionnalité.** Pour mapper un champ sur une fonctionnalité, faites un double clic sur cette colonne (ou appuyez sur la barre d'espace) sur la ligne du champ et choisissez une fonctionnalité dans la liste. Si aucune fonctionnalité appropriée n'est disponible, choisissez <Ajouter une nouvelle fonctionnalité...> pour afficher la fenêtre de configuration du référentiel où vous pouvez définir une nouvelle fonctionnalité pour ce type d'entité. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.
- **Utilisation.** Indique le contexte d'un champ spécifique où plusieurs contextes sont possibles, par exemple, des numéros de téléphone du bureau ou du domicile. Des types d'utilisation prédéfinis sont disponibles pour les fonctionnalités ADRESSE et TÉLÉPHONE et vous pouvez créer vos propres types d'utilisation pour toutes les fonctionnalités. Pour définir une utilisation autre que celle par défaut (Auto), cliquez sur la ligne souhaitée de cette colonne et choisissez un des types d'utilisation existants (le cas échéant) ou cliquez sur <Ajouter l'utilisation...> pour en créer une nouvelle. Pour plus d'informations, reportez-vous à la section [Gestion des types d'entités](#) sur p. 28.

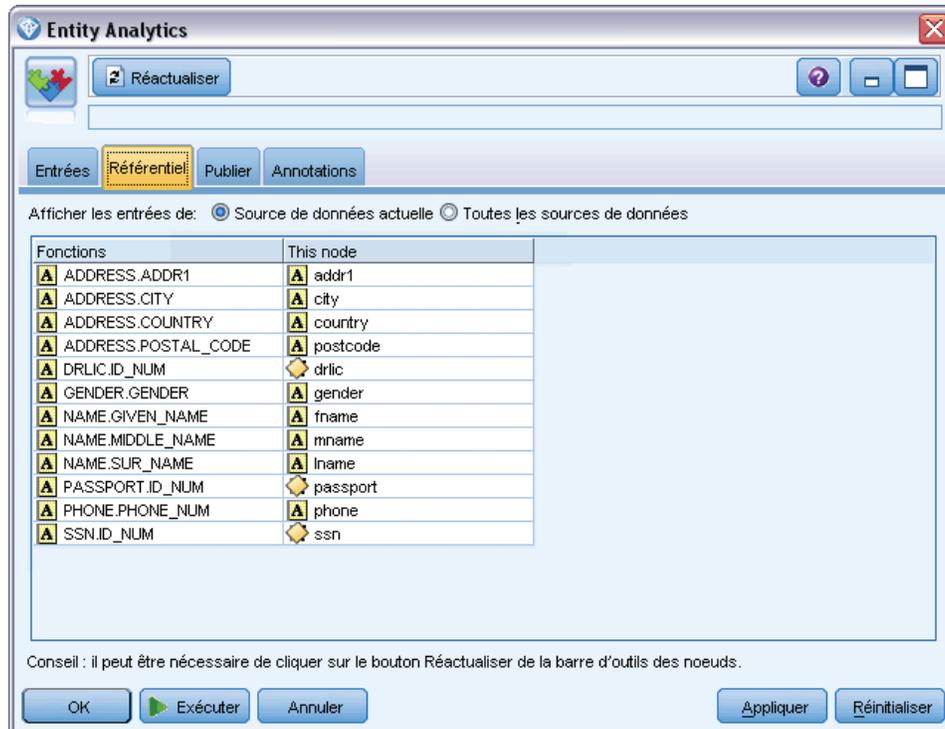
Importer le mappage. Importe un ensemble précédemment exporté de mappages de champs sur fonctionnalités depuis un fichier XML externe. Cette fonctionnalité peut être utile si vous disposez de différentes sources de données avec les mêmes paramètres de mappage car elle permet de ne pas avoir à redéfinir les mêmes mappages pour différentes sources.

Exporter le mappage. Exporte vers un fichier XML externe l'ensemble des mappages champs sur fonctionnalités affiché dans la table de mappage.

Affichage des mappages de champs (nœud Export EA)

Sur l'onglet Référentiel, cliquez sur le bouton Rafraîchir pour voir les fonctionnalités du référentiel sur lesquelles des champs d'entrée sont mappés. Vous pouvez faire cela pour la source de données actuelle (celle contrôlée par le nœud source relié à ce nœud Export) ou pour toutes les sources de données.

Figure 3-4
Affichage des mappages de champs



Afficher les entrées pour. Choisissez une option pour afficher les mappages de la source de données actuelle ou de toutes les sources de données connues du référentiel.

Actualiser. Met à jour l'affichage de l'option d'entrée sélectionnée.

Caractéristiques. Une liste de toutes les fonctionnalités ayant des mappages dans les sources de données affichées. Les fonctionnalités non mappées ne sont pas affichées.

<Source de données>. Chaque colonne répertorie les champs mappés dans une source de données spécifique pour chaque fonctionnalité pour laquelle un mappage a été défini.

Configuration d'un référentiel d'entité

Vous pouvez gérer le contenu du référentiel depuis la fenêtre Configuration du référentiel qui propose une interface visuelle simple pour le référentiel entier.

si vous pensez utiliser plusieurs référentiels avec les mêmes configurations ou des configurations similaires, vous pouvez définir une configuration de base et l'exporter vers un fichier que vous pourrez ensuite importer dans d'autres référentiels. Pour plus d'informations, reportez-vous à la section [Réutilisation d'une configuration de référentiel](#) sur p. 31.

Avertissement : Si vous souhaitez modifier et enregistrer la configuration d'un référentiel qui contient déjà des données, nous vous recommandons fortement de purger le contenu du référentiel et de charger de nouveau les données. Cela évite de conserver des incohérences dans le référentiel.

Pour définir une configuration de référentiel

1. Ouvrez un nœud Entity Analytics.
2. Cliquez sur la liste Référentiel d'entité.
3. Cliquez sur <Parcourir...> pour afficher la boîte de dialogue Instances de résolution d'entités.
4. Dans la boîte de dialogue Instances de résolution d'entités, cliquez sur la liste Nom du référentiel.
5. Sélectionnez le référentiel pour lequel vous souhaitez définir la configuration.
6. Si vous n'êtes pas encore connecté, saisissez le nom et le mot de passe administrateur et cliquez sur Se connecter.
7. Lorsque le bouton Configurer le référentiel est activé, cliquez dessus pour afficher la fenêtre Configuration du référentiel.
8. Créez les informations de configuration comme expliqué dans les sections suivantes.

Le volet de navigation à gauche de la fenêtre Configuration du référentiel contient une structure en arborescence à partir de laquelle vous pouvez gérer les différentes caractéristiques du référentiel.

Table 3-1

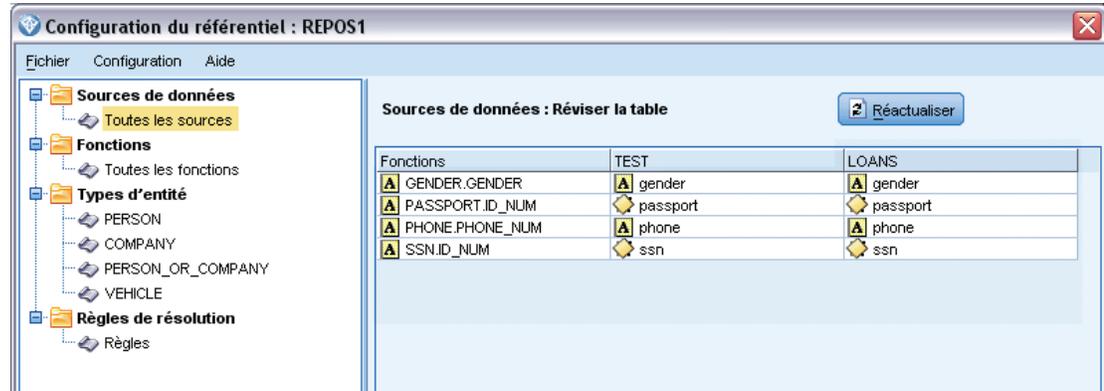
Éléments principaux de la fenêtre Configuration du référentiel

Section	Description	
Sources de données	Affichez les mappages de toutes les sources de données sur les différentes fonctionnalités du référentiel.	Pour plus d'informations, reportez-vous à la section Affichage des mappages de sources de données sur p. 24.
fonctionnalités	Créez une nouvelle fonctionnalité, ou dupliquez, modifiez ou supprimez une fonctionnalité existante.	Pour plus d'informations, reportez-vous à la section Gestion des fonctionnalités du référentiel sur p. 24.
Types d'entités	Créez un nouveau type d'entité ou gérez des types existants (dupliquez, renommez, reliez ou supprimez des fonctionnalités).	Pour plus d'informations, reportez-vous à la section Gestion des types d'entités sur p. 28.
Règles de résolution	Définissez le seuil de la mise en correspondance d'entités.	Pour plus d'informations, reportez-vous à la section Définition du seuil de mise en correspondance des entités sur p. 31.

Affichage des mappages de sources de données

Dans la section Sources de données de la fenêtre Configuration du référentiel, l'entrée Toutes les sources propose un affichage en lecture seule des mappages de toutes les sources de données sur les différentes fonctionnalités du référentiel.

Figure 3-5
Mappages des sources de données dans la fenêtre de configuration



Cliquez sur Actualiser pour mettre à jour la liste si de nouvelles sources de données ont été ajoutées au référentiel.

Remarque : vous ne pouvez pas ajouter de source de données au référentiel ici. Les sources de données peuvent uniquement être ajoutées en créant un nœud source SPSS Modeler et en le connectant à un nœud Export Entity Analytics. Pour plus d'informations, reportez-vous à la section [Connexion à une source de données](#) sur p. 15.

Gestion des fonctionnalités du référentiel

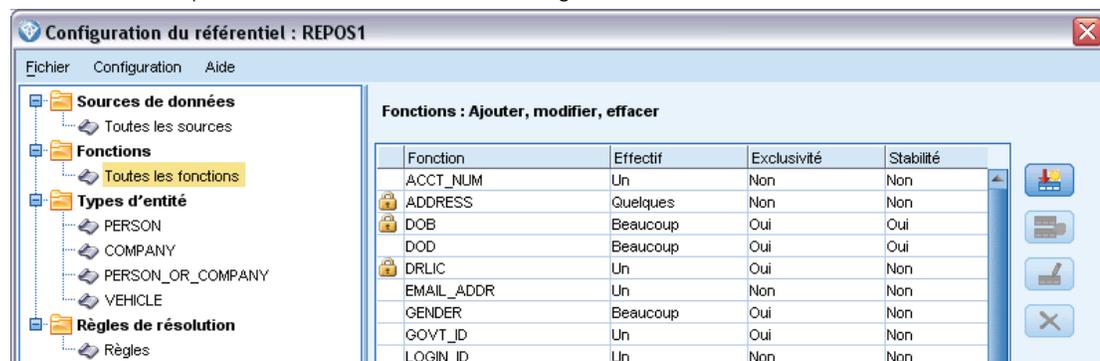
Une fonctionnalité de référentiel est un type d'informations individuel pouvant être utilisé avec une source de données d'entités. Certaines fonctionnalités (par exemple, Prénom, Nom, Date de naissance, etc.) peuvent être utilisées avec de nombreuses sources de données différentes alors que d'autres fonctionnalités seront spécifiques à une source de données particulière. Une fonctionnalité peut contenir un ou plusieurs éléments ; chaque élément est généralement l'équivalent d'un champ dans un enregistrement de données ou d'une colonne dans une table de base de données.

Dans la section fonctionnalités de la fenêtre Configuration du référentiel, l'entrée Toutes les fonctionnalités propose un moyen de gérer toutes les fonctionnalités du référentiel. Vous pouvez effectuer les actions suivantes.

- Créer une nouvelle fonctionnalité.
- Dupliquer une fonctionnalité existante (par exemple, créer une nouvelle fonctionnalité basée sur une fonctionnalité existante)
- Modifier une fonctionnalité existante
- Supprimer une fonctionnalité existante

Les instructions pour ces tâches sont données plus loin dans cette section.

Figure 3-6
fonctionnalités répertoriées dans la fenêtre de configuration



La liste des fonctionnalités présente toutes les fonctionnalités qui ont été définies dans ce référentiel. Les colonnes dans la liste indiquent les différentes propriétés d'une fonctionnalité.

Caractéristique. le nom de la caractéristique. Un symbole de cadenas en regard du nom d'une fonctionnalité indique que la fonctionnalité est verrouillée. Les fonctionnalités verrouillées ne peuvent être ni supprimées, ni dupliquées et il est impossible d'y apporter des modifications.

Effectif. Indique le nombre d'entités pouvant avoir la même valeur pour cette caractéristique. Les valeurs valides sont Aucun, Nom, Un (par exemple pour un numéro de passeport), Quelques (par exemple pour une adresse) ou Nombreux (par exemple pour une date de naissance).

Exclusivité. Indique qu'une entité devrait généralement avoir un seul de ce type de fonctionnalité. Par exemple, une date de naissance ou un numéro d'identité aurait ici la valeur Oui alors que l'adresse ou un numéro de carte bancaire aurait la valeur Non (car une entité peut avoir plusieurs adresses ou cartes bancaires).

Stabilité. Indique la valeur de stabilité de cette fonction (c'est à dire, dans quelle mesure cette fonction n'est pas susceptible de subir des modifications au cours de la vie d'une entité). Par exemple, une fonction Date de naissance aura la valeur Oui car elle ne change jamais, mais une fonction Adresse aura la valeur Non car il est plus probable qu'elle change et est par conséquent moins stable. *Remarque :* Le sexe est généralement stable pendant toute une vie, mais parce qu'il est souvent mal défini en raison de données incorrectes, la configuration par défaut lui donne la valeur Non.

Pour créer une nouvelle fonctionnalité

1. Effectuez l'une des actions suivantes :
 - cliquez sur le bouton Créer une nouvelle fonctionnalité (bouton du haut à droite de l'écran).
 - Faites un clic droit sur Toutes les fonctionnalités dans le volet de navigation à gauche de l'écran et choisissez Nouvelle fonctionnalité.
2. Remplissez la boîte de dialogue Ajouter/modifier une fonctionnalité. Pour plus d'informations, reportez-vous à la section [Ajout ou modification d'une fonctionnalité](#) sur p. 26.

Pour dupliquer une fonctionnalité existante

1. Dans la colonne fonctionnalité du tableau à droite de l'écran, sélectionnez la fonctionnalité à dupliquer.
2. Cliquez sur le bouton Dupliquer la fonctionnalité sélectionnée (deuxième bouton à droite de l'écran).
3. Remplissez la boîte de dialogue Ajouter/modifier une fonctionnalité. Pour plus d'informations, reportez-vous à la section [Ajout ou modification d'une fonctionnalité](#) sur p. 26.

Pour modifier une fonctionnalité existante

Avertissement : si vous modifiez ou supprimez une fonctionnalité ou un élément de fonctionnalité alors que le référentiel contient déjà des données, vous devez alors purger le référentiel et recharger les données. Cela évite de conserver des incohérences dans le référentiel.

1. Dans la colonne fonctionnalité du tableau à droite de l'écran, sélectionnez la fonctionnalité à modifier. *Remarque* : Vous pouvez modifier uniquement les fonctionnalités que vous avez créées, et pas les fonctionnalités du système.
2. Cliquez sur le bouton Modifier la fonctionnalité sélectionnée (troisième bouton à droite de l'écran).
3. Remplissez la boîte de dialogue Ajouter/modifier une fonctionnalité. Pour plus d'informations, reportez-vous à la section [Ajout ou modification d'une fonctionnalité](#) sur p. 26.

Pour supprimer une fonctionnalité existante

Avertissement : si vous modifiez ou supprimez une fonctionnalité ou un élément de fonctionnalité alors que le référentiel contient déjà des données, vous devez alors purger le référentiel et recharger les données. Cela évite de conserver des incohérences dans le référentiel.

1. Dans la colonne fonctionnalité du tableau à droite de l'écran, sélectionnez la fonctionnalité à supprimer. *Remarque* : Vous pouvez supprimer uniquement les fonctionnalités que vous avez créées, et pas les fonctionnalités du système.
2. Effectuez l'une des actions suivantes :
 - Cliquez sur le bouton Supprimer la fonctionnalité sélectionnée (bouton du bas à droite de l'écran).
 - Faites un clic droit sur Toutes les fonctionnalités dans le volet de navigation à gauche de l'écran et choisissez Supprimer.
3. Cliquez sur Continuer pour confirmer la suppression de la fonctionnalité. *Remarque* : Il est impossible d'annuler la suppression d'une fonctionnalité.

Ajout ou modification d'une fonctionnalité

Avertissement : si vous modifiez ou supprimez une fonctionnalité ou un élément de fonctionnalité alors que le référentiel contient déjà des données, vous devez alors purger le référentiel et recharger les données. Cela évite de conserver des incohérences dans le référentiel.

Dans la boîte de dialogue Ajouter/modifier une fonctionnalité, vous pouvez créer une nouvelle fonctionnalité de référentiel, ou dupliquer ou modifier une fonctionnalité existante.

Figure 3-7
Modification d'une fonctionnalité

Élément	Description	Type de d...	Rechercher	Comparer	Description
1 ADDR_TYPE	Address t...	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2 FULL_ADDR	Full address	Chaîne	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 ADDR1	Addr line 1	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4 ADDR2	Addr line 2	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5 ADDR3	Addr line 3	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
6 CITY	City	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7 STATE	State	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8 POSTAL_...	Postal code	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9 COUNTRY	Country	Chaîne	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Type de fonctionnalité. Une étiquette indiquant le type d'informations auquel est associée la fonctionnalité. Cette étiquette forme la première partie de l'identifiant de la fonctionnalité.

Description : une courte description du type de fonctionnalité, dans un but informatif uniquement.

Effectif. Indique le nombre d'entités pouvant avoir la même valeur pour cette caractéristique. Les valeurs valides sont Aucun, Nom, Un (par exemple pour un numéro de passeport), Quelques (par exemple pour une adresse) ou Nombreux (par exemple pour une date de naissance).

Exclusivité. Indique qu'une entité devrait généralement avoir un seul de ce type de fonctionnalité. Par exemple, une date de naissance ou un numéro d'identité aurait ici la valeur Oui alors que l'adresse ou un numéro de carte bancaire aurait la valeur Non (car une entité peut avoir plusieurs adresses ou cartes bancaires).

Stabilité. Indique la valeur de stabilité de cette fonction (c'est à dire, dans quelle mesure cette fonction n'est *pas susceptible* de subir des modifications au cours de la vie d'une entité). Par exemple, une fonction Date de naissance aura la valeur Oui car elle ne change jamais, mais une fonction Adresse aura la valeur Non car il est plus probable qu'elle change et est par conséquent moins stable. *Remarque :* Le sexe est généralement stable pendant toute une vie, mais parce qu'il est souvent mal défini en raison de données incorrectes, la configuration par défaut lui donne la valeur Non.

Conserver l'historique. Si vous sélectionnez Oui (option par défaut), l'historique des modifications de la valeur du champ est conservé. Ceci est utile, par exemple, pour le champ adresse d'un client. Dans un tel cas, la conservation de l'historique permet d'améliorer la correspondance si

un client a changé d'adresse plusieurs fois. Cependant, pour un champ solde du compte par ex., vous préférerez définir cette option sur Non pour éviter de conserver de trop grandes quantités de données.

Table des éléments. Une liste des éléments que cette fonctionnalité contient.

- **Élément.** Le nom de l'élément.
- **Description :** une courte description de ce que l'élément propose.
- **Type de données.** Le type de données pouvant être utilisé pour cet élément.
- **Rechercher.** Si cette option est sélectionnée, cela signifie que cet élément de la fonctionnalité peut être recherché, tout seul ou en combinaison avec d'autres éléments de recherche. Par exemple, supposons qu'une fonctionnalité PASSEPORT contiennent deux éléments, NUM_ID et PAYS. Si vous souhaitez pouvoir effectuer une recherche uniquement sur NUM_ID, vous devez sélectionner sa case Rechercher et désélectionner la case Rechercher de PAYS. Toute modification de la fonctionnalité, tel que le pays ou l'état, la date d'émission ou de validité, ne doit pas faire partie de la recherche.
- **Comparer.** Indique quels éléments d'une fonctionnalité seront envoyés à la routine de comparaison pour la résolution des entités. Continuons avec l'exemple d'une fonctionnalité PASSEPORT qui contient deux éléments, NUM_ID et PAYS, étant donné que PAYS n'est pas toujours indiqué, seul NUM_ID sera envoyé à la routine de comparaison.
- **Description.** Indique quels éléments de la fonctionnalité doivent apparaître dans sa description interne. Dans l'exemple précédent d'une fonctionnalité PASSEPORT qui contient deux éléments, NUM_ID et PAYS, si vous souhaitez que les deux éléments apparaissent dans la description de la fonctionnalité, vous devez sélectionner la case Description des deux éléments. Supposons maintenant que la fonctionnalité contient un élément supplémentaire DT_EMISSION. Vous décocheriez la case Description correspondante pour vous assurer qu'elle n'apparaît pas dans la description de la fonctionnalité

Bouton Ajouter un nouvel élément. Ajoute une ligne à la table des éléments, pour que vous puissiez définir un nouvel élément.

Bouton Supprimer un élément. Supprime une ligne sélectionnée de la table des éléments. Vous ne pouvez pas annuler cette opération.

Avertissement : si vous modifiez ou supprimez une fonctionnalité ou un élément de fonctionnalité alors que le référentiel contient déjà des données, vous devez alors purger le référentiel et recharger les données. Cela évite de conserver des incohérences dans le référentiel.

Gestion des types d'entités.

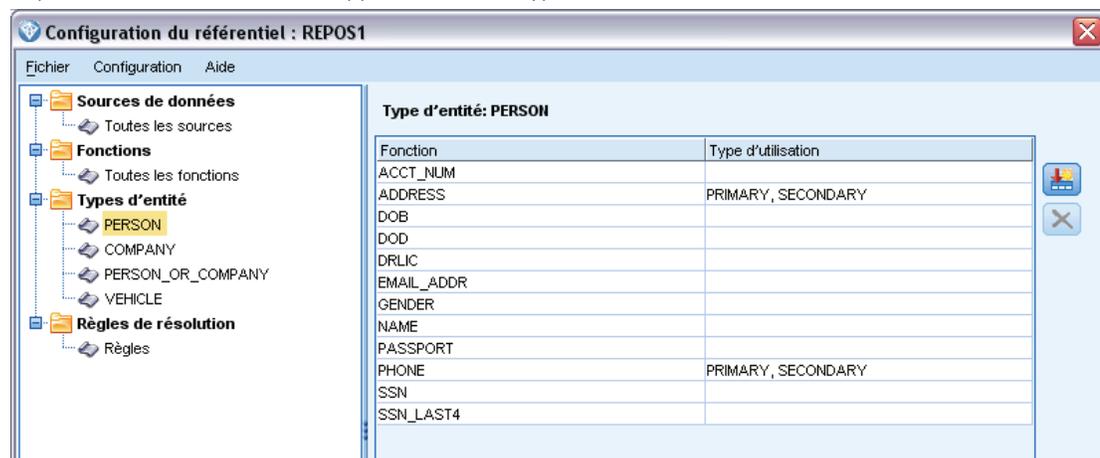
Un **type d'entité** est un ensemble nommé de fonctionnalités de référentiel qui sont regroupées par ressemblance. Par exemple, un type d'entité devant être utilisé avec un ensemble de données client peut être composé de fonctionnalités telles que Nom, Date de naissance, Sexe, Adresse, Numéro de téléphone, etc.

Le référentiel IBM SPSS Modeler Entity Analytics est fourni avec un ensemble de types d'entités standard auquel vous pouvez ajouter les vôtres.

La section Types d'entité de la fenêtre Configuration du référentiel répertorie les différents types d'entité qui ont été créés. Vous pouvez effectuer les actions suivantes.

- Créer un nouveau type d'entité
- Dupliquer un type d'entité existant (par exemple, créer un nouveau type d'entité basé sur un type existant)
- Relier les fonctionnalités à un type d'entité
- Supprimer les fonctionnalités d'un type d'entité
- Renommer un type d'entité
- Supprimer un type d'entité

Figure 3-8
Répertorier les fonctionnalités appartenant à un type d'entité



Type d'entité. Le nom du type d'entité sélectionné.

Caractéristique. La liste des fonctionnalités valides que ce type d'entité comprend.

Type d'utilisation. (En option) Indique différents contextes dans lesquels cette caractéristique peut être utilisée. Faites un double clic sur cette colonne pour ajouter ou modifier le type d'utilisation, en séparant les types d'utilisation par une virgule et un espace. Les valeurs spécifiées ici définissent les valeurs affichées sur le nœud Export EA ou le nœud de processus EA quand un utilisateur clique sur la colonne Utilisation d'une caractéristique dans l'onglet Entrées. Pour plus d'informations, reportez-vous à la section [Options d'entrée du référentiel pour le mappage](#) sur p. 20.

Pour créer un nouveau type d'entité

1. Faites un clic droit sur Types d'entité dans le volet de navigation à gauche de l'écran.
2. Choisissez Nouveau type d'entité.
3. Saisissez un nom unique pour le type d'entité et cliquez sur OK.
4. Reliez les fonctionnalités au type d'entité (voir section suivante).

Pour relier les fonctionnalités à un type d'entité

1. Sélectionnez le type d'entité dans le volet de navigation à gauche de l'écran.
2. Cliquez sur le bouton Joindre une fonctionnalité (bouton du haut à droite de l'écran).
3. Dans la liste des fonctionnalités disponibles, choisissez une ou plusieurs fonctionnalités (utilisez Ctrl-Clic pour choisir plusieurs fonctionnalités) et cliquez sur OK.

Pour supprimer les fonctionnalités d'un type d'entité

1. Sélectionnez le type d'entité dans le volet de navigation à gauche de l'écran.
2. Sélectionnez une ou plusieurs fonctionnalités dans la table des fonctionnalités reliées à droite de l'écran. Utilisez Ctrl-Clic pour choisir plusieurs fonctionnalités.
3. Cliquez sur le bouton Annuler l'adjonction de la fonctionnalité (bouton du bas à droite de l'écran).

Pour dupliquer un type d'entité existant

1. Dans le volet de navigation à gauche de l'écran, faites un clic droit sur le type d'entité à dupliquer.
2. Choisissez Dupliquer le type d'entité.
3. Saisissez un nom unique pour le nouveau type d'entité et cliquez sur OK.
4. Reliez des fonctionnalités au type d'entité ou supprimez des fonctionnalités du type d'entité selon les besoins (voir instructions précédentes).

Pour renommer un type d'entité

Avertissement : si vous modifiez ou supprimez une fonctionnalité ou un élément de fonctionnalité alors que le référentiel contient déjà des données, vous devez alors purger le référentiel et recharger les données. Cela évite de conserver des incohérences dans le référentiel.

1. Dans le volet de navigation à gauche de l'écran, faites un clic droit sur le type d'entité à renommer.
2. Choisissez Renommer.
3. Saisissez le nouveau nom pour le type d'entité et cliquez sur OK.

Pour supprimer un type d'entité

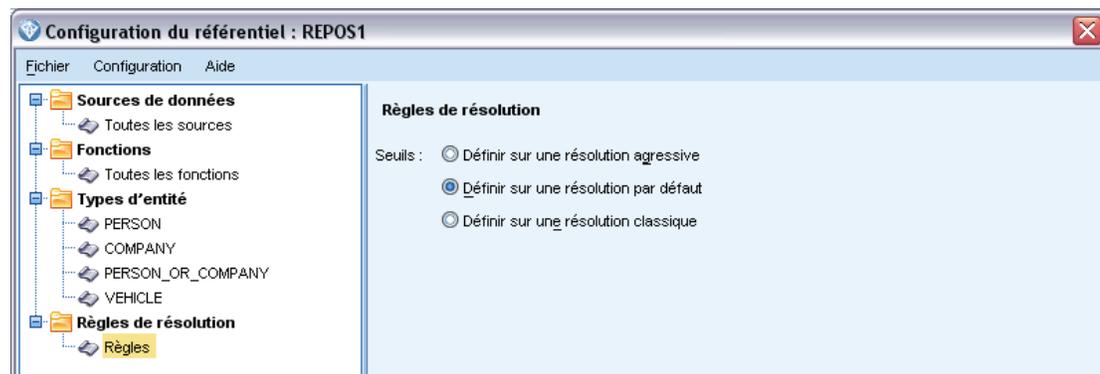
Avertissement : si vous modifiez ou supprimez une fonctionnalité ou un élément de fonctionnalité alors que le référentiel contient déjà des données, vous devez alors purger le référentiel et recharger les données. Cela évite de conserver des incohérences dans le référentiel.

1. Dans le volet de navigation à gauche de l'écran, faites un clic droit sur le type d'entité à supprimer.
2. Choisissez Supprimer.
3. Cliquez sur OK pour confirmer la suppression du type d'entité. *Avertissement* : Vous ne pouvez pas annuler la suppression d'un type d'entité.

Définition du seuil de mise en correspondance des entités

Dans la section Règles de résolution de la fenêtre Configuration du référentiel, choisissez le seuil auquel la mise en correspondance des entités aura lieu.

Figure 3-9
Définition des règles de résolution des entités



Lorsque vous créez le référentiel, la mise en correspondance est prédéfinie sur le seuil par défaut.

Choisissez Définir la résolution agressive si vous ne trouvez pas assez de correspondances dans vos enregistrements pour effectuer la résolution d'entités.

Choisissez Définir à la résolution par défaut pour revenir au seuil par défaut d'un des autres paramètres.

Choisissez Définir à la résolution classique si vous trouvez trop de correspondances.

Réutilisation d'une configuration de référentiel

Si vous avez déjà défini une configuration et que vous souhaitez l'utiliser pour un autre référentiel, vous pouvez exporter la configuration existante vers un fichier XML et importer le fichier dans l'autre référentiel (cible).

Pour réutiliser une configuration existante

1. Affichez la fenêtre Configuration de référentiel pour le référentiel dont vous souhaitez utiliser la configuration. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.
2. Dans le menu de cette fenêtre, choisissez Configuration > Exporter la configuration.
3. Dans la boîte de dialogue Enregistrer sous, choisissez le nom et l'emplacement du fichier d'exportation XML.
4. Affichez la fenêtre Configuration du référentiel du répertoire cible.

5. Dans le menu de cette fenêtre, choisissez Configuration > Importer la configuration.
6. Dans la boîte de dialogue Ouvrir, choisissez le nom et l'emplacement du fichier XML exporté précédemment et cliquez sur Ouvrir.

Enregistrement de vos modifications de configuration

Pour enregistrer les modifications de la configuration

Dans le menu de la fenêtre Configuration du référentiel, choisissez Fichier > Enregistrer.

Fermeture de la fenêtre de configuration

Pour quitter la fenêtre de configuration

Dans le menu de la fenêtre Configuration du référentiel, choisissez Fichier > Quitter.

S'il reste des modifications de la configuration non enregistrées, cliquez sur OK pour enregistrer les changements et quitter ou sur Annuler pour quitter sans effectuer d'enregistrement.

Analyse des identités résolues (nœud source Entity Analytics (EA))

Avec au moins une entrée de source de données dans le référentiel, vous pouvez utiliser le nœud source Entity Analytics (EA) pour transmettre les identités résolues à d'autres nœuds IBM® SPSS® Modeler pour une analyse ou un traitement ultérieur, tels que la création d'un rapport répertoriant les identités résolues.

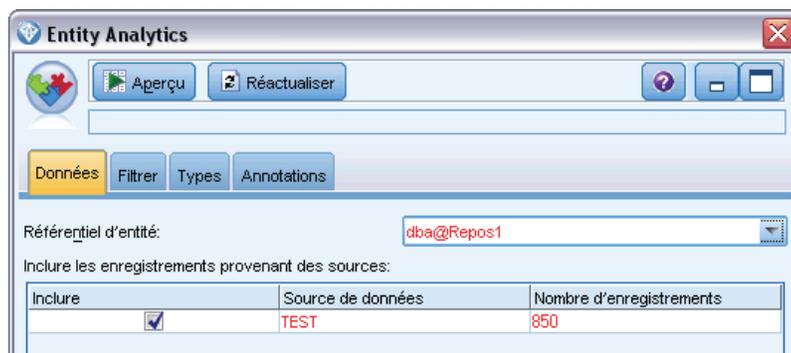
Pour analyser les identités résolues

1. Ajoutez un nœud source Entity Analytics (EA) à un flux.
2. Ouvrez un nœud source Entity Analytics (EA).
3. Dans l'onglet Données, sélectionnez le référentiel d'entité et une ou plusieurs sources de données (cliquez sur Actualiser pour mettre à jour le nombre d'enregistrements). Pour plus d'informations, reportez-vous à la section [Sélection d'une source de données](#) sur p. 32.
4. Ajoutez d'autres nœuds au flux pour effectuer le traitement désiré. Pour plus d'informations, reportez-vous à la section [Ajout de nœuds au flux](#) sur p. 34.

Sélection d'une source de données

Dans l'onglet Données, sélectionnez au moins une source de données dans le référentiel sur lequel effectuer un traitement supplémentaire. Cliquez sur Actualiser pour mettre à jour le nombre d'enregistrements des sources de données répertoriées.

Figure 3-10
Sélection d'une source de données sur le nœud Source



Référentiel d'entité. Affiche le référentiel d'entités actuel s'il existe. Pour choisir un autre référentiel s'il en existe plusieurs, sélectionnez-en un dans la liste. Pour créer un nouveau référentiel, choisissez <Parcourir...> pour afficher une boîte de dialogue à partir de laquelle vous pourrez créer le référentiel. Pour plus d'informations, reportez-vous à la section [Options du référentiel d'entité](#) sur p. 18.

Inclure les enregistrements provenant des sources Cette table répertorie les différentes sources de données qui ont des entrées dans le référentiel, ainsi que le nombre d'enregistrements dans chaque source. Sélectionnez la case Inclure pour les sources de données que vous souhaitez utiliser pour une analyse et un traitement ultérieurs.

Nouveau nom des champs de données

Vous pouvez utiliser l'onglet Filtrer pour donner un nouveau nom aux champs d'identités résolues qui sont transmis en aval pour un traitement ultérieur. Il est possible que vous désiriez renommer un champ d'identités résolues, par exemple, pour conserver la compatibilité des noms des champs lors de la fusion en aval avec un autre ensemble de données.

Les champs avec leurs noms d'origine sont les suivants.

Table 3-2
Champs d'identités résolues

Champ	Description
\$EA-ID	Identifiant d'entité
\$EA-SRC	Balise source identifiant la source de données d'où proviennent les enregistrements
\$EA-KEY	Champ désigné comme clé unique dans le fichier source de données

Remarque : bien que vous puissiez également utiliser l'onglet Filtrer pour filtrer les champs, ne l'utilisez pas ici car les champs d'identités résolues sont le strict minimum nécessaire au traitement des analyses d'entités.

Définition des informations de type pour les champs de données

Dans l'onglet Types, vous pouvez consulter ou modifier les différentes propriétés des champs d'identités résolues qui sont transmis en aval pour un traitement ultérieur.

Les propriétés que vous pouvez modifier sont les mêmes que celles de l'onglet Types d'un nœud Typier SPSS Modeler ordinaire et sont les suivantes.

Table 3-3
Propriétés de types pour les champs

Propriété	Description
Mesure	Le niveau de mesure (c'est-à-dire le type de données) utilisé pour décrire les caractéristiques des données dans le champ.
Valeurs	Propose des options de lecture des valeurs de données de l'ensemble de données.
Manquantes	Permet de spécifier le traitement des valeurs manquantes du champ.
Vérifier	Les options de validation permettant de vérifier que les valeurs des champs sont conformes aux valeurs ou intervalles spécifiés.
Rôle	Spécifie la façon dont le champ sera utilisé si les données sont transmises au nœud de modélisation ou au nugget de modèle.

Ajout de nœuds au flux

Vous pouvez ajouter différents nœuds SPSS Modeler au flux pour effectuer les opérations d'analyse ou de traitement sur la sortie du nœud source Entity Analytics (EA). Par exemple, vous pouvez ajouter un ou plusieurs des nœuds suivants.

- Agrégez le nœud pour résumer la sortie qui peut être de taille importante
- Sélectionnez le nœud pour sélectionner un sous-ensemble de la sortie
- Nœud Table pour visualiser la sortie du nœud source Entity Analytics (EA)
- Nœud Rapport pour imprimer la sortie d'un rapport
- Nœud Export SPSS Modeler pour exporter la sortie dans un format différent, tel qu'une feuille de calcul ou une base de données

Pour plus d'informations, consultez les sections sur les nœuds Opérations d'enregistrement, Sortie et Export dans le *Guide des nœuds Source, Processus et Sortie de IBM SPSS Modeler*.

Comparaison de nouvelles observations au contenu du référentiel (nœud Mise en flux EA)

Lorsque vous avez déjà exécuté des résolutions d'identités dans le référentiel, vous pouvez utiliser le nœud Mise en flux EA pour comparer de nouvelles observations rencontrées ultérieurement au contenu du référentiel. Ce nœud traite les enregistrements provenant d'une nouvelle source de données, les compare aux entités déjà résolues dans le référentiel et transmet ces données au flux pour un traitement supplémentaire. Il est possible de définir les correspondances afin qu'elles soient exactes ou associées d'une manière ou d'une autre aux entités existantes.

Tout comme le nœud Export EA, le nœud Mise en flux EA utilise un nœud source SPSS Modeler unique comme entrée. Cependant, le nœud Mise en flux EA diffère sur les points suivants. Alors que le nœud Export produit des données pour toutes les entités associées à ses enregistrements d'entrée, le nœud Mise en flux EA produit des données uniquement pour les entités associées aux entités déjà résolues dans le référentiel. Pour plus d'informations, reportez-vous à la section [Sortie du nœud Mise en flux EA](#) sur p. 40.

Pour comparer de nouvelles observations au contenu du référentiel

1. Connectez-vous à la source de données contenant les nouveaux enregistrements que vous souhaitez comparer aux entités existantes. Pour plus d'informations, reportez-vous à la section [Connexion à une source de données](#) sur p. 15.
2. Dans l'onglet Options d'enregistrement, reliez un nœud Mise en flux EA au nœud source de données.
3. Faites un double clic sur le nœud Export Entity Analytics pour ouvrir sa boîte de dialogue.
4. Cliquez sur la liste Référentiel d'entité.
5. Cliquez sur <Parcourir...> pour afficher la boîte de dialogue Référentiels d'entités.
6. Dans la boîte de dialogue Référentiels d'entités, cliquez sur le champ Nom du référentiel.
7. Cliquez sur le nom du référentiel à utiliser.
8. Entrez le nom d'utilisateur et le mot de passe du référentiel et cliquez sur Connecter. Cliquez sur OK lorsque le référentiel est connecté.
9. Dans la boîte de dialogue Mise en flux EA, sélectionnez le type d'entité à utiliser. Pour plus d'informations, reportez-vous à la section [Gestion des types d'entités](#) sur p. 28.
10. Mapper les champs d'entrées dans la source de données sur les fonctionnalités du référentiel. Pour plus d'informations, reportez-vous à la section [Mappage de champs d'entrées sur les fonctionnalités \(nœud Mise en flux EA\)](#) sur p. 36.
11. Vous pouvez cliquer sur l'onglet Référentiel pour afficher une table contenant le mappage de cette source de données (ou toutes les sources de données ayant été mappées). Pour plus d'informations, reportez-vous à la section [Affichage des mappages de champs \(nœud Mise en flux EA\)](#) sur p. 37.
12. Cliquez sur l'onglet Afficher pour consulter les détails des différentes sources de données qui ont été ajoutées au référentiel et définissez les critères de sélection pour récupérer les entités existantes. Pour plus d'informations, reportez-vous à la section [Affichage des sources de données \(nœud Mise en flux EA\)](#) sur p. 38.
13. Cliquez sur OK lorsque le nœud est correctement configuré.
14. Reliez un nœud Table au nœud Mise en flux EA et exécutez le flux.

La fenêtre des résultats du nœud Table répertorie toutes les entités récupérées correspondant aux nouveaux enregistrements de la source de données. Le préfixe \$EA- est ajouté aux champs de sortie. Pour plus d'informations, reportez-vous à la section [Sortie du nœud Mise en flux EA](#) sur p. 40.

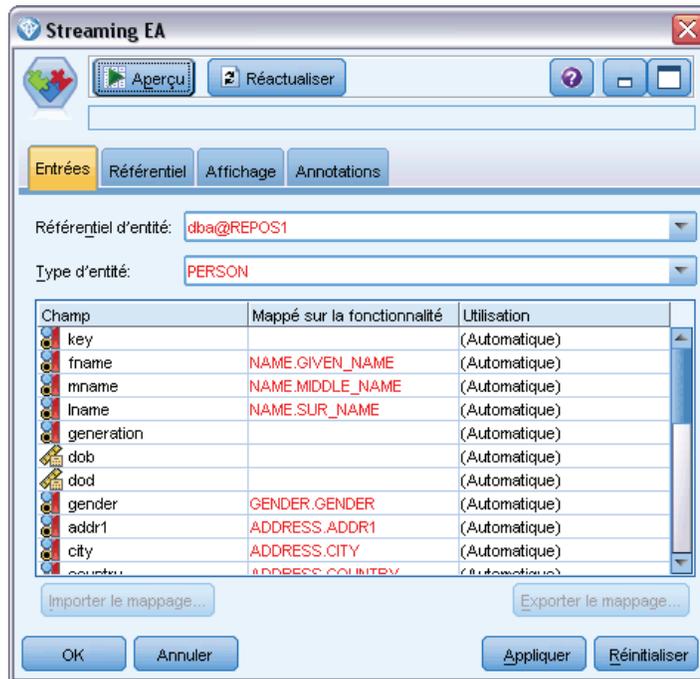
Mappage de champs d'entrées sur les fonctionnalités (nœud Mise en flux EA)

L'onglet Entrées contient les options de mappage des champs dans l'entrée de ce nœud sur les fonctionnalités du référentiel. Configurez les tâches de mappage dans cet onglet, ou l'onglet Afficher pour afficher des détails sur toutes les sources de données dans le référentiel, puis cliquez sur OK.

Si vous avez déjà stocké un ensemble de mappages dans un fichier XML, vous pouvez les utiliser en cliquant sur Importer le mappage.

Figure 3-11

Mappage des champs des nouveaux enregistrements sur les fonctionnalités dans le référentiel



Référentiel d'entité. Affiche le référentiel d'entités actuel s'il existe. Pour choisir un autre référentiel s'il en existe plusieurs, sélectionnez-en un dans la liste. Pour créer un nouveau référentiel, choisissez <Parcourir...> pour afficher une boîte de dialogue à partir de laquelle vous pourrez créer le référentiel. Pour plus d'informations, reportez-vous à la section [Options du référentiel d'entité](#) sur p. 18.

Type d'entité. Une liste de types d'entité (c'est-à-dire des ensembles de fonctionnalités) définis dans le référentiel. En choisir un dans la liste ou choisir <Ajouter un nouveau type d'entité...> pour afficher la fenêtre de configuration du référentiel, où vous pouvez définir un nouveau type d'entité. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.

Table de mappage. Dans cette table, vous pouvez mapper chaque champ d'entrée sur une fonctionnalité correspondante dans le référentiel. Si aucune fonctionnalité adaptée n'existe dans le type d'entité sélectionné, vous pouvez créer une nouvelle fonctionnalité ici.

- **Champ.** L'ensemble des champs d'entrée dans la source de données sélectionnée. Chaque champ contient une icône indiquant le niveau de mesure (c'est-à-dire le type de données) du champ.
- **Mappé sur fonctionnalité.** Pour mapper un champ sur une fonctionnalité, faites un double clic sur cette colonne (ou appuyez sur la barre d'espace) sur la ligne du champ et choisissez une fonctionnalité dans la liste. Si aucune fonctionnalité appropriée n'est disponible, choisissez <Ajouter une nouvelle fonctionnalité...> pour afficher la fenêtre de configuration du référentiel où vous pouvez définir une nouvelle fonctionnalité pour ce type d'entité. Pour plus d'informations, reportez-vous à la section [Configuration d'un référentiel d'entité](#) sur p. 22.
- **Utilisation.** Indique le contexte d'un champ spécifique où plusieurs contextes sont possibles, par exemple, des numéros de téléphone du bureau ou du domicile. Pour plus d'informations, reportez-vous à la section [Gestion des types d'entités](#) sur p. 28.

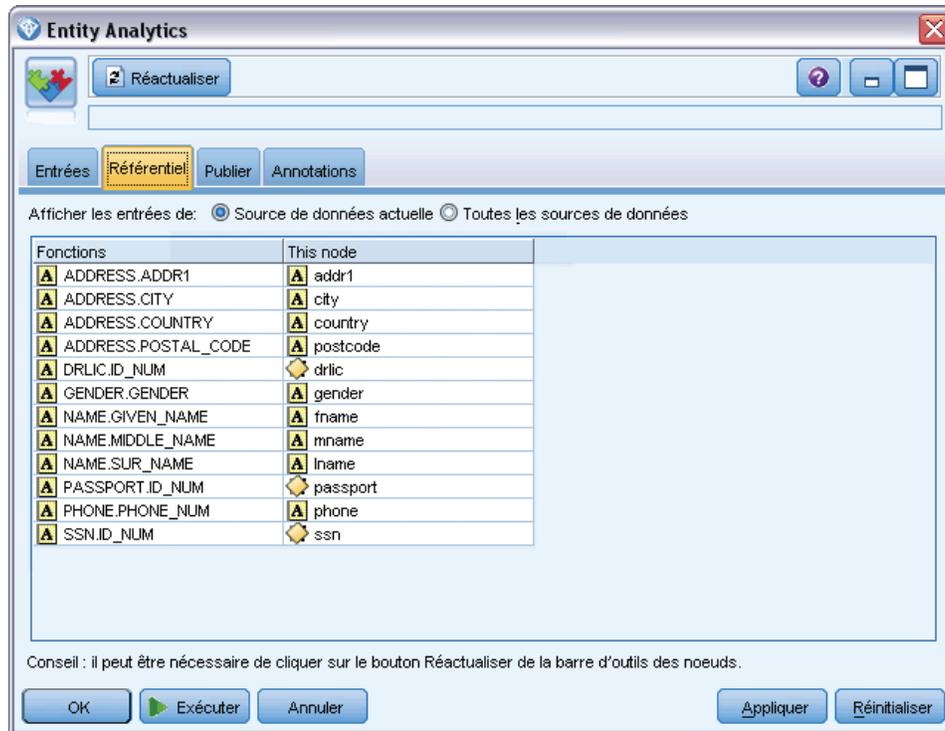
Importer le mappage. Importe un ensemble précédemment exporté de mappages de champs sur fonctionnalités depuis un fichier XML externe. Cette fonctionnalité peut être utile si vous disposez de différentes sources de données avec les mêmes paramètres de mappage car elle permet de ne pas avoir à redéfinir les mêmes mappages pour différentes sources.

Exporter le mappage. Exporte vers un fichier XML externe l'ensemble des mappages champs sur fonctionnalités affiché dans la table de mappage.

Affichage des mappages de champs (nœud Mise en flux EA)

Sur l'onglet Référentiel, cliquez sur le bouton Actualiser pour voir les fonctionnalités du référentiel sur lesquelles des champs d'entrée sont mappés. Vous pouvez faire cela pour la source de données actuelle (celle contrôlée par le nœud source relié à ce nœud Export) ou pour toutes les sources de données.

Figure 3-12
Affichage des mappages de champs



Afficher les entrées pour. Choisissez une option pour afficher les mappages de la source de données actuelle ou de toutes les sources de données connues du référentiel.

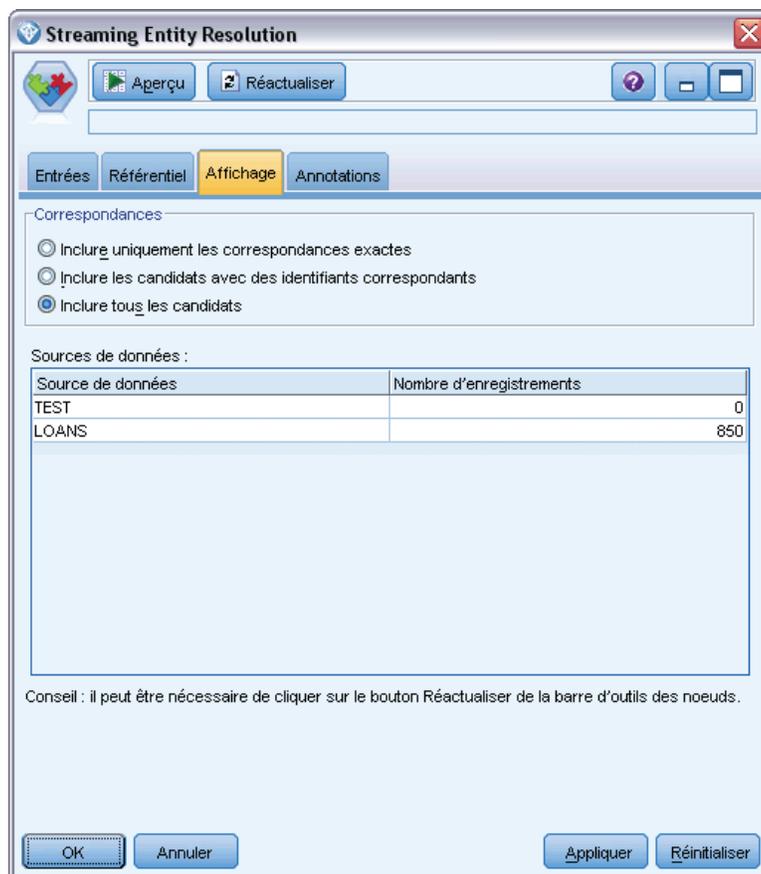
Caractéristiques. Une liste de toutes les fonctionnalités ayant des mappages dans les sources de données affichées. Les fonctionnalités non mappées ne sont pas affichées.

Ce nœud. Chaque colonne répertorie les champs mappés dans une source de données spécifique pour chaque fonctionnalité pour laquelle un mappage a été défini.

Affichage des sources de données (nœud Mise en flux EA)

Dans l'onglet Afficher, vous pouvez voir les détails des différentes sources de données ayant été ajoutées au référentiel. Il s'agit de sources de données par rapport auxquelles l'entrée de ce nœud est traitée, afin de rechercher et de récupérer des entités correspondantes. Cliquez sur Actualiser pour mettre à jour le nombre d'enregistrements.

Figure 3-13
Affichage des détails de sources de données



Correspondances. Ces options spécifient à quel niveau les informations de mappage de champs sur fonctionnalités spécifiées dans l'onglet Entrées doivent être mises en correspondance avec les enregistrements candidats (c'est-à-dire tout le contenu du référentiel). Plus les critères ont un niveau élevé de correspondance, moins il y aura d'entités récupérées.

- **Inclure uniquement les correspondances exactes.** Il s'agit de l'option de correspondance la plus élevée qui génère le plus petit nombre d'enregistrements. Utilisez cette option lorsque vous souhaitez renvoyer uniquement les entités considérées comme des correspondances exactes.
- **Inclure les candidats avec des identifiants correspondant.** Utilisez ce paramètre lorsque vous souhaitez renvoyer à la fois des entités correspondantes et des entités qui partagent les mêmes identifiants (ceux ayant des fonctionnalités configurées avec une valeur de fréquence de Un, par exemple ceux qui correspondent aux numéros de carte bancaire, aux numéros de TVA, etc.).
- **Inclure tous les candidats.** Utilisez cette option lorsque vous souhaitez voir le plus grand nombre possible d'entités partageant des fonctionnalités dans le référentiel. Il s'agit du critère de correspondance le plus flou qui renvoie le plus grand nombre d'enregistrements sélectionnés. Cette option renvoie des correspondances exactes et des entités partageant quasiment toutes les fonctionnalités (généralement celles avec une valeur de fréquence de

Un ou de Quelques). Par exemple, deux entités avec le même numéro de TVA et des entités avec des adresses similaires seraient incluses.

Sources de données. Cette table répertorie les balises sources de données qui ont été chargées dans le référentiel, ainsi que le nombre d'enregistrements dans chaque source.

Sortie du nœud Mise en flux EA

La sortie du nœud Mise en flux EA est composée des champs suivants pour chaque enregistrement récupéré.

Champ	Description
<i>Champ1</i> [, <i>Champ2</i> [, ... <i>ChampN</i>]]	Champs provenant de la source de données qui contient les nouveaux enregistrements.
\$EA-ID	Identifiant d'entité de l'enregistrement dans le référentiel.
\$EA-SC	Champ indiquant la proximité de la correspondance entre l'enregistrement et l'entité observée dans le référentiel, valeur comprise entre 1,0 (correspondance faible) et 10,0 (bonne correspondance).
\$EA-SRC	Balise source identifiant la source de données d'où proviennent les enregistrements
\$EA-KEY	Valeur de la clé unique de l'enregistrement dans le fichier de source de données.
<i>\$EA-fonctionnalité1</i> [, <i>\$EA-fonctionnalité2</i> [, ... <i>\$EA-fonctionnalitéN</i>]]	Valeurs des fonctionnalités mappées de l'enregistrement dans le référentiel.

Utiliser IBM SPSS Modeler Entity Analytics avec les autres produits IBM SPSS

Des programmes d'installation sont disponibles afin de vous permettre d'utiliser IBM SPSS Modeler Entity Analytics avec les produits suivants :

- IBM SPSS Collaboration and Deployment Services
- IBM SPSS Modeler Batch pour Windows
- IBM SPSS Modeler Solution Publisher

Vous devrez exécuter ces programmes d'installation avant de pouvoir utiliser les fonctionnalités de IBM SPSS Modeler Entity Analytics avec ces produits. Pour plus d'informations, reportez-vous au guide d'*Installation d'IBM SPSS Modeler Premium*.

Tâches administratives

Pour les référentiels qui sont créés dans Entity Analytics, un nouveau service de base de données est créé avec le produit IBM solidDB. Il existe peu de tâches administratives associées à solidDB. Ces tâches généralement exécutées par l'administrateur de la base de données ou l'administrateur système sont les suivantes :

- Configuration des tâches de port
- Gestion des identifiants de l'administrateur pour la base de données du référentiel

D'autres tâches administratives pouvant s'avérer nécessaires et s'appliquant à tous les référentiels, sont les suivantes :

- Déplacement du répertoire de stockage du référentiel
- Définition des options flux pour les champs date/heure et horodatage
- Exécution de IBM SPSS Modeler Entity Analytics avec le client SPSS Modeler et SPSS Modeler Server installés sur le même système Windows
- Purge d'un référentiel d'entité
- Suppression d'un référentiel d'entité
- Suppression d'un référentiel quand aucune connexion n'est possible

Configuration des tâches de port

Chaque service de base de données solidDB doit avoir un port qui lui est alloué et ce port ne peut être alloué à d'autres services s'exécutant sur l'ordinateur. Les services de la base de données se trouvent sur le même ordinateur que celui qui exécute IBM® SPSS® Modeler Server (ou, quand IBM® SPSS® Modeler est utilisé sans connexion à SPSS Modeler Server, l'ordinateur qui exécute IBM® SPSS® Modeler).

Par défaut, Entity Analytics attribue les ports allant de 1320 à 1520, en commençant au port 1320 pour le premier référentiel créé. En cas de conflit, vous pouvez configurer les tâches des ports en modifiant le fichier : `<modeler server installation path>/ext/bin/pasw.entityanalytics/ea.cfg` et en définissant les valeurs appropriées des paramètres `min_port` et `max_port`. Le contenu par défaut de ce fichier apparaît ci-dessous :

```
# port range configuration for entity analytics
#
# this port range controls which ports SolidDB databases
# (created to store Entity Analytics Repositories in)
# may use. Configure this if the default port range will
# introduce a conflict on your system.
#
```

```
# default min_port = 1320

# default max_port = 1520

min_port, 1320

max_port, 1520
```

Gestion des identifiants de l'administrateur pour la base de données du référentiel

Le nom d'utilisateur et le mot de passe de l'administrateur de la base de données solidDV qui héberge un référentiel d'entités sont définis lorsque le référentiel est créé. Si vous connaissez les identifiants actuels, vous pouvez modifier ces informations à l'aide de l'éditeur SQL solidDB.

Pour démarrer l'éditeur SQL solidDB

1. Sur un ordinateur client, ouvrez une fenêtre d'invite de commande.
2. Entrez :

```
cd modeler_install_dir\ext\bin\pasw.entityanalytics\solidDB\bin
```

où *modeler_install_dir* est le répertoire d'installation de SPSS Modeler.

3. Entrez :

```
solsql -c "C:\Documents and Settings\All Users\Application  
Data\IBM\SPSS\Modeler\version\EA\repositories\repos_name
```

où *version* est le numéro de version de l'installation SPSS Modeler et *repos_name* est le nom du référentiel.

4. A l'invite, entrez le nom d'utilisateur et le mot de passe actuels de l'administrateur de la base de données pour afficher l'invite `solsql>`.

Pour modifier le mot de passe de l'administrateur de la base de données

1. À l'invite `solsql>`, entrez :

```
alter user username identifié par password;  
commit work;
```

où *username* est le nom d'utilisateur actuel de l'administrateur de la base de données et *password* est le nouveau mot de passe.

2. Entrez `exit` (quitter) pour fermer l'éditeur.
3. Redémarrez le client SPSS Modeler.

Pour exécuter d'autres tâches administratives

Pour des informations sur d'autres tâches administratives à exécuter avec la base de données solidDB, consultez la documentation pour la version appropriée de IBM solidDB sur <http://publib.boulder.ibm.com/>.

Déplacement du répertoire de stockage du référentiel

Par défaut, les fichiers du référentiel sont stockés dans un répertoire nommé *EA* à l'emplacement suivant :

- C:\Documents and Settings\All Users\ApplicationData\IBM\SPSS\Modeler\version\EA (système Windows)
- *modeler_install_directory*/ext/bin/pasw.entityanalytics/EA (systèmes UNIX)

Les fichiers de stockage du référentiel pouvant devenir très volumineux, il se peut que vous deviez les déplacer sur un disque différent ou sur une partition différente afin de libérer de l'espace. Pour ce faire, suivez la procédure suivante.

1. Quittez SPSS Modeler.
2. Déplacez le répertoire *EA* de son emplacement d'origine (indiqué précédemment) vers son nouvel emplacement. Par exemple, sous Windows vous pouvez souhaiter le déplacer vers un nouvel emplacement, tel que *F:\data\EA*.
3. Modifiez le fichier *<modeler_server_installation_path>/ext/bin/pasw.entityanalytics/ea.cfg* en ajoutant l'option suivante :

repository_data_directory, new_location

où *new_location* est le répertoire vers lequel vous déplacez le répertoire *EA*, par exemple, *F:\data\EA*.

Définition des options flux pour les champs date/heure et horodatage

Si votre source de données inclut des champs heure/date ou d'horodatage, vérifiez que les options de flux correspondantes sont définies dans un format reconnu par IBM SPSS Modeler Entity Analytics.

Pour définir le format des options de flux

1. Dans le menu SPSS Modeler principal, sélectionnez :
Outils > Propriétés du flux > Options.
2. Sélectionnez Date/heure.
3. Définissez le format de date sur AAA-MM-JJ.
4. Définissez le format de l'heure sur HH:MM:SS.
5. Cliquez sur OK.

Exécution de IBM SPSS Modeler Entity Analytics avec le client SPSS Modeler et SPSS Modeler Server installés sur le même système Windows

Si vous avez installé IBM SPSS Modeler Entity Analytics à la fois dans le client SPSS Modeler dans SPSS Modeler Server sur le même système Windows, par défaut le client et le serveur partageront le même référentiel. Si vous souhaitez qu'ils utilisent des référentiels distincts, vous devez modifier le fichier de configuration *ea.cfg* sur l'un des systèmes, de sorte à le configurer pour qu'il utilise une plage de port et un dossier de référentiel différents.

Remarque : notamment si vous utilisez un client SPSS Modeler de 32 bits et un SPSS Modeler Server de 64 bits (ou vice versa), vous devrez suivre cette procédure.

1. Ouvrez le fichier `<modeler [server] installation path>/ext/bin/pasw.entityanalytics/ea.cfg` pour l'édition.
2. Modifiez les paramètres `min_port` et `max_port` afin que les systèmes utilisent des ports différents. Pour plus d'informations, reportez-vous à la section [Configuration des tâches de port](#) sur p. 41.
3. Modifiez le paramètre `repository_data_directory` afin que les systèmes utilisent des répertoires différents.
4. Enregistrez le fichier *ea.cfg*, puis fermez-le.

Purge d'un référentiel d'entité

Si vous souhaitez nettoyer les enregistrements de données d'un référentiel d'entité, tout en conservant les informations de configuration, utilisez l'option de purge du référentiel.

Pour purger un référentiel

1. Ouvrez un nœud Entity Analytics.
2. Cliquez sur la liste Référentiel d'entité.
3. Cliquez sur `<Parcourir...>` pour afficher la boîte de dialogue Instances de résolution d'entités.
4. Dans la boîte de dialogue Instances de résolution d'entités, cliquez sur la liste Nom du référentiel.
5. Sélectionnez le référentiel à purger.
6. Si vous n'êtes pas encore connecté, saisissez le nom et le mot de passe administrateur et cliquez sur `Se connecter`.
7. Lorsque le bouton Purger le référentiel est activé, cliquez dessus.
8. Cliquez sur Purger le référentiel dans la boîte de dialogue Purger les sources de données pour confirmer la purge du référentiel.

Suppression d'un référentiel d'entité

Lorsque vous n'avez plus besoin d'un référentiel, vous pouvez le supprimer entièrement.

Avertissement : supprimer signifie supprimer. **Vous ne pouvez pas annuler cette opération.** Si vous n'êtes pas sûr, utilisez le bouton Purger pour supprimer toutes les sources de données. Cela ne supprime pas la configuration du référentiel. Pour plus d'informations, reportez-vous à la section [Purge d'un référentiel d'entité](#) sur p. 44.

Remarque : La procédure suivante suppose que vous connectiez le référentiel de SPSS Modeler et que vous connaissiez le nom d'utilisateur et le mot de passe de la base de données qui héberge le référentiel. Si ce n'est pas le cas, suivez la procédure pour supprimer un référentiel lorsque vous ne parvenez pas à vous y connecter. Pour plus d'informations, reportez-vous à la section [Suppression d'un référentiel quand aucune connexion n'est possible](#) sur p. 45.

Pour supprimer un référentiel

1. Ouvrez un nœud Entity Analytics.
2. Cliquez sur la liste Référentiel d'entité.
3. Cliquez sur <Parcourir...> pour afficher la boîte de dialogue Instances de résolution d'entités.
4. Dans la boîte de dialogue Instances de résolution d'entités, cliquez sur la liste Nom du référentiel.
5. Sélectionnez le référentiel à supprimer.
6. Si vous n'êtes pas encore connecté, saisissez le nom et le mot de passe administrateur et cliquez sur Se connecter.
7. Lorsque le bouton Supprimer le référentiel en entier est activé, cliquez dessus.
8. Cliquez sur Supprimer pour confirmer la suppression du référentiel.
9. Cliquez sur OK pour confirmer que votre suppression a réussi.

Suppression d'un référentiel quand aucune connexion n'est possible

Utilisez la procédure suivante si vous souhaitez effacer un référentiel entier mais que vous ne parvenez pas à vous y connecter soit en raison de problèmes de connectivité de SPSS Modeler soit parce que vous avez oublié le nom d'utilisateur et le mot de passe.

Exécutez cette procédure sur l'ordinateur qui héberge la base de données du référentiel.

Systèmes Windows

1. Ouvrez une fenêtre d'invite de commande.
2. Entrez :

```
cd modeler_install_dir  
cd ext\bin\pasw.entityanalytics\tools  
delete_repository.bat repos_name
```

où *modeler_install_dir* est le répertoire d'installation de SPSS Modeler et *repos_name* est le nom du référentiel.

3. Continuez à partir de la rubrique « Terminer la procédure » plus tard dans cette section.

Systèmes UNIX

1. Ouvrez un fichier shell.
2. Entrez :

```
cd modeler_server_install_dir  
cd ext/bin/pasw.entityanalytics/tools  
./delete_repository.sh repos_name
```

où *modeler__server_install_dir* est le répertoire d'installation de SPSS Modeler Server et *repos_name* est le nom du référentiel.

Terminer la procédure (tous les systèmes)

1. A l'invite, confirmez la suppression du référentiel en entrant Y.
2. Lorsque le référentiel a été effacé, le message suivant apparaît :

Informations - supprimez les fichiers du référentiel dans le répertoire suivant :
directory_path
(Remarque ; il peut être nécessaire d'effectuer un redémarrage avant de pouvoir supprimer les fichiers du référentiel).
3. Supprimez le répertoire qui porte le même nom que le référentiel supprimé. Si vous ne parvenez pas à supprimer le répertoire, redémarrez l'ordinateur et réessayez.

Fonctionnement des analyses d'entités

A propos de cet exemple

Dans cet exemple, nous verrons en quoi ajouter les analyses d'entités permet d'améliorer les excellents résultats obtenus en utilisant IBM® SPSS® Modeler.

Cet exemple utilise le flux *loan_entity_analytics.str*, qui fait référence au fichier de données *loan_applications.csv*. Ces fichiers sont disponibles dans le répertoire *Demos* de toute installation SPSS Modeler qui contient également IBM® SPSS® Modeler Entity Analytics. Ce répertoire *Demos* est accessible à partir du groupe de programmes SPSS Modeler dans le menu Démarrer de Windows. Le fichier *loan_entity_analytics.str* se trouve dans le répertoire *Entity_Analytics*.

Remarque : Avant de pouvoir exécuter cet exemple de flux, vous devez créer un référentiel sur votre système. Pour ce faire, poursuivez avec cet exemple. Pour plus d'informations, reportez-vous à la section [Création d'un référentiel](#) dans le chapitre 3 sur p. 16.

Commençons par une situation familière : les dirigeants d'une banque s'inquiètent de savoir si oui ou non ses clients vont être dans l'incapacité de rembourser leurs prêts qui sont en cours d'acceptation. Le service informatique de la banque utilise SPSS Modeler depuis un moment et a déjà créé un flux et construit un modèle prédictif à partir des données existantes concernant environ 700 prêts accordés dans le passé. Ces prêts ont soit été remboursés soit les clients n'ont pas effectué leurs remboursements.

Le modèle d'origine

Voici comment le service informatique de la banque a créé son modèle et ce qu'il en a appris.

Figure 4-1
Flux d'origine avec noeud de modélisation



En plus des informations sur les prêts passés, l'ensemble de données *loan_applications.csv* contient des informations sur 150 clients dont les demandes de prêts sont encore en attente, soit un total de 850 enregistrements.

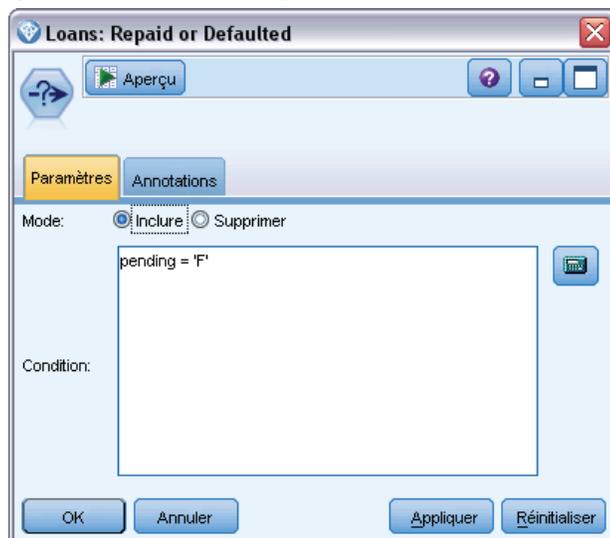
Tous les champs de l'ensemble de données ne sont pas utiles pour les prévisions, par exemple, les champs de nom peuvent être ignorés. Le noeud Typer filtre les champs à ignorer en définissant leur rôle sur Aucun. Les champs à utiliser pour effectuer des prévisions ont leur rôle défini sur Entrée et le champ dont le modèle essaie de prévoir la valeur a son rôle défini sur Cible.

Figure 4-2
Rôles des champs définis dans le noeud Typer



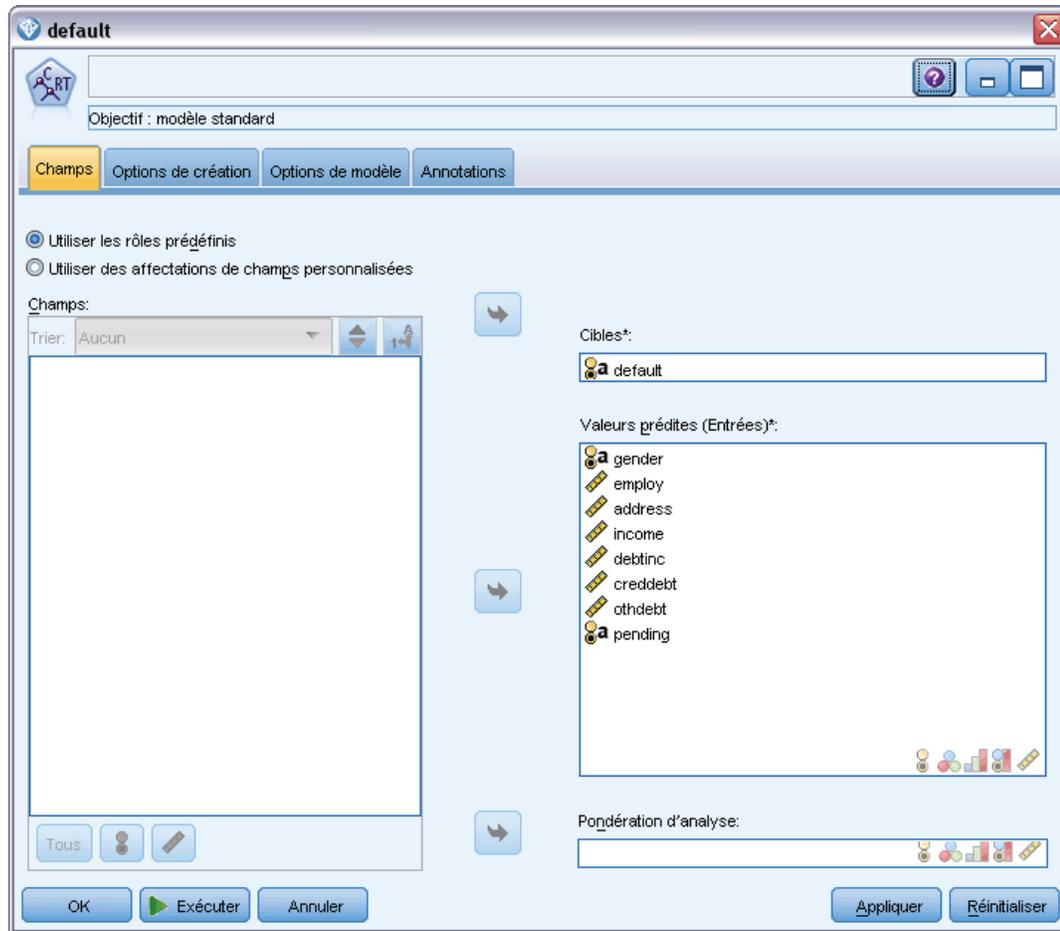
Lorsque le modèle doit effectuer ses prévisions uniquement en fonction des données passées, le flux contient un noeud Sélectionner qui inclut uniquement les prêts qui ne sont *pas* signalés comme En attente et ignore les 150 prêts en attente.

Figure 4-3
Ignorer les demandes de prêt en attente



Les prêts en attente ayant été ignorés, seules les informations sur les 700 prêts restants (ayant été remboursés ou non) sont transmises au noeud de modélisation. La banque aurait pu utiliser un des nombreux algorithmes SPSS Modeler pour produire un modèle efficace. Dans ce cas, elle a utilisé un noeud Arbre CRT qui sera utilisé pour créer un modèle permettant de prévoir les mauvais payeurs potentiels en fonction des performances passées de ses clients.

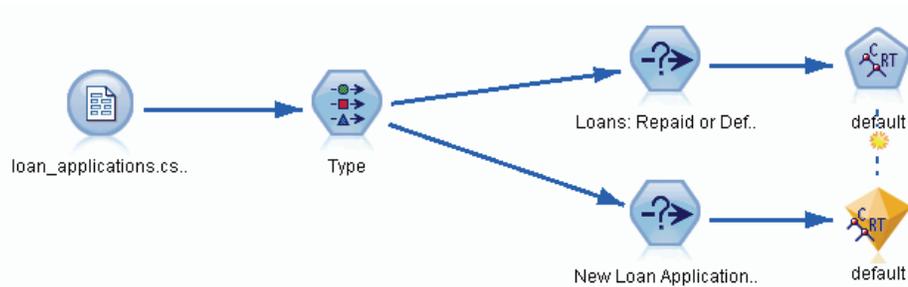
Figure 4-4
Attribution des champs de prédicteur et champs cible



Les champs utilisés pour effectuer la prévision sont créés en tant que champs de prédicteur et le champ dont le modèle essaie de prédire la valeur—par défaut dans ce cas—est défini en tant que champ cible comme le noeud Typer l’a défini auparavant.

Exécuter ce flux produit un nugget de modèle contenant le modèle qui a été créé à partir des champs de prédicteur.

Figure 4-5
Flux avec nugget de modèle ajouté



Désormais, l'analyste de la banque peut utiliser le modèle pour commencer à prévoir si les clients avec des remboursements en attente sont susceptibles de ne pas les rembourser. A l'aide de l'ensemble de données d'origine, l'analyste insère un noeud Sélectionner qui, cette fois, contient seulement les 150 enregistrements de prêt En attente, au lieu de les ignorer. L'analyste transmet ces enregistrements directement au modèle et ajoute un noeud Proportion pour une représentation visuelle des prévisions du modèle.

Figure 4-6
Flux sélectionnant les nouvelles demandes de prêts avec noeud Proportion ajouté



Le noeud Proportion affiche la proportion des valeurs du champ \$R-default dans le modèle. Ce champ est ajouté au modèle de données par le noeud Arbre CRT lorsqu'il est exécuté. Le champ contient la prévision qui établit si chaque nouveau demandeur remboursera ou non son prêt et nous utiliserons ce champ ultérieurement pour comparer ce que l'ajout d'analyses d'entité a comme effet.

En exécutant cette partie du flux, l'analyste apprend grâce aux résultats du noeud Proportion qu'il est prévu que 137 des 150 nouveaux demandeurs remboursent leurs prêts. Il est prévu que les 13 demandeurs restants n'effectuent pas leur remboursement. Par conséquent, l'analyste recommandera probablement à la banque de refuser leur demande.

Figure 4-7
Sortie du noeud Proportion sans analyses d'entités

Le screenshot montre une fenêtre intitulée 'Proportion de \$R-default'. Elle contient un menu (Fichier, Edition, Générer, Affichage) et des onglets (Table, Graphiques, Annotations). Une table de données est visible :

Valeur	Proportion	%	Comptage
Default		8.67	13
Repaid		91.33	137

Ajout d'analyses d'entités

Voyons à présent si la situation peut être améliorée en ajoutant une analyse d'entités à l'équation. Imaginons que vous êtes un spécialiste des analyses d'entités, appelé par la banque pour effectuer des recherches sur des entrées frauduleuses potentielles dans les enregistrements des clients au sein des données source. Il est possible qu'il existe des enregistrements en double générés par des erreurs de saisie de données, mais il est également possible qu'un demandeur de prêt essaie de masquer son identité. Dans les deux cas, la banque doit savoir ce qu'il se passe.

Dans cet exemple, nous supposons qu'un référentiel d'entité a déjà été créé. Pour plus d'informations, reportez-vous à la section [Création d'un référentiel](#) dans le chapitre 3 sur p. 16.

Ajout des données source au référentiel

Vous devez d'abord ajouter un nœud Export EA au nœud source de données pour pouvoir exporter les données sources dans le référentiel d'entités.

Figure 4-8

Liaison d'un nœud Export EA au nœud source de données



Avant de pouvoir exporter les données, vous devez mapper les champs de la source de données sur les fonctions dans le référentiel d'entité. Cette opération est nécessaire parce que différentes sources de données peuvent utiliser différents noms de champ pour le même type d'informations. Le référentiel d'entité propose un ensemble de types d'informations standard (appelé "fonctions") pour éviter la duplication.

Figure 4-9

Mappages des champs sur les fonctions

The screenshot shows the 'Entity Analytics' application window. The 'Entrées' tab is active, and the 'Référéntiel' sub-tab is selected. The configuration is as follows:

- Mode:** Ajouter au référentiel Purger le référentiel avant l'exportation
- Référentiel d'entité:** Repos@REPOS1
- Etiquette source:** TEST
- Type d'entité:** PERSON
- Clé unique:** key

Champ	Mappé sur la fonctionnalité	Utilisation
key		(Automatique)
fname	NAME.GIVEN_NAME	(Automatique)
mname	NAME.MIDDLE_NAME	(Automatique)
lname	NAME.SUR_NAME	(Automatique)
generation	NAME.NAME_GEN	(Automatique)
dob	DOB.DOB	(Automatique)
dod		(Automatique)
gender	GENDER.GENDER	(Automatique)
addr1	ADDRESS.ADDR1	(Automatique)
city	ADDRESS.CITY	(Automatique)
country	ADDRESS.COUNTRY	(Automatique)
postcode	ADDRESS.POSTAL_CODE	(Automatique)
phone	PHONE.PHONE_NUM	(Automatique)

Buttons at the bottom: Importer le mappage..., Exporter le mappage..., OK, Exécuter, Annuler, Appliquer, Réinitialiser.

Dans le nœud export EA, renseignez les informations concernant le référentiel : les informations de connexion, la balise source (pour identifier la source de données – TEST dans ce cas), le type d'entité (l'ensemble de fonctions que nous utilisons, celui appelé PERSON), et le champ-clé unique (pour identifier chaque enregistrement de manière unique). Dans ce cas, utilisez le champ clé comme clé unique.

Vous pouvez maintenant définir les mappages. Dans l'ensemble de fonctions que vous utilisez, il existe des fonctions qui correspondent aux champs *fname*, *mname*, *lname*, *generation*, *dob*, *gender*, *addr1*, *city*, *country*, *postcode*, *phone*, *email*, *ssn*, *drlic*, et *passport*.

Commencez par définir le mappage pour *fname*. Faites un double clic sur la colonne Mappé à la fonction dans le tableau à la ligne *fname*, recherchez l'entrée NAME.GIVEN_NAME et cliquez dessus pour créer le mappage.

Maintenant, mappez les champs restants qui ont des fonctions correspondantes afin que l'ensemble complet des mappages ressemble à ceci.

Table 4-1
Champs mappés sur les fonctions du référentiel

Champ	Mappé sur fonction
<i>fname</i>	NAME.GIVEN_NAME
<i>mname</i>	NAME.MIDDLE_NAME
<i>lname</i>	NAME.SUR_NAME
<i>génération</i>	NAME.NAME_GEN
<i>dob</i>	DOB.DOB
<i>gender</i>	GENDER.GENDER
<i>addr1</i>	ADDRESS.ADDR1
<i>city</i>	ADDRESS.CITY
<i>country</i>	ADDRESS.COUNTRY
<i>postcode</i>	ADDRESS.POSTAL_CODE
<i>phone</i>	PHONE.PHONE_NUM
<i>adresse électronique</i>	EMAIL_ADDR.ADDR
<i>ssn</i>	SSN.ID_NUM
<i>drlic</i>	DRLIC.ID_NUM
<i>passport</i>	PASSPORT.ID_NUM

Cliquez sur Exécuter pour exporter les données dans le référentiel. Cela prend un moment et quand la boîte de dialogue Commentaires sur l'exécution se ferme, l'exportation est terminée.

Lecture des identités résolues

Quand vous exportez des données dans le référentiel, les analyses d'entités commencent à résoudre les conflits d'identité possibles en attribuant un identifiant d'entité unique que vous verrez plus tard en tant que champ *\$EA-ID*. (*Remarque* : il ne s'agit pas du même champ que le champ Clé unique dans le nœud Export EA —ce champ permet d'identifier les enregistrements de source de données de manière unique.)

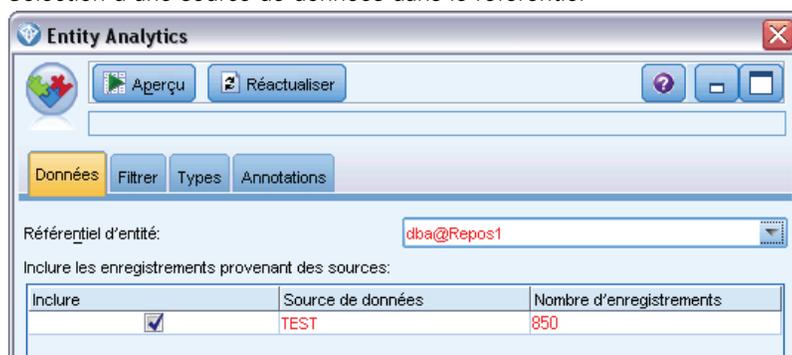
La première étape pour lire les identités résolues est d'ajouter un nœud source Entity Analytics(EA) au flux. Ce nœud source ne doit pas être connecté à quoi que ce soit d'autre à ce stade.

Figure 4-10
Nœud source Entity Analytics(EA)



Ouvrez le nœud source Entity Analytics(EA) et définissez les informations relatives au référentiel d'entité. Une liste apparaît alors : elle contient les sources de données qui ont été exportées vers le référentiel (dans ce cas, il n'y en a qu'une seule).

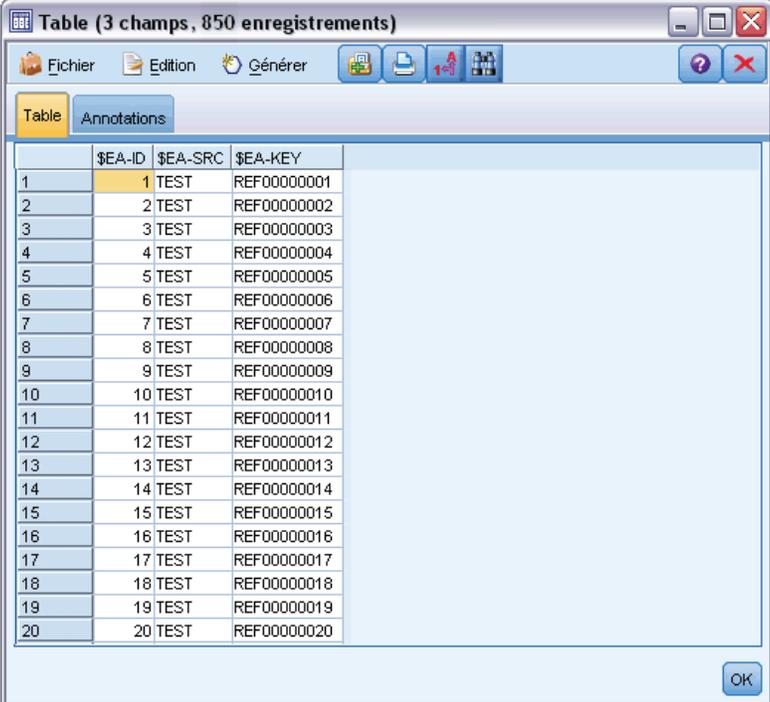
Figure 4-11
Sélection d'une source de données dans le référentiel



Sélectionnez la case de la source de données TEST et cliquez sur OK.

Examinons ce que le système d'analyses d'entités a eu comme effet sur les données. Reliez un nœud Table au nœud source Entity Analytics(EA), ouvrez le nœud Table et cliquez sur Exécuter pour afficher la fenêtre de sortie du nœud Table.

Figure 4-12
Sortie du noeud Table



	\$EA-ID	\$EA-SRC	\$EA-KEY
1	1	TEST	REF00000001
2	2	TEST	REF00000002
3	3	TEST	REF00000003
4	4	TEST	REF00000004
5	5	TEST	REF00000005
6	6	TEST	REF00000006
7	7	TEST	REF00000007
8	8	TEST	REF00000008
9	9	TEST	REF00000009
10	10	TEST	REF00000010
11	11	TEST	REF00000011
12	12	TEST	REF00000012
13	13	TEST	REF00000013
14	14	TEST	REF00000014
15	15	TEST	REF00000015
16	16	TEST	REF00000016
17	17	TEST	REF00000017
18	18	TEST	REF00000018
19	19	TEST	REF00000019
20	20	TEST	REF00000020

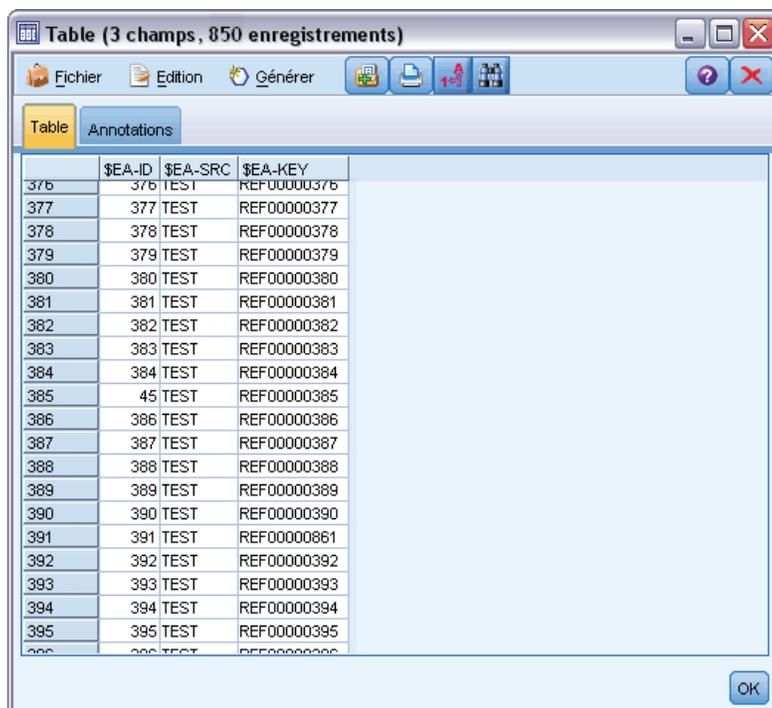
Seul un champ a une apparence familière – celui appelé *\$EA-KEY*. Il s'agit en fait du champ *clé* des données source qui se trouve là parce que vous l'avez choisi comme champ Clé unique dans le nœud Export EA.

Cependant, le système a ajouté deux autres champs. Le champ *\$EA-ID* est l'identifiant unique, ne constitue pas des enregistrements source, mais des identités résolues. Nous étudierons la différence dans un instant. Le champ *\$EA-SRC* identifie d'où viennent les données – ici, il s'appelle TEST parce qu'il s'agit de la balise source que vous lui avez attribuée dans le nœud Export EA.

Qu'est-il arrivé à tous les autres champs dans les données sources ? Ne vous inquiétez pas, ils se trouvent toujours dans le référentiel. Simplement, pour des raisons de performances, le nœud source Entity Analytics(EA) transmet uniquement l'ensemble de champs minimum en aval pour un traitement ultérieur.

Maintenant, recherchez la sortie du noeud Table à la ligne 385.

Figure 4-13
Différences entre les lignes de sortie Table et les nombres \$EA-ID

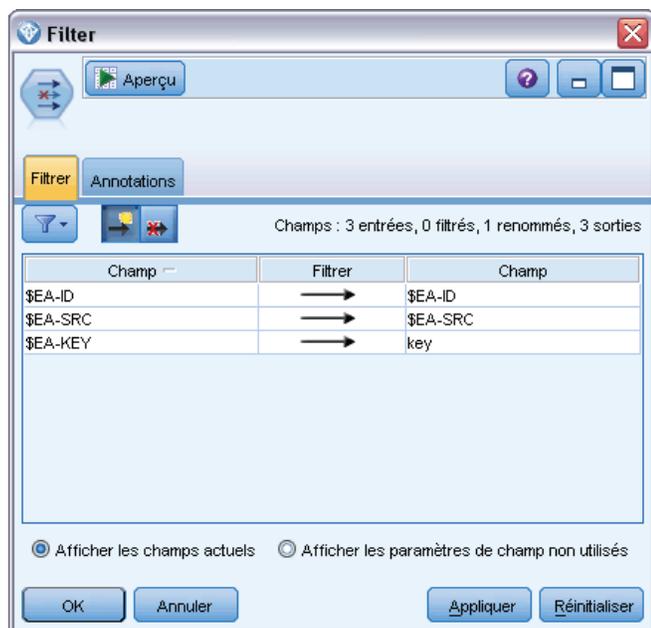


	\$EA-ID	\$EA-SRC	\$EA-KEY
376	376	TEST	REF00000376
377	377	TEST	REF00000377
378	378	TEST	REF00000378
379	379	TEST	REF00000379
380	380	TEST	REF00000380
381	381	TEST	REF00000381
382	382	TEST	REF00000382
383	383	TEST	REF00000383
384	384	TEST	REF00000384
385	45	TEST	REF00000385
386	386	TEST	REF00000386
387	387	TEST	REF00000387
388	388	TEST	REF00000388
389	389	TEST	REF00000389
390	390	TEST	REF00000390
391	391	TEST	REF00000861
392	392	TEST	REF00000392
393	393	TEST	REF00000393
394	394	TEST	REF00000394
395	395	TEST	REF00000395

Veillez noter comment le nombre *\$EA-ID* semble hors plage ici. Le système d'analyses d'entités a déterminé que l'enregistrement REF00000385 fait référence à la personne identifiée comme l'entité 45 qui a également l'enregistrement REF0000045. Si vous descendez jusqu'à la sortie, vous trouverez d'autres nombres hors plage, par exemple les lignes 485, 517, 520, etc. Examinons cela de plus près.

D'abord, il est nécessaire de souligner que le champ *\$EA-KEY* contient les données du champ *clé* dans les données sources en le renommant *clé*. Reliez un nœud Filtrer au nœud source Entity Analytics(EA) et ouvrez le nœud Filtrer. Faites un double clic sur la chaîne *\$EA-KEY* dans la deuxième colonne Champ et saisissez key.

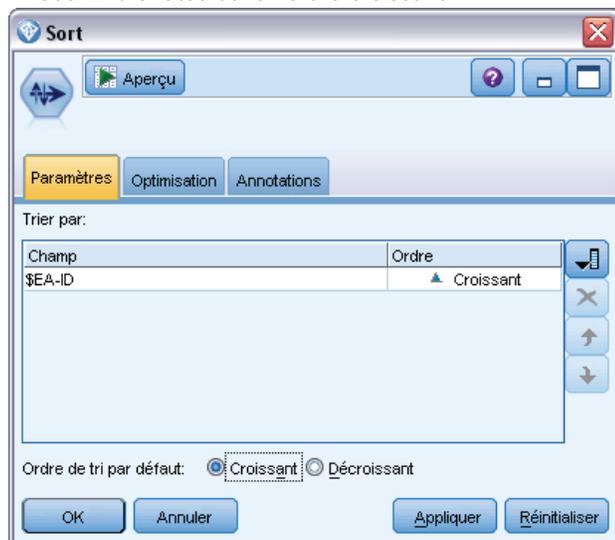
Figure 4-14
Nouveau nom du champ \$EA-KEY



Cliquez sur OK pour fermer le noeud Filtrer.

Nous devons à présent trier les ID de l'entité *\$EA-ID* dans l'ordre croissant. Reliez un noeud Trier au noeud Filtrer. Ouvrez le noeud Trier, cliquez sur le bouton du haut à côté de la table Trier par, sélectionnez *\$EA-ID* et cliquez sur OK.

Figure 4-15
Tri des ID d'entités dans l'ordre croissant



Laissez l'ordre de tri croissant et cliquez sur OK.

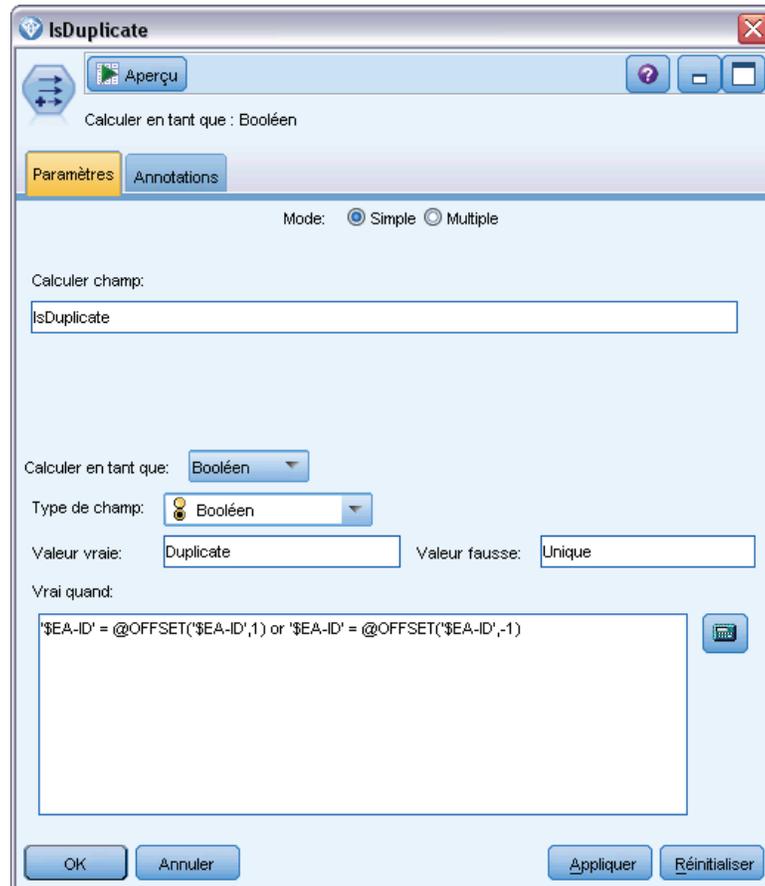
Vous devez maintenant créer un champ supplémentaire qui indique si un enregistrement est unique ou s'il s'agit d'un doublon. Reliez un noeud Calculer au noeud Trier. Ouvrez le noeud Calculer et définissez le nom Champ Calculer sur `IsDuplicate`. Dans la liste Calculer comme, choisissez Booléen ce qui définit également le Type de champ sur Booléen. Définissez le champ Valeur réelle sur Dupliquer et le champ Valeur fausse sur Unique.

Pour rechercher des enregistrements en double, utilisez une fonction de séquence spécifique appelée `@OFFSET`, fournie avec SPSS Modeler.

Saisissez ce qui suit dans le champ If :

```
'$EA-ID' = @OFFSET('$EA-ID',1) or '$EA-ID' = @OFFSET('$EA-ID',-1)
```

Figure 4-16
Définition de la condition dans le noeud Calculer



Une fois les ID d'entités triés par ordre croissant, la fonction `@OFFSET` vérifie que les ID d'entités adjacentes sont identiques auquel cas les enregistrements sont des doublons. Si c'est le cas, leur valeur *EstDoublon* est définie sur *Dupliquer*, sinon, elle est définie sur *Unique*.

Cliquez sur OK pour fermer le noeud.

Pour voir l'effet du noeud Calculer, reliez un noeud Table au noeud Calculer, ouvrez le noeud Table et cliquez sur Exécuter. Faites défiler la fenêtre de sortie du noeud Table jusqu'à la ligne 45.

Figure 4-17
Sortie du noeud Calculer

	\$EA-ID	\$EA-SRC	key	IsDuplicate
39	39	TEST	REF00000039	Unique
40	40	TEST	REF00000040	Unique
41	41	TEST	REF00000041	Unique
42	42	TEST	REF00000042	Unique
43	43	TEST	REF00000043	Unique
44	44	TEST	REF00000044	Unique
45	45	TEST	REF00000045	Duplicate
46	45	TEST	REF00000385	Duplicate
47	46	TEST	REF00000046	Unique
48	47	TEST	REF00000047	Unique
49	48	TEST	REF00000048	Unique
50	49	TEST	REF00000049	Unique
51	50	TEST	REF00000050	Unique
52	51	TEST	REF00000051	Unique
53	52	TEST	REF00000052	Unique
54	53	TEST	REF00000053	Unique
55	54	TEST	REF00000054	Unique
56	55	TEST	REF00000055	Unique
57	56	TEST	REF00000056	Unique
58	57	TEST	REF00000057	Unique

Souvenez-vous de quand nous avons examiné la sortie directement depuis le noeud source Entity Analytics(EA). Le système a déjà identifié que l'enregistrement REF00000385 fait référence au même individu que l'entité 45. Nous avons affiné cette opération et avons signalé le fait que les enregistrements REF0000045 et REF00000385 sont des doublons car ils font tous deux référence à l'entité 45.

Continuez à faire défiler la fenêtre de sortie vers le bas et vous verrez d'autres enregistrements signalés comme doublons.

Pour obtenir un rapport répertoriant les enregistrements en double, reliez un noeud Rapport (dans l'onglet Sortie de la palette des noeuds) au noeud Calculer *EstDoublon*. Ouvrez le noeud Rapport, copiez le texte suivant dans le champ entrée de l'onglet Modèle et cliquez sur Exécuter.

```
<html>
<h1>List of duplicate customer records.

<h2>This report was generated: [@TODAY]

<h2>Duplicate records
<table>
<tr>
<td>Entity ID</td>
<td>Key</td>
```

```
</tr>

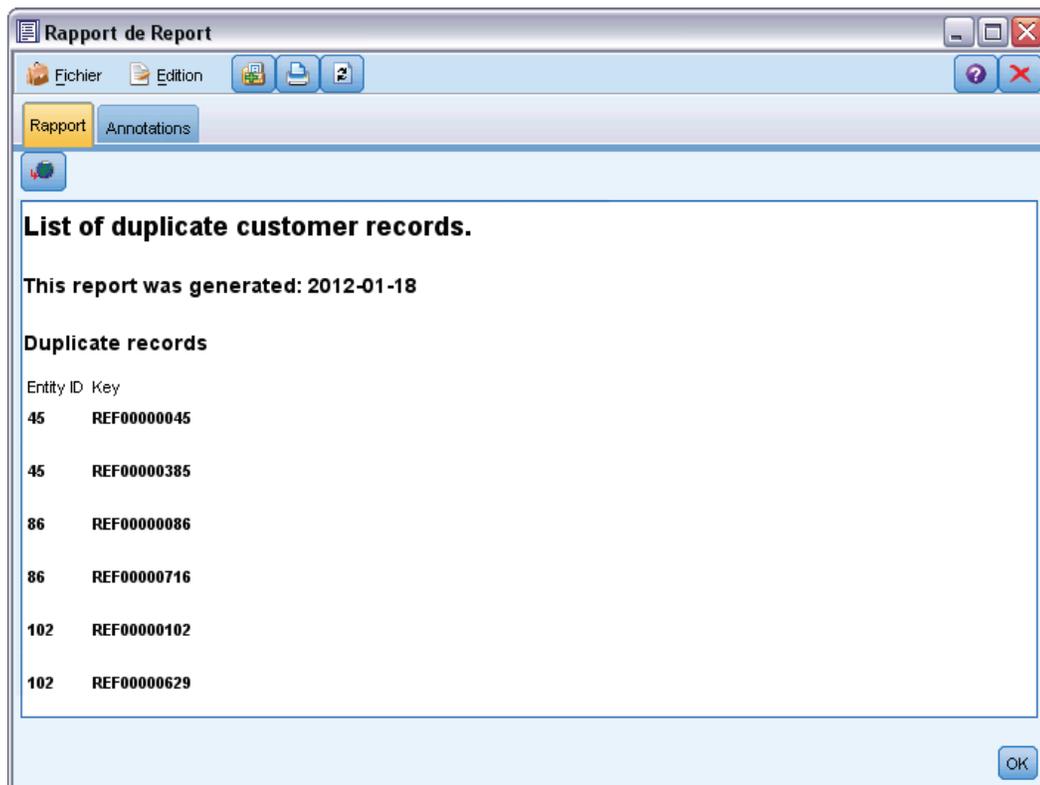
#WHERE IsDuplicate = "Duplicate"

<tr>
  <td>['$EA-ID']</td>
  <td>[key]</td>
<tr>
#
</table>

</html>
```

Cela donne la sortie suivante.

Figure 4-18
Sortie du noeud Rapport



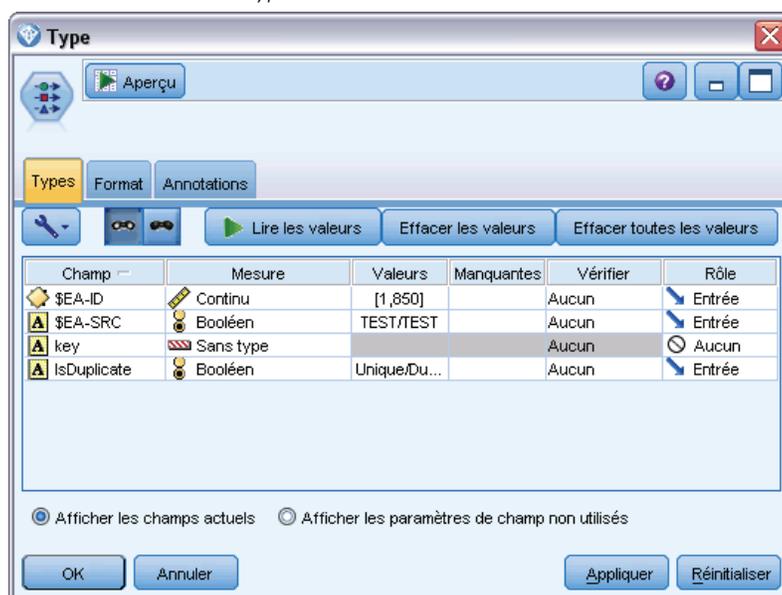
Le rapport utilise le format HTML dans ce cas bien que vous puissiez également utiliser le format XML ou ASCII.

Comparaison de la sortie des analyses d'entités au modèle d'origine

La dernière étape de cet exemple est de savoir si l'ajout des analyses d'entités fait une différence quant aux prévisions d'origine de la banque. Vous vous souvenez peut-être que le modèle d'origine avait prédit 13 mauvais payeurs sur les 150 demandes en attente. Vous allez utiliser un noeud Fusionner pour fusionner la sortie de ce modèle avec des informations sur les enregistrements en double des analyses d'entités pour savoir si cela modifie les prévisions.

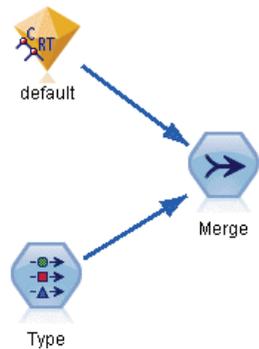
D'abord, vous devez vérifier que les nouveaux champs ajoutés par le système d'analyses d'entités ont des types de données corrects ou les *niveaux de mesure* comme ils sont appelés dans SPSS Modeler. Reliez un noeud Typer au noeud Calculer EstDouble, ouvrez le noeud Typer et cliquez sur le bouton Lire les valeurs.

Figure 4-19
Paramètres du noeud Typer



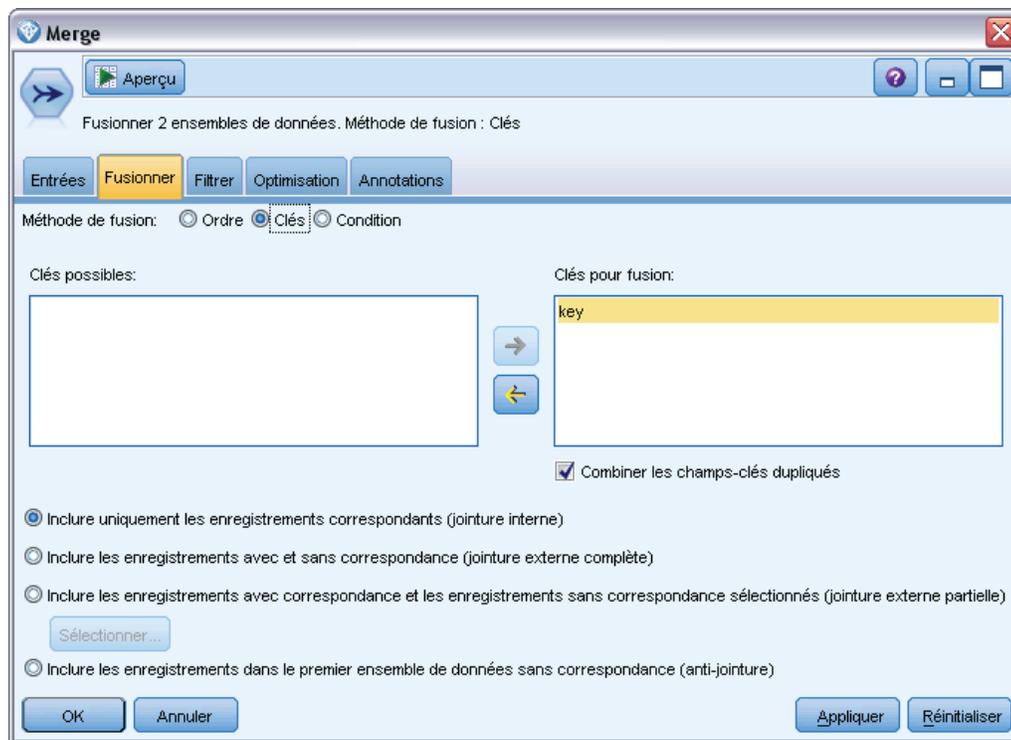
Vous pouvez maintenant ajouter le noeud Fusionner. Reliez-le au noeud Typer et connectez-le également au nugget doré qui contient le modèle d'origine. Pour ce faire, faites un clic droit sur le nugget doré, choisissez Connecter puis cliquez sur le noeud Fusionner qui devrait désormais contenir deux flèches d'entrée.

Figure 4-20
Entrées du noeud Fusionner



Ouvrez le noeud Fusionner, définissez la Méthode de fusion sur Clés, et cliquez sur la flèche de droite pour déplacer le champ clé de Clés possibles à Clés à fusionner puis cliquez sur OK.

Figure 4-21
Spécification du champ-clé pour la fusion



Vous êtes presque prêt à effectuer la comparaison. Cependant, si vous deviez relier un noeud Proportion et l'exécuter à ce stade, vous ne verriez aucun changement par rapport à la prévision d'origine. Bien que le flux fusionne désormais la sortie du nugget de modèle d'origine avec les nouveaux champs créés par le système d'analyses d'entités, le champ de prévision lui-même (*\$R-default*) dans le modèle de données a été mis à jour avec les nouvelles informations.

Pour ce faire, utilisez un noeud Remplacer qui peut remplacer les valeurs de champ. Reliez un noeud Remplacer au noeud Fusionner et ouvrez le noeud Remplacer.

Cliquez sur le bouton du haut à droite de Remplir les champs, descendez au bas de la liste, choisissez \$R-default et cliquez sur OK. Il s'agit du champ dont les valeurs doivent être changées si la condition spécifiée dans le reste de la boîte de dialogue est remplie.

Pour spécifier la condition, vérifiez que Remplacer est défini sur Basé sur une condition, puis dans le champ Condition, saisissez :

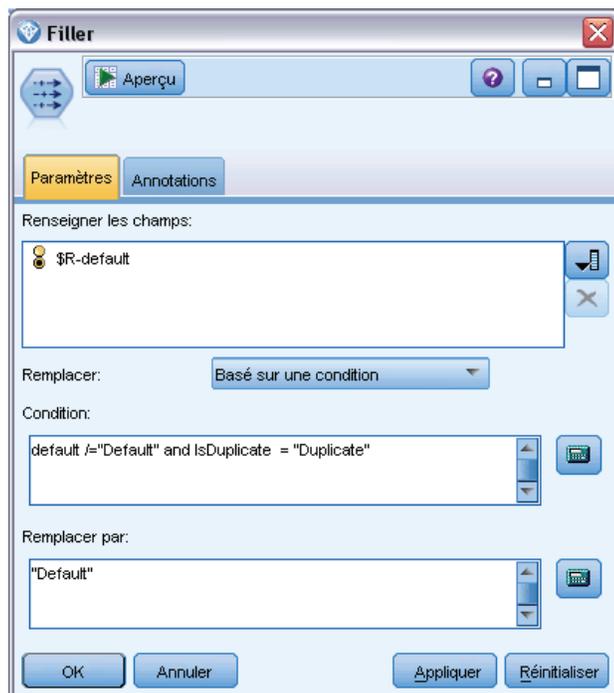
```
default != "default" and IsDuplicate = "Duplicate"
```

Dans le champ Remplacer par, saisissez :

```
"default"
```

Figure 4-22

Spécification de la condition de remplacement des valeurs de champs



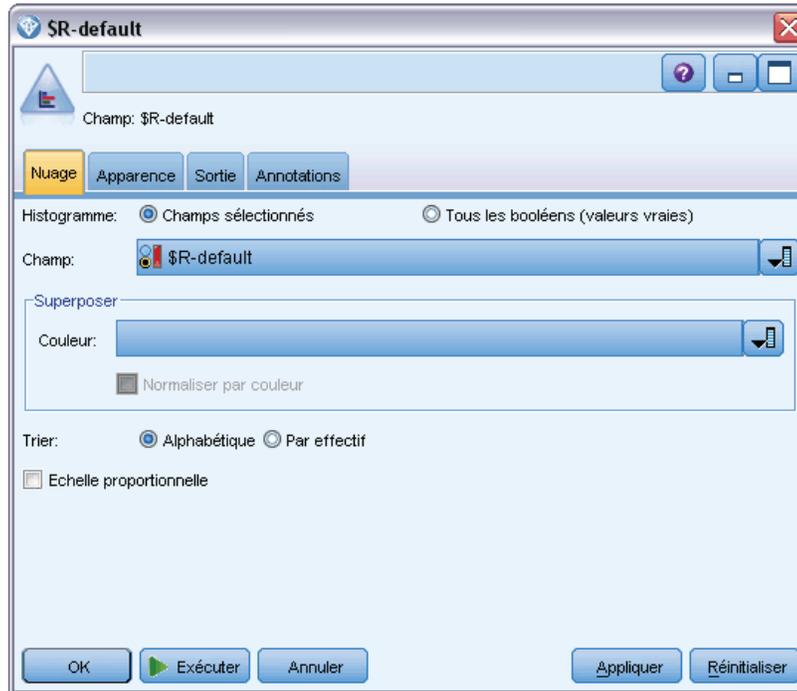
Ces paramètres nécessitent une courte explication. La condition stipule que pour chaque enregistrement où la valeur du champ *par défaut* de l'ensemble de données d'origine n'est pas égale à la valeur par défaut et où l'enregistrement a été signalé comme doublon, alors la valeur du champ *\$R-default* dans le modèle est définie sur la valeur par défaut.

Le champ *\$R-default* est le champ dans le modèle qui contient la prédiction concernant la probabilité d'un client à ne pas rembourser un prêt. Ainsi, les clients avec des enregistrements en double sont ajoutés au modèle comme mauvais payeurs potentiels.

Cliquez sur OK pour fermer le noeud Remplacer.

Vous êtes maintenant prêt à voir la différence générée par les analyses d'entités. Dans la palette Graphiques, reliez un noeud Proportion à un noeud Remplacer et ouvrez le noeud Proportion. Cliquez sur la liste Champ et choisissez \$R-default.

Figure 4-23
Paramètres du noeud Proportion



Cliquez sur Exécuter pour générer le diagramme de la nouvelle prédiction.

Figure 4-24
Sortie du noeud Proportion après analyses d'entités

Valeur	Proportion	%	Comptage
Default		10.67	16
Repaid		89.33	134

Il y a maintenant 16 demandes risquées au lieu des 13 précédentes. Ces demandes supplémentaires pourraient coûter cher si les individus étaient réellement de mauvais payeurs et à l'aide d'un graphique, vous pouvez démontrer à la banque l'avantage qu'apporte l'utilisation des analyses d'entités à leurs opérations d'évaluation des risques.

Récapitulatif

Cet exemple a démontré comment, grâce aux analyses d'entités, vous pouvez éliminer les enregistrements en double dans les données sur les individus et les organisations et améliorer ainsi la qualité des prévisions.

Remarque : idéalement, vous pourriez éliminer les enregistrements en double avant tout traitement. Vous pourriez poursuivre cette opération en utilisant un noeud de préparation automatisée de données (ADP) pour analyser vos données, identifier des corrections, filtrer des champs qui sont problématiques ou qui sont peu susceptibles d'être utiles, créer de nouveaux attributs le cas échéant et améliorer la performance au moyen de techniques de filtrage intelligentes.

Allier les analyses d'entités et la préparation automatisée de données vous permet de travailler avec des données aussi propres que possible.

Remarques

Ces informations ont été développées pour les produits et services offerts dans le monde.

Il est possible qu'IBM n'offre pas dans les autres pays les produits, services et fonctionnalités décrits dans ce document. Contactez votre représentant local IBM pour obtenir des informations sur les produits et services actuellement disponibles dans votre région. Toute référence à un produit, programme ou service IBM n'implique pas que les seuls les produits, programmes ou services IBM peuvent être utilisés. Tout produit, programme ou service de fonctionnalité équivalente qui ne viole pas la propriété intellectuelle IBM peut être utilisé à la place. Cependant l'utilisateur doit évaluer et vérifier l'utilisation d'un produit, programme ou service non IBM.

IBM peut posséder des brevets ou des applications de brevet en attente qui couvrent les sujets décrits dans ce document. L'octroi de ce document n'équivaut aucunement à celui d'une licence pour ces brevets. Vous pouvez envoyer par écrit des questions concernant la licence à :

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, États-Unis

Pour obtenir des informations de licence concernant la configuration de caractères codés sur deux octets (DBCS), veuillez contacter dans votre pays le département chargé de la propriété intellectuelle chez IBM ou envoyez vos commentaires par écrit à :

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japon.

Le paragraphe suivant ne s'applique pas au Royaume-Uni ni à aucun pays dans lequel ces dispositions sont contraires au droit local : INTERNATIONAL BUSINESS MACHINES FOURNIT CETTE PUBLICATION « EN L'ÉTAT » SANS GARANTIE D'AUCUNE SORTE, IMPLICITE OU EXPLICITE, Y COMPRIS, MAIS SANS ETRE LIMITE AUX GARANTIES IMPLICITES DE NON VIOLATION, DE QUALITE MARCHANDE OU D'ADAPTATION POUR UN USAGE PARTICULIER. Certains états n'autorisent pas l'exclusion de garanties explicites ou implicites lors de certaines transactions, par conséquent, il est possible que cet énoncé ne vous concerne pas.

Ces informations peuvent contenir des erreurs techniques ou des erreurs typographiques. Ces informations sont modifiées de temps en temps ; ces modifications seront intégrées aux nouvelles versions de la publication. IBM peut apporter des améliorations et/ou modifications des produits et/ou des programmes décrits dans cette publications à tout moment sans avertissement préalable.

Toute référence dans ces informations à des sites Web autres qu'IBM est fournie dans un but pratique uniquement et ne sert en aucun cas de recommandation pour ces sites Web. Le matériel contenu sur ces sites Web ne fait pas partie du matériel de ce produit IBM et l'utilisation de ces sites Web se fait à vos propres risques.

IBM peut utiliser ou distribuer les informations que vous lui fournissez, de la façon dont il le souhaite, sans encourir aucune obligation envers vous.

Les personnes disposant d'une licence pour ce programme et qui souhaitent obtenir des informations sur celui-ci pour activer : (i) l'échange d'informations entre des programmes créés de manière indépendante et d'autres programmes (notamment celui-ci) et (ii) l'utilisation mutuelle des informations qui ont été échangées, doivent contacter :

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, États-Unis.

Ces informations peuvent être disponibles, soumises à des conditions générales, et dans certains cas payantes.

Le programme sous licence décrit dans ce document et toute la documentation sous licence disponible pour ce programme sont fournis par IBM en conformité avec les conditions de l'accord du client IBM, avec l'accord de licence du programme international IBM et avec tout accord équivalent entre nous.

Toutes les données sur les performances contenues dans le présent document ont été obtenues dans un environnement contrôlé. Par conséquent, les résultats obtenus dans d'autres environnements d'exploitation peuvent varier de manière significative. Certaines mesures peuvent avoir été effectuées sur des systèmes en cours de développement et il est impossible de garantir que ces mesures seront les mêmes sur les systèmes commercialisés. De plus, certaines mesures peuvent avoir été estimées par extrapolation. Les résultats réels peuvent être différents. Les utilisateurs de ce document doivent vérifier les données applicables à leur environnement spécifique.

Les informations concernant les produits autres qu'IBM ont été obtenues auprès des fabricants de ces produits, leurs annonces publiques ou d'autres sources publiques disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances, leur compatibilité ou toute autre fonctionnalité associée à des produits autres qu'IBM. Les questions sur les capacités de produits autres qu'IBM doivent être adressées aux fabricants de ces produits.

Toutes les déclarations concernant la direction ou les intentions futures d'IBM peuvent être modifiées ou retirées sans avertissement préalable et représentent uniquement des buts et des objectifs.

Ces informations contiennent des exemples de données et de rapports utilisés au cours d'opérations quotidiennes standard. Pour les illustrer le mieux possible, ces exemples contiennent des noms d'individus, d'entreprises, de marques et de produits. Tous ces noms sont fictifs et toute ressemblance avec des noms et des adresses utilisés par une entreprise réelle ne serait que pure coïncidence.

Si vous consultez la version papier de ces informations, il est possible que certaines photographies et illustrations en couleurs n'apparaissent pas.

Marques commerciales

IBM, le logo IBM, ibm.com et SPSS sont des marques commerciales d'IBM Corporation, déposées dans de nombreuses juridictions du monde entier. Une liste à jour des marques IBM est disponible sur Internet à l'adresse <http://www.ibm.com/legal/copytrade.shtml>.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, et Pentium sont des marques commerciales ou des marques déposées de Intel Corporation ou de ses filiales aux États-Unis et dans d'autres pays.

Linux est une marque déposée de Linus Torvalds aux États-Unis et/ou dans d'autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques commerciales de Microsoft Corporation aux Etats-Unis et/ou dans d'autres pays.

UNIX est une marque déposée de The Open Group aux Etats-Unis et dans d'autres pays.

Java et toutes les marques et logos Java sont des marques commerciales de Sun Microsystems, Inc. aux Etats-Unis et/ou dans d'autres pays.

Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés.



Index

- analyses d'entités
 - comparées aux analyses prédictives, 3
 - défini, 1
 - utilisation avec IBM SPSS Modeler, 4
- balises source
 - référentiel d'entité, 20
- clés uniques
 - analyses d'entités, 7
 - référentiel d'entité, 20
- configuration
 - référentiel d'entité, 22, 31–32
- exportation
 - des données vers un référentiel d'entité, 8
- fonctionnalités
 - référentiel d'entité, 7, 19–20, 22, 24, 26, 37
- identités résolues, analyse, 32
- informations de type, définition pour les analyses d'entités, 34
- mappage de champs
 - sur les fonctionnalités du référentiel d'entité, 7, 19–20, 22, 24, 37
- marques commerciales, 67
- mentions légales, 66
- mise en correspondance des entités, définition du seuil, 31
- noeuds
 - ajout à un flux d'analyses d'entités, 34
- nœuds d'exportation
 - Entity Analytics, 8
- noeuds source
 - Entity Analytics, 11
- purge
 - référentiel d'entité, 44
- référentiel
 - administration des analyses d'entités, 41
 - analyses d'entités, 6–8, 14, 16–20, 22, 24, 26, 28, 31–32, 37, 44–45
- référentiel d'entité, 14
 - configuration, 22, 31–32
 - configuration des tâches de port, 41
 - connexion à IBM SPSS Modeler, 7
 - Création, 6, 16–17
 - définition, 14
 - fonctionnalités, 26
 - gestion, 24
 - options, 18
 - purge, 44
 - Suppression, 44–45
 - tâches administratives, 41
- renommer
 - champs de données pour les analyses d'entités, 33
 - résolution des identités, analyses d'entités, 8
- source de données, sélection d'analyses d'entités, 32
- sources de données
 - affichage des analyses d'entités, 18, 38
 - connexion aux analyses d'entités, 5, 15
- Suppression
 - référentiel d'entité, 44–45
- tâches de port
 - configuration pour les analyses d'entités, 41
- types d'entités
 - analyses d'entités, 28
 - référentiel d'entité, 20
- types d'utilisation, analyses d'entités, 28