IBM SPSS Modeler Entity Analytics 17.1 - Guide d'utilisation

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 59.

Certaines illustrations de ce manuel ne sont pas disponibles en français à la date d'édition.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- http://www.fr.ibm.com (serveur IBM en France)
- http://www.ibm.com/ca/fr (serveur IBM au Canada)
- http://www.ibm.com (serveur IBM aux Etats-Unis)

Compagnie IBM France
Direction Qualité
17, avenue de l'Europe
92275 Bois-Colombes Cedex

Cette édition s'applique à la version 17.1.0 d'IBM(r) SPSS(r) Modeler et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

Table des matières

Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	**IBM Canada**
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.

OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

France	Canada	Etats-Unis
(Pos1)		Home
Fin	Fin	End
(PgAr)		PgUp
(PgAv)		PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
(Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

Préface

IBM® SPSS Modeler est le puissant utilitaire d'exploration de données 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 prestations de services publics.

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 offre 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 d'IBM Business Analytics

Le logiciel IBM Business Analytics propose des informations complètes, cohérentes et précises auxquelles les preneurs de décisions peuvent se fier pour améliorer les performances de leur entreprise. Un porte-feuilles étendu de fonctions de veille économique, d'analyses prédictives, de fonctions degestion des performances et de stratégie financière et d'applications analytiques vous offre des informations claires, immédiates et décisionnelles sur les performances actuelles et vous permet de prévoir les résultats futurs. Ce logiciel intègre des solutions dédiées à l'industrie, des pratiques éprouvées et des services professionnels qui permettent aux organisations de toute taille de maximiser leur productivité, d'automatiser leurs décisions sans risque et de proposer de meilleurs résultats.

Ce porte-feuilles intègre le logiciel IBM SPSS Predictive Analytics qui aide les organisations à prévoir les événements à venir et à réagir en fonction des informations afin d'améliorer leurs résultats. Les clients de l'industrie du commerce, de l'éducation et des administrations du monde entier font confiance à la technologie IBM SPSS qui offre un avantage concurrentiel en attirant et fidélisant les clients et en améliorant la base de données de la clientèle tout en diminuant la fraude et en réduisant les risques. En utilisant le logiciel IBM SPSS dans leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives, capables de diriger et d'automatiser les décisions pour répondre aux objectifs commerciaux et obtenir un avantage concurrentiel mesurable. Pour des informations supplémentaires ou pour joindre un représentant, consultez http://www.ibm.com/spss.

Assistance technique

L'assistance technique est disponible pour les clients du service de maintenance. 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 contacter l'assistance technique, rendez-vous sur le site Web IBM Corp. à l'adresse http://www.ibm.com/support. Lorsque vous contactez l'assistance technique, soyez prêt à indiquer votre identité, le nom de votre société et votre contrat d'assistance.

Chapitre 1. Analyses d'entités

A propos des analyses d'entités

IBM SPSS Modeler Entity Analytics ajoute une dimension supplémentaire 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 disposiez des enregistrements client suivants provenant de deux sources différentes et que vous ne sachiez pas s'ils font référence à la même personne ou non.

Source 1

```
Record no.: 70001
Name:  Jon Smith
Address: 123 Main Street
Tax Reference: 555-00-1111
Driv. License: 0001133107
Cred. Card: 10229127
```

Source 2

```
Record no.: 9103
Name:  JOHNATHAN Smith
Date of Birth: 06/17/1934
Telephone: 555-1212
Cred. Card: 10229128
Email:  jls@mail.com
IP address: 9.50.18.77
```

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

Source 3

```
Record no.: 6251
Name:  Jon Smith
Telephone: 555-1212
Driv. License: 0001133107
Cred. Card: 10229132
```

Le numéro de permis de conduire (Driv. License) elie les enregistrements de la Source 1 et de la Source 3 alors que le numéro de téléphone relie les Sources 2 et 3.

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

Source 4

```
Record no.: S45286
Name:  John T Smith Jr
Address: 456 Main Street
Telephone: 703-555-2000
```

```
Date of birth: 03/12/1984


Record no.: S45287
Name:  John T Smith
Address: 456 Main Street
Telephone: 703-555-2000
Driv. License: 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

```
Record no.: 769582-2
Name:  John T Smith Sr
Address: 456 Main Street
Telephone: 703-555-2000
Driv. License: 009900991
Date of birth: 06/25/1959
```

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 IBM SPSS Modeler, effectuer le 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 différentes qui enregistrent de nombreux attributs différents avec quelques 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 (Entity Analytics - EA) 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'analyse.

Tableau 1. Différences entre analyses prédictives et analyses d'entités.

Caractéristiques	Analyses prédictives	Analyses d'entités
Types de données d'apprentissage	Basé sur des ensembles et des intervalles numériques relativement petits	Peut exploiter les grands ensembles (champs sans type) tels que des noms et des adresses
Taille des données d'apprentissage	Ignore généralement les grands ensembles (champs sans type)	Toutes les données sont utilisées

Tableau 1. Différences entre analyses prédictives et analyses d'entités (suite).

Caractéristiques	**Analyses prédictives**	**Analyses d'entités**
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 connus ou s'ils proviennent des mêmes individus mais avec des identités différentes

Chapitre 2. 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. Par exemple, certaines personnes peuvent apparaître plusieurs fois, ou différentes personnes peuvent sembler fusionnées ou manquantes. En quoi le produit 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.

- Lisez les données source dans IBM SPSS Modeler
- Créez un référentiel prêt à héberger les données
- Connectez IBM SPSS Modeler au référentiel
- Mappez les champs de données aux fonctions du référentiel
- Exportez les données dans le référentiel et résolvez les identités
- Analysez les identités résolues
- Résolvez les nouvelles observations en fonction du référentiel
- Générez les alertes nécessaires (par lots ou en temps réel)

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

IBM SPSS Modeler propose de nombreux noeuds qui couvrent toutes les fonctionnalités d'exploration de données standard. IBM SPSS Modeler Entity Analytics ajoute des noeuds spécifiques aux analyses d'entités. Il s'agit du noeud Export EA, du noeud source Entity Analytics (EA) et du noeud de processus Flux EA.

La figure suivante illustre ce processus.

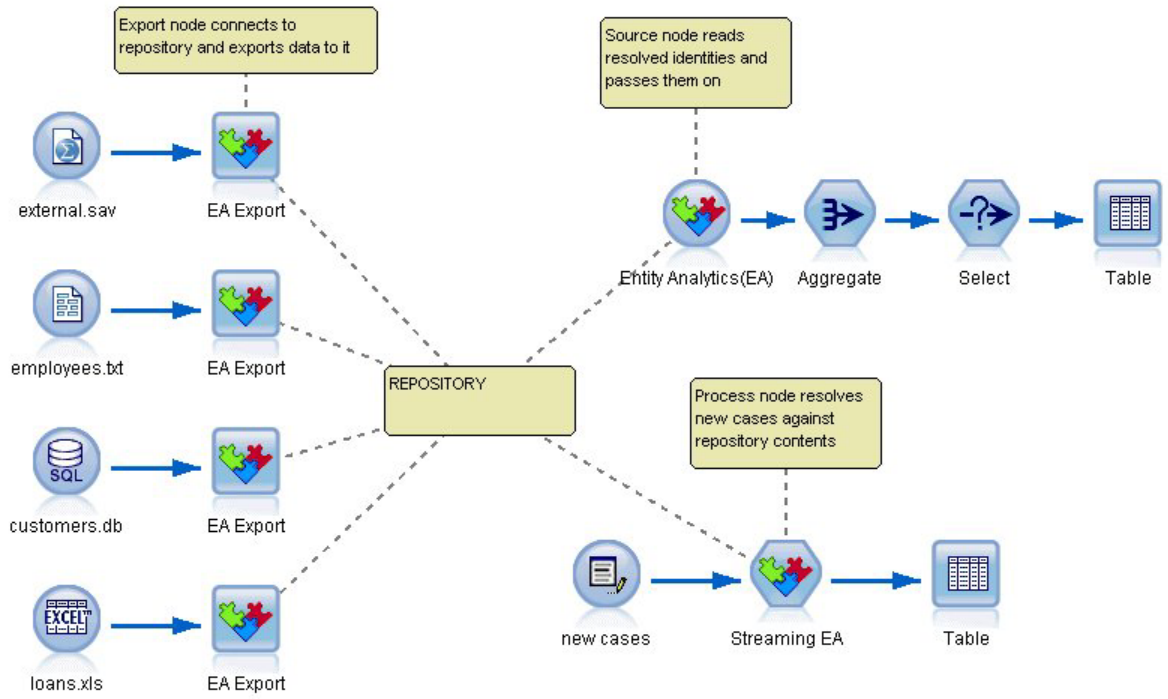


Figure 1. Le processus d'analyses d'entités

Etape 1 : Lecture des données source dans SPSS Modeler

Votre première tâche est de lire vos données dans SPSS Modeler au moyen d'un ou plusieurs noeuds source, signalés dans SPSS Modeler par une icône arrondie.

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 noeud source SPSS Modeler correspondant. Dans l'illustration, il s'agit d'un noeud source de base de données.

Chaque fichier de 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, voir la rubrique «Ajout d'un identificateur d'enregistrement unique», à la page 14.

Pour plus d'informations, voir «Connexion à une source de données», à la page 14.

Remarque : Les données de type caractères non latins ne sont pas prises en charge. Lorsque des données sont constituées d'un mélange d'enregistrements en jeu de caractères latins (pour l'Europe occidentale, par exemple) et non latins, seules les entrées correspondant aux caractères latins sont résolues.

Etape 2 : Création du 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 noeud Export EA, représenté par l'icône carrée.

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

Le processus de création d'un référentiel est décrit en détail ultérieurement. Pour plus d'informations, voir la rubrique «Définition d'un référentiel d'entités (noeud Export EA)», à la page 13.

Remarque : Si vous travaillez en mode serveur distant, vous devez créer le référentiel sur le poste Modeler Server (c'est-à-dire que vous devez être connecté à Modeler Server depuis Modeler Client lors de la création du référentiel de sorte que le référentiel EA soit créé sur la machine serveur).

Lorsque vous avez configuré un référentiel, vous pouvez gérer son contenu de plusieurs façons. Pour plus d'informations, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Etape 3 : Connexion de SPSS Modeler au référentiel

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

Pour plus d'informations, voir la rubrique «Options du référentiel d'entités», à la page 16.

Etape 4 : Mappage des champs d'entrée aux fonctions 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 donnés d'entités, alors que d'autres peuvent être propres à une source de données particulière. Dans un référentiel d'entités, ces différents types d'informations s'appellent des **fonctions**. Le référentiel propose de nombreuses fonctions standard et vous pouvez également créer vos propres fonctions.

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 sont propres 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.

Une fois que vous avez créé un référentiel et que vous 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 à leurs fonctions correspondantes dans le référentiel. Le mappage vers des fonctions prédéfinies indique au référentiel d'entités les champs à comparer et, surtout, comment les comparer. Le noeud Export EA propose une table de mappage dans laquelle vous pouvez créer des mappages.

Pour plus d'informations, voir la rubrique «Mappage de champs d'entrées à des fonctions (noeud Export EA)», à la page 17.

Etape 5 : Exportation des données vers le référentiel et résolution des identités

Chaque noeud source de données nécessite son propre noeud Export EA. Par conséquent, si vos données sont éparpillées entre plusieurs sources différentes, votre flux peut comporter plusieurs sources de données, chacune étant connectée à un noeud Export EA distinct. Pour plus d'informations, voir la rubrique «Utilisation des analyses d'entités avec IBM SPSS Modeler», à la page 5.

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. Lorsque deux enregistrements ou plus, associés à des identités ambiguës précédentes, peuvent 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 noeud source de données à son propre noeud Export EA, mappez les champs d'entrée aux fonctions du référentiel puis exécutez le flux pour exporter les données de SPSS Modeler vers le 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

Tableau 2. Données externes

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

Employés

Tableau 3. Employés

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

Clients

Tableau 4. Clients

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

Prêts

Tableau 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 *$EA ID*) et d'un champ d'indicateur source (appelé *$EA-SRC*), qui affiche la source de données d'où provient l'enregistrement. Ainsi, dans notre exemple, une fois que vous avez exporté les quatre sources de données, le contenu du référentiel ressemble à ceci.

Tableau 6. Exemple de contenu de référentiel après l'exportation.

$EA-ID	$EA-SRC	Nom	Téléphone	Adresse	Risque de crédit	Epargne	Prêt
1	Employés	Michael	555-1234	1234 5th St			
1	Externe	Mike	555-1234		560		
2	Clients	Joe		777 Oak St		5 $	
2	Externe	Joe	555-4567		780		

Tableau 6. Exemple de contenu de référentiel après l'exportation (suite).

$EA-ID	$EA-SRC	Nom	Téléphone	Adresse	Risque de crédit	Epargne	Prêt
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 le jeu de données *externe* est la même personne que *Michael* dans le jeu de données *Employés* en fonction d'un numéro de téléphone commun. Il leur affecte l'identifiant 1.

Le cas de *Joe* dans le jeu 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*. Le numéro de téléphone de Joseph est le même que celui de Joe dans le jeu de données *externe*. En fonction de cette information, le système détermine qu'ils sont une seule et même personne et leur donne l'identifiant 2.

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

Remarque : Nous avons choisi un exemple de correspondance positive pour les besoins de l'illustration. Vous auriez pu choisir un ensemble de règles plus pessimiste, dans lequel 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.

Etape 6 : Analyse des identités résolues

Après avoir résolu les conflits d'identités 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 noeud source Entity Analytics (EA) puis reliez-le au référentiel.

La sortie commune du noeud 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

En outre, si vous recherchez des relations, la sortie ci-après est générée. Pour plus d'informations, voir la rubrique «Sélection d'une source de données», à la page 26.

- Degré de séparation entre les entités (*$EA-DEGREE*)
- Champ parent (*$EA-PARENT*)
- Champ enfant (*$EA-CHILD*)
- Règle qui identifie la relation (*$EA-RULE*)

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

Le noeud source Entity Analytics(EA) sera décrit en détail ultérieurement. Pour plus d'informations, voir la rubrique «Analyse des identités résolues (noeud source Entity Analytics (EA))», à la page 26.

Etape 7 : Résolution de nouvelles observations en fonction 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 une meilleure évaluation ? C'est là que le noeud Flux EA entre en jeu.

D'abord, ajoutez un nouveau noeud source de données SPSS Modeler pour lire vos nouvelles données dans le flux. Ensuite, connectez ce noeud source à un noeud Flux EA. Pour consulter les résultats, vous pouvez ajouter un noeud Table, comme précédemment.

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

Pour illustrer cette situation, imaginons que le référentiel est constitué d'un contenu qui a été produit par le noeud source Entity Analytics (EA). Voir tableau 6, à la page 8.

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

Tableau 7. 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 noeud 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 noeud 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 ce qui suit.

Tableau 8. Sortie du noeud Flux EA.

$EA-ID	$EA-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 en utilisant un identifiant d'analyses d'entités comme clé d'agrégation et transmise à d'autres noeuds en aval pour un traitement ultérieur.

Le noeud Flux EA sera décrit en détail ultérieurement.

Etape 8 : Génération d'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 métier, il peut être nécessaire d'examiner ces données de plus près.

Vous pouvez, par exemple, joindre et exécuter un noeud Table ou un noeud Rapport SPSS Modeler, 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 noeud 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 métier. Une autre possibilité est d'exporter la sortie vers une base de données ou un autre support pour un traitement ultérieur. Avec IBM SPSS Modeler, vous disposez d'un large choix d'actions correspondant à vos besoins spécifiques.

Chapitre 3. Tâches liées aux analyses d'entités

A propos des tâches

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

- Définition d'un référentiel d'entités
- Configuration d'un référentiel d'entités
- Analyse des identités résolues
- Résolution de nouvelles observations en fonction du référentiel d'entités
- Purge d'un référentiel d'entités
- Suppression d'un référentiel d'entités
- Utilisation d'analyses d'entités avec d'autres produits IBM SPSS
- Administration des analyses d'entités

Définition d'un référentiel d'entités (noeud Export EA)

Le processus de définition d'un référentiel d'entités est composé des tâches ci-après.

1. Connexion à une source de données. Pour plus d'informations, voir la rubrique «Connexion à une source de données», à la page 14.
2. Création du référentiel. Pour plus d'informations, voir la rubrique «Création du référentiel», à la page 14.
3. Mappage des champs d'entrée de la source de données aux fonctions du référentiel. Pour plus d'informations, voir la rubrique «Mappage de champs d'entrées à des fonctions (noeud Export EA)», à la page 17.

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, voir la rubrique «Affichage des mappages de champs (noeud Export EA)», à la page 18.

Remarque : A partir de la version 16, SPSS Entity Analytics prend en charge les référentiels sur le produit IBM DB2. Etant donné qu'un référentiel est spécifique à une version de SPSS Modeler et qu'il ne peut pas être importé depuis une version antérieure, si vous disposez d'un référentiel existant, et que vous effectuez une mise à niveau vers SPSS Entity Analytics version 16, vous devrez recréer ce référentiel dans la nouvelle base de données DB2.

Le référentiel d'entités

Le référentiel fournit une zone de stockage centrale, qui tient lieu de cache de données pour toutes les informations sur les entités. Etant donné 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'entités. 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, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Important : A partir de la version 16, IBM SPSS Modeler Entity Analytics prend en charge les référentiels sur le produit IBM DB2 ; les versions précédentes de SPSS Entity Analytics prenaient en charge les

référentiels hébergés sur IBM solidDB. Si vous disposez d'un référentiel solidDB existant, vous devrez recréer ce référentiel dans la nouvelle base de données DB2 si vous effectuez une mise à niveau vers SPSS Entity Analytics version 16 ou suivante.

Remarque : La version d'IBM SPSS Modeler Entity Analytics fournie avec IBM SPSS Modeler Premium prend uniquement en charge un seul référentiel hébergé sur le produit IBM DB2 qui est livré avec SPSS Entity Analytics. Dans cette version, vous devez supprimer un référentiel existant avant de pouvoir en créer un nouveau. Une mise à niveau vers SPSS Entity Analytics faisant l'objet d'une licence distincte (appelée IBM SPSS Modeler Entity Analytics Unleashed) est disponible et vous permet de faire coexister plusieurs référentiels sur le même système ; chaque référentiel peut contenir plus de 10 millions de lignes et utiliser plus de quatre coeurs de processeur. Contactez votre représentant local du support 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 noeud source.

Pour vous connecter à une source de données

1. Dans l'onglet Sources de la palette de noeuds 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 noeud source à la grille d'écran.
2. Dans la grille d'écran, faites un double clic sur l'icône pour ouvrir la boîte de dialogue correspondante.
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 vos 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 identifiant chaque enregistrement de manière unique, ajoutez-en un à l'aide du noeud Dériver. Pour plus d'informations, voir la rubrique «Ajout d'un identificateur d'enregistrement unique».

Remarque : Les données de type caractères non latins ne sont pas prises en charge. Lorsque des données sont constituées d'un mélange d'enregistrements en jeu de caractères latins (pour l'Europe occidentale, par exemple) et non latins, seules les entrées correspondant aux caractères latins sont résolues.

Ajout d'un identificateur d'enregistrement unique

Chaque fichier de 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 de source de données ne contient pas ce champ, vous pouvez en ajouter un à l'aide du noeud Dériver de SPSS Modeler.

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

1. Dans la grille d'écran, cliquez sur le noeud source que vous avez ajouté lors de la tâche précédente.
2. Dans l'onglet **Ops sur champs** de la palette de noeuds, faites un double clic sur l'icône **Dériver** pour joindre un noeud Dériver au noeud source.
3. Dans la grille d'écran, faites un double clic sur le noeud Dériver pour ouvrir la boîte de dialogue correspondante.
4. Dans le champ **Dériver**, remplacez le nom par défaut par un nom significatif (comme **ID**) pour le champ d'identification que vous ajoutez.
5. Vérifiez que le champ **Dériver en tant que** est défini sur **Formule**.
6. Définissez le **Type de champ** sur `Continu`.
7. Dans la zone de texte **Formule**, entrez `@INDEX` et cliquez sur **OK**.

Création du référentiel

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

Remarque : Si vous travaillez en mode serveur distant, vous devez créer le référentiel sur le poste Modeler Server (c'est-à-dire que vous devez être connecté à Modeler Server depuis Modeler Client lors de la création du référentiel de sorte que le référentiel EA soit créé sur la machine serveur).

Pour créer un référentiel

1. Dans l'onglet Exporter de la palette de noeuds SPSS Modeler, placez un noeud Export EA dans le canevas de flux.

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

   a. Faites un clic droit sur le noeud source SPSS Modeler.
   b. Sélectionnez Connecter.
   c. Cliquez sur le noeud Export EA.
2. Faites un double clic sur le noeud Export EA pour ouvrir la boîte de dialogue correspondante.
3. Cliquez sur la liste **Référentiel d'entité**.
4. Cliquez sur **<Parcourir...>** pour afficher la boîte de dialogue Référentiels d'entité.
5. Dans la boîte de dialogue Référentiels d'entité, cliquez sur le champ Nom du référentiel.
6. Sélectionnez **<Créer un nouveau référentiel...>** pour afficher l'assistant Créer un référentiel.

Assistant Créer un référentiel

Etape 1

Au cours de cette étape, vous choisissez ou non de créer un référentiel local, à l'aide du produit IBM DB2 fourni avec IBM SPSS Modeler Entity Analytics, ou d'utiliser une base de données externe comme référentiel.

Créer un référentiel local. Indiquez un d'utilisateur et un mot de passe d'administrateur de base de données IBM DB2 qui doit héberger le référentiel que vous êtes en train de créer. Confirmez le mot de passe et cliquez sur **Suivant**.

Remarque : Vous ne pouvez utiliser aucun symbole de trait ou de trait de soulignement dans le nom d'utilisateur.

Les données d'identification qui doivent être utilisées pour la base de données IBM DB2 dépendent de votre système d'exploitation. Les utilisateurs UNIX doivent utiliser le nom d'utilisateur g2user et le mot de passe G2password.

Les tâches d'administration de référentiel au sein des noeuds Analyse des entités (telles que la création ou la destruction d'un référentiel) nécessitent des droits supplémentaires. Sur UNIX, l'utilisateur connecté dans IBM SPSS Modeler Server doit être le superutilisateur ou l'utilisateur g2user, et un membre du groupe db2iadm1. Sous Windows, l'utilisateur qui se connecte dans IBM SPSS Modeler Server doit être membre du groupe DB2ADMNS afin de pouvoir effectuer l'administration du référentiel.

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, voir la rubrique «Gestion des identifiants de l'administrateur pour la base de données de référentiel», à la page 33.

Remarque : Seule une combinaison nom d'utilisateur-mot de passe est possible. Tous les utilisateurs qui se connectent au référentiel partagent le même nom d'utilisateur et le même mot de passe.

Ajouter un référentiel externe. Utilisez cette option si vous souhaitez utiliser 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**.

Etape 2

Nouveau nom de référentiel. Indiquez 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, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Si vous choisissez un référentiel existant, entrez 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'entités

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

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 un nom d'utilisateur valide pour le référentiel sélectionné.
- **Mot de passe.** Mot de passe correspondant à ce nom d'utilisateur.
- **Connecter.** Cliquez pour vous connecter au référentiel actuel.

Gestion du référentiel. 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 les informations sur la source de données et 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 tout.** 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 que vous souhaitez supprimer tous les enregistrements de données du référentiel. Pour plus d'informations, voir la rubrique «Purge d'un référentiel d'entités», à la page 35.
- **Supprimer les éléments inutilisés.** Supprime les données source du référentiel mises en évidence mais conserve toutes les informations de configuration. Pour plus d'informations, voir la rubrique «Suppression des sources de données non utilisées d'un référentiel», à la page 36.
- **Renommer la source.** Ouvre une boîte de dialogue dans laquelle vous pouvez modifier le nom de la source de données mise en évidence.

  Remarque : La source de données est renommée à l'intérieur du référentiel ; vous devrez resélectionner ce nouveau nom de source de données dans les noeuds d'exportation ou de flux à partir duquel est elle référencée.

Détruire tout le référentiel. Détruit entièrement le contenu du référentiel actuel et ses informations de configuration. Pour plus d'informations, voir la rubrique «Suppression d'un référentiel d'entités», à la page 36.

Configurer le référentiel. Affiche une fenêtre dans laquelle vous pouvez configurer le référentiel actuel. Pour plus d'informations, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Mappage de champs d'entrées à des fonctions (noeud Export EA)

Le référentiel propose un certain nombre de fonctions prédéfinies standard. Des sources de données différentes peuvent utiliser des noms de champ différents, **Address1** ou **Address Line 1**) pour des types d'information correspondant à la même fonction. Pour éviter les doublons, il est nécessaire de mapper les champs de source d'entrée à des fonctions spécifiques du référentiel. Vous n'avez pas besoin de mapper chaque champ du jeu de données, uniquement ceux susceptibles de correspondre à la même fonction dans d'autres jeux de données.

Lorsqu'une source de données utilise des champs correspondant à d'autres types d'information qui ne sont pas prédéfinis dans le référentiel, vous pouvez créer de nouvelles fonctions depuis la fenêtre Configuration du référentiel. Pour plus d'informations, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Pour mapper des champs d'entrée à des fonctions

1. Joignez un noeud Export EA à un noeud source de données dans le canevas de flux. Chaque noeud source de données que vous utilisez doit être joint à son propre noeud Export EA.
2. Ouvrez le noeud Export EA pour afficher l'onglet Entrées qui contient les options de mappage des champs d'entrée. Pour plus d'informations, voir la rubrique «Options d'entrée du référentiel pour le mappage».
3. Dans le noeud Export EA, sélectionnez l'onglet Référentiel pour consulter les affectations 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 affectations de mappage (par exemple, à utiliser avec un autre noeud d'exportation pour une autre source de données), cliquez sur **Exporter le mappage**.

Lorsque vous avez terminé le mappage du premier noeud source de données, répétez le processus pour tous les autres noeuds 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 aux fonctions du référentiel prêtes à être exportées vers le référentiel. Configurez les affectations de mappage dans cet onglet et, facultativement, 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 **Importation du mappage**.

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'entités. 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, voir la rubrique «Options du référentiel d'entités», à la page 16.

Mapper vers type d'entité. Liste des types d'entité (c'est-à-dire des ensembles de fonctions) définis dans le référentiel. Choisissez-en un dans la liste ou sélectionnez **<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, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Balise source. Liste des balises indiquant les sources de données actuellement connues du référentiel. Choisissez-en une dans la liste ou sélectionnez **<Ajouter une nouvelle balise source...>** pour créer une balise pour une nouvelle source de données.

Clé unique. (obligatoire) 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 à une fonction correspondante dans le référentiel. S'il n'existe aucune fonction adaptée dans le type d'entité sélectionné, vous pouvez créer une nouvelle fonction ici.

- **Champ.** 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 à une fonction.** Pour mapper un champ à une fonction, faites un double clic sur cette colonne (ou appuyez sur la barre d'espace) sur la ligne du champ et choisissez une fonction dans la liste. Si aucune fonction appropriée n'est disponible, choisissez **<Ajouter une nouvelle fonction...>** pour afficher la fenêtre de configuration du référentiel où vous pouvez définir une nouvelle fonction pour ce type d'entité. Pour plus d'informations, voir la rubrique «Configuration d'un référentiel d'entités».
- **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 fonctions ADRESSE et TELEPHONE et vous pouvez créer vos propres types d'utilisation pour toutes les fonctions. 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 une utilisation...>** pour en créer une nouvelle. Pour plus d'informations, voir la rubrique «Gestion des types d'entités», à la page 23.

Importation du mappage. Importe un ensemble précédemment exporté de mappages de champs à des fonctions depuis un fichier XML externe. Cette fonction 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.

Exportation du mappage. Exporte vers un fichier XML externe l'ensemble des mappages de champs à des fonctions affiché dans la table de mappage.

Affichage des mappages de champs (noeud Export EA)

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

Afficher les entrées de. 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 pour l'option d'entrée sélectionnée.

Fonctions. Liste de toutes les fonctions ayant des mappages dans les sources de données affichées. Les fonctions non mappées ne sont pas affichées.

<Source de données>. Chaque colonne répertorie les champs mappés d'une source de données particulière pour chaque fonction pour laquelle un mappage a été défini.

Configuration d'un référentiel d'entités

Vous pouvez gérer le contenu du référentiel à partir de la fenêtre Configuration du référentiel qui propose une interface visuelle simple à utiliser pour la totalité du référentiel.

Si vous pensez utiliser plusieurs référentiels avec des configurations identiques ou 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, voir la rubrique «Réutilisation d'une configuration de référentiel», à la page 25.

Remarque : A partir de la version 16, SPSS Entity Analytics prend en charge les référentiels sur le produit IBM DB2. Etant donné qu'un référentiel est spécifique à une version de SPSS Modeler et qu'il ne peut pas être importé depuis une version antérieure, si vous disposez d'un référentiel existant, et que vous effectuez une mise à niveau vers SPSS Entity Analytics version 16, vous devrez recréer ce référentiel dans la nouvelle base de données DB2.
ATTENTION :
Si vous souhaitez modifier et enregistrer la configuration d'un référentiel qui contient déjà des données, vous serez peut-être invité à purger le contenu du référentiel et à charger de nouveau les données. Cela évite de conserver des incohérences dans le référentiel.

Pour définir une configuration de référentiel

1. Ouvrez un noeud 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 **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.

Tableau 9. Eléments principaux de la fenêtre Configuration du référentiel.

Section	Description	
Sources de données	Affiche les mappages de toutes les sources de données sur les différentes fonctions du référentiel.	Pour plus d'informations, voir la rubrique «Affichage des mappages de sources de données».
Fonctions	Crée une nouvelle fonction, ou duplique, modifie ou supprime une fonction existante.	Pour plus d'informations, voir la rubrique «Gestion des fonctions du référentiel», à la page 20.
Types d'entités	Crée un nouveau type d'entité ou gère des types existants (dupliquer, renommer, joindre ou supprimer des fonctions).	Pour plus d'informations, voir la rubrique «Gestion des types d'entités», à la page 23.
Règles de résolution	Définit le seuil de la mise en correspondance d'entités.	Pour plus d'informations, voir la rubrique «Définition du seuil de mise en correspondance des entités», à la page 25.

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 aux différentes fonctions du référentiel.

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 une source de données au référentiel à cet endroit. Vous ne pouvez ajouter des sources de données qu'en créant un noeud source SPSS Modeler et en le connectant à un noeud Export Entity Analytics. Pour plus d'informations, voir la rubrique «Connexion à une source de données», à la page 14.

Gestion des fonctions du référentiel

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 sont propres à une source de données particulière. Une fonction 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 Fonctions de la fenêtre Configuration du référentiel, l'entrée Toutes les fonctions propose un moyen de gérer toutes les fonctions du référentiel. Vous pouvez effectuer les actions suivantes.

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

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

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

Fonction. Nom de la fonction. Un symbole de cadenas situé en regard du nom d'une fonction indique que cette dernière est verrouillée. Les fonctionnalités verrouillées ne peuvent être ni supprimées, ni dupliquées, la seule modification apportée que vous pouvez enregistrer est la modification de l'attribut d'anonymisation.

Fréquence. Indique le nombre d'entités pouvant avoir la même valeur pour cette caractéristique. Les valeurs valides sont **Un** (pour un numéro de passeport, par exemple), **Quelques-uns** (pour une adresse, par exemple) ou **Beaucoup** (pour une date de naissance, par exemple).

Exclusivité. Indique qu'une entité doit généralement n'avoir qu'un seul de ce type de fonction. Par exemple, une date de naissance ou un numéro d'identité aura ici la valeur **Oui** alors que l'adresse ou un numéro de carte bancaire aura 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 probable qu'elle change et elle 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**.

Anonymiser. Indique si la fonction a été anonymisée. Les entrées sont **Oui** ou **Non**. Pour plus d'informations, voir la rubrique «Anonymisation des fonctionnalités du référentiel», à la page 22.

Pour créer une nouvelle fonction

1. Effectuez l'une des actions suivantes :
   - Cliquez sur le bouton Créer une nouvelle fonction (bouton en haut à droite de l'écran).

- Faites un clic droit sur **Toutes les fonctions** dans le volet de navigation à gauche de l'écran et choisissez **Nouvelle fonction**.

2. Remplissez la boîte de dialogue Ajouter/modifier une fonction. Pour plus d'informations, voir la rubrique «Ajout ou modification d'une fonction».

Pour dupliquer une fonction existante

1. Dans la colonne **Fonction** du tableau à droite de l'écran, sélectionnez la fonction à dupliquer.
2. Cliquez sur le bouton Dupliquer la fonction sélectionnée (deuxième bouton à droite de l'écran).
3. Remplissez la boîte de dialogue Ajouter/modifier une fonction. Pour plus d'informations, voir la rubrique «Ajout ou modification d'une fonction».

Pour modifier une fonction existante

ATTENTION Si vous modifiez, supprimez ou anonymisez 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 **Fonction** du tableau à droite de l'écran, sélectionnez la fonction à modifier. *Remarque* : Vous ne pouvez modifier que les fonctions que vous avez créées, pas les fonctions fournies par le système.
2. Cliquez sur le bouton Modifier la fonction sélectionnée (troisième bouton à droite de l'écran).
3. Remplissez la boîte de dialogue Ajouter/modifier une fonction. Pour plus d'informations, voir la rubrique «Ajout ou modification d'une fonction».

Pour supprimer une fonction existante

ATTENTION Si vous modifiez, supprimez ou anonymisez 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 **Fonction** du tableau à droite de l'écran, sélectionnez la fonction à supprimer. *Remarque* : Vous ne pouvez supprimer que les fonctions que vous avez créées, pas les fonctions fournies par le système.
2. Effectuez l'une des actions suivantes :
   - Cliquez sur le bouton Supprimer la fonction sélectionnée (bouton en bas à droite de l'écran).
   - Faites un clic droit sur **Toutes les fonctions** dans le volet de navigation à gauche de l'écran et choisissez **Supprimer**.
3. Cliquez sur **Continuer** pour confirmer la suppression de la fonction.

   ATTENTION :
   Il est impossible d'annuler la suppression d'une fonctionnalité.

Ajout ou modification d'une fonction

ATTENTION Si vous modifiez, supprimez ou anonymisez 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 fonction, vous pouvez créer une nouvelle fonction de référentiel ou dupliquer ou modifier une fonction existante.

Remarque : Si une fonction existante est verrouillée, vous ne pouvez pas éditer ses détails dans cette boîte de dialogue.

Type de fonction. Libellé indiquant le type d'informations auquel est associée la fonction. Ce libellé forme la première partie de l'identifiant de la fonction.

Description. Courte description du type de fonction, dans un but informatif uniquement.

Fréquence. Indique le nombre d'entités pouvant avoir la même valeur pour cette caractéristique. Les valeurs valides sont **Un** (pour un numéro de passeport, par exemple), **Quelques-uns** (pour une adresse, par exemple) ou **Beaucoup** (pour une date de naissance, par exemple).

Exclusivité. Indique qu'une entité doit généralement n'avoir qu'un seul de ce type de fonction. Par exemple, une date de naissance ou un numéro d'identité aura ici la valeur **Oui** alors que l'adresse ou un numéro de carte bancaire aura 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 probable qu'elle change et elle 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**.

Table des éléments. Liste des éléments que cette fonction contient.

- **Elément.** Nom de l'élément.
- **Description.** Courte description de ce que l'élément propose.
- **Type de données.** Type de données pouvant être utilisé pour cet élément. Les types disponibles sont les suivants : Chaîne, Entier, Réel et Date.

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 dans la table des éléments. Vous ne pouvez pas annuler cette opération.

ATTENTION Si vous modifiez, supprimez ou anonymisez 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.

Anonymiser. Pour la protection des données, vous pouvez choisir d'anonymiser les données lors de leur ajout dans un référentiel ; pour activer cela pour une fonction, sélectionnez **Oui**. Pour plus d'informations, voir la rubrique «Anonymisation des fonctionnalités du référentiel».

Anonymisation des fonctionnalités du référentiel

Dans le cadre de la sécurité des données, vous pouvez si vous le souhaitez anonymiser les données au fur et à mesure de leur ajout dans le référentiel afin de réduire le risque de divulgation par inadvertance des informations d'identification.

Lorsque des données anonymisées sont exportés vers un référentiel, une méthode d'anonymisation est requise et elle doit permettre la résolution d'entité à l'aide des données anonymisées. Par exemple, si deux enregistrements de données des détails de carte de crédit d'une personne sont anonymisés sous la forme "anon_s21" et "anon_s9271", ils perdent leur relation ; toutefois, si un lien interne, d'arrière-plan, est utilisé entre les deux enregistrements, le système peut encore comprendre que l'un de ces noms est une version abrégée de l'autre.

Les liens et identificateurs d'arrière-plan qui permettent de relier vos données anonymisées sont générés lors de la création d'un référentiel et ils sont uniques pour ce dernier. Les données chiffrées sont enregistrées en interne et lues lorsqu'un flux se connecte à un référentiel.

Lorsque vous configurez votre référentiel, vous pouvez indiquer si chaque fonction individuelle sera ou non anonymisée. Si une fonction est anonymisée, tous ses éléments sont anonymisées et elle est toujours anonymisée quel que soit son type d'utilisation. Pour plus d'informations, voir la rubrique «Ajout ou modification d'une fonction», à la page 21.

Remarque : Assurez-vous de ne pas anonymiser tous les champs de SPSS Entity Analytics ou vous ne pourrez pas identifier les données en retour. Nous vous recommandons de laisser au moins un champ (même s'il s'agit uniquement d'un numéro de ligne) sans anonymat de manière à pouvoir contrôler ultérieurement une nouvelle fusion de vos données d'origine.

Une colonne de la liste de fonctions dans la fenêtre Configuration du référentiel indique les fonctions qui sont définies pour l'anonymisation. Les entrées sont **Oui** ou **Non**.

Remarque : Si un référentiel existant contient des données avant l'anonymisation des données, vous devez purger d'abord toutes les données, sinon il n'y aura aucune correspondance entre entre les fonctions anonymisées et les fonctions non anonymisées.

Gestion des types d'entités

Un **type d'entité** est un ensemble nommé de fonctions de référentiel qui sont regroupées par ressemblance. Par exemple, un type d'entité devant être utilisé avec un jeu de données client peut être composé de fonctions 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)
- Joindre des fonctions à un type d'entité
- Supprimer des fonctions d'un type d'entité
- Renommer un type d'entité
- Supprimer un type d'entité

Type d'entité. Nom du type d'entité sélectionné.

Fonction. Liste des fonctions valides incluses dans ce type d'entité.

Type d'utilisation. (Facultatif) Indique les différents contextes dans lesquels cette fonction 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 indiquées ici définissent les valeurs affichées sur le noeud Export EA ou Flux EA quand un utilisateur clique sur la colonne Utilisation d'une fonction dans l'onglet Entrées. Pour plus d'informations, voir la rubrique «Options d'entrée du référentiel pour le mappage», à la page 17.

Informations générales sur les types d'utilisation :

- • Les types d'utilisation sont des libellés arbitraires.
- • Vous pouvez créer un type d'utilisation à partir de presque tout champ d'entrée ; cependant, vous ne pouvez pas entrer des espaces et des caractères non valides.
- • Ce que vous entrez est automatiquement converti en majuscules au fur et à mesure de la saisie.
- • Vous pouvez avoir autant de types d'utilisation que vous le souhaitez.

- • Il n'est pas nécessaire que les types d'utilisation soient significatifs, mais cela peut vous être utile ultérieurement si vous utilisez une convention de dénomination qui ait un sens pour vous et pour les autres utilisateurs.
- • Lorsque vous effectuez un mappage, un avertissement s'affiche dans une police de couleur rouge si vous avez mappé les éléments avec différents types d'utilisation.

Généralement, une erreur s'affiche si vous essayez de mapper deux champs au même élément de fonction. Les types d'utilisation constituent un moyen de mapper plus de deux champs au même élément de fonction afin de pouvoir les mettre en relation.

Par exemple, si vous avez défini deux fonctions distinctes : *HOMEADDRESS* et *WORKADDRESS*, il n'y aurait aucune relation entre elles. Si une entité a une fonction *HOMEADDRESS* qui est identique à la fonction *WORKADDRESS* d'une autre entité, il n'y a pas de relation car il s'agit de fonctions différentes. Toutefois, si vous réutilisez une seule fonction avec différents types d'utilisation, la résolution comprend que la fonction *ADDRESS.WORK* est identique à la fonction *ADDRESS.HOME*.

Vous pouvez réutiliser les types d'utilisation de différentes fonctions ou en avoir des différents ; par exemple, *HM* et *WK* pour le téléphone et *HOME* et *WORK* pour l'adresse. Cela n'a pas d'importance car nous n'établissons pas de relations entre les téléphones et les adresses ; toutefois, si vous maintenez une certaine cohérence, cela vous aidera plus tard à identifier et à regrouper des champs.

Lorsque plusieurs types d'entité sont entrés dans un référentiel unique, si vous utilisez la même fonction, peut importe leurs types d'utilisation. Par exemple, si vous définissez *WK* et *HM* comme types d'utilisation de *ADDRESS* pour le type d'entité *COMPANY*, il y aura encore une relation avec *WORK* et *HOME* comme types d'utilisation de *ADDRESS* pour *PERSON*.

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. Joignez des fonctions au type d'entité (voir section suivante).

Pour joindre des fonctions 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 fonction (bouton en haut à droite de l'écran).
3. Dans la liste des fonctions disponibles, choisissez une ou plusieurs fonctions (utilisez Ctrl-Clic pour choisir plusieurs fonctions) et cliquez sur OK.

Pour supprimer des fonctions 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 fonctions dans la table des fonctions jointes à droite de l'écran. Utilisez Ctrl-Clic pour sélectionner plusieurs fonctions.
3. Cliquez sur le bouton Annuler l'adjonction de la fonction (bouton en 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. Joignez des fonctions au type d'entité ou supprimez-en du type d'entité en fonction de vos besoins (voir instructions précédentes).

Pour renommer un type d'entité

ATTENTION Si vous modifiez, supprimez ou anonymisez 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

ATTENTION Si vous modifiez, supprimez ou anonymisez 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.
   ATTENTION :
   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.

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

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

Choisissez **Définir sur une résolution par défaut** pour revenir au seuil par défaut d'un des autres paramètres.

Choisissez **Définir sur une résolution classique** si vous trouvez trop de correspondances.

Pour générer un référentiel à la fois pour les entités et les relations, sélectionnez **Relations d'inclusion**. Cette option n'est disponible que si vous disposez de la mise à niveau faisant l'objet d'une licence distincte (appelée IBM SPSS Modeler Entity Analytics Unleashed).

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). Cela est possible uniquement au sein d'une configuration existante. Par exemple, vous ne pouvez pas migrer une configuration de référentiel d'une version de IBM SPSS Modeler vers une autre, ou d'un type de base de données vers un autre.

Pour réutiliser une configuration existante

1. Affichez la fenêtre Configuration du référentiel pour le référentiel dont vous souhaitez utiliser la configuration. Pour plus d'informations, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.
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 (noeud source Entity Analytics (EA))

Une fois les données exportées, vous pouvez utiliser le noeud source Entity Analytics (EA) pour transmettre les identités résolues à d'autres noeuds 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 noeud source Entity Analytics (EA) à un flux.
2. Ouvrez le noeud Entity Analytics(EA).
3. Dans l'onglet Données, sélectionnez le référentiel d'entités et une ou plusieurs sources de données (cliquez sur **Actualiser** pour mettre à jour le nombre d'enregistrements). Pour plus d'informations, voir la rubrique «Sélection d'une source de données».
4. Ajoutez d'autres noeuds au flux pour effectuer le traitement désiré. Pour plus d'informations, voir la rubrique «Ajout de noeuds au flux», à la page 28.

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.

Référentiel d'entités. 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, voir la rubrique «Options du référentiel d'entités», à la page 16.

Inclure les enregistrements provenant des sources de données. Cette table répertorie les différentes sources de données qui ont été entrées dans le référentiel, ainsi que le nombre d'enregistrements dans chaque source. Cochez la case **Inclure** associée aux sources de données que vous souhaitez utiliser pour des analyses et des traitements ultérieurs. Pour sélectionner ou désélectionner toutes les sources de données, cliquez sur **Tout inclure** ou sur **Tout exclure** respectivement.

Relations. Sélectionnez le type de relations à inclure dans le référentiel. Cette option n'est disponible que si vous disposez de la mise à niveau faisant l'objet d'une licence distincte (appelée IBM SPSS Modeler Entity Analytics Unleashed) et si le référentiel est configuré pour l'inclusion des relations.

- **Pas de relations.** Les détails de relation ne sont pas utilisés.
- **Relations proches.** Sélectionne uniquement les entités étroitement liées. Le caractère proche d'une relation dépend de nombreuses variables, par exemple les propriétés des fonctions mappées, les fonctions qui partagées et la nature de la résolution (classique ou agressive).
- **Toutes les relations.** Sélectionne toutes les entités liées.

Degré maximal de séparation. Disponible uniquement si **Fermer les relations** ou **Toutes les relations** est sélectionné. Sélectionnez le nombre de degrés de séparation à utiliser pour identifier une relation. Par exemple, si Ann et Bob ne se connaissent pas mais que John connaît à la fois Ann et Bob, alors Ann et Bob sont liées par deux degrés de séparation.

Type d'entité de sortie. Par défaut, si le référentiel contient des détails, le premier type d'entité répertorié dans le référentiel est affiché. Si le référentiel en comporte plusieurs, la sélection ici d'un type d'entité modifié les fonctions affichées sous l'onglet Filtrer pour l'affichage des fonctions de ce type. Vous pouvez sélectionner l'un des types d'entité utilisés dans le référentiel.

Renommer des champs de données

Vous pouvez utiliser l'onglet Filtre 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 jeu de données.

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

Tableau 10. Champs d'identités résolues

Champ	Description
$EA-ID	Identificateur 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 de source de données

Remarque : Bien que vous puissiez également utiliser l'onglet Filtre 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 en vue d'un traitement ultérieur.

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

Tableau 11. Propriétés de types pour les champs

Propriété	Description
Mesure	Niveau de mesure (c'est-à-dire le type de données) utilisé pour décrire les caractéristiques des données du champ.
Valeurs	Fournit des options de lecture des valeurs de données du jeu de données.
Manquant	Permet de spécifier le traitement des valeurs manquantes du champ.

Tableau 11. Propriétés de types pour les champs (suite)

Propriété	Description
Vérifier	Options de validation permettant de vérifier que les valeurs des champs sont conformes aux valeurs ou intervalles spécifiés.
Rôle	Indique la façon dont le champ sera utilisé si les données sont transmises au noeud de modélisation ou au nugget de modèle.

Ajout de noeuds au flux

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

- Noeud Agréger ou Distinguer pour résumer la sortie qui peut être de taille importante
- Noeud Sélectionner pour sélectionner un sous-ensemble de la sortie
- Noeud Table pour visualiser la sortie du noeud source Entity Analytics (EA)
- Noeud Rapport pour imprimer la sortie dans un rapport
- Noeud 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 noeuds Opérations d'enregistrement, Sortie et Export dans le manuel *IBM SPSS Modeler Source, Process and Output Nodes Guide*.

Comparaison de nouvelles observations au contenu du référentiel (noeud Flux EA)

Lorsque vous avez déjà exécuté des résolutions d'identités dans le référentiel, vous pouvez utiliser le noeud Flux EA pour comparer de nouvelles observations rencontrées ultérieurement au contenu du référentiel. Ce noeud 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 noeud Export EA, le noeud Flux EA utilise un noeud source SPSS Modeler unique comme entrée. Cependant, le noeud Flux EA diffère sur les points suivants. Alors que le noeud Export EA génère des enregistrements pour toutes les entités associées à ses enregistrements d'entrée, le noeud Flux EA génère des enregistrements uniquement pour les entités associées aux entités déjà résolues dans le référentiel. Pour plus d'informations, voir la rubrique «Sortie du noeud Flux EA», à la page 31.

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, voir la rubrique «Connexion à une source de données», à la page 14.
2. Dans l'onglet Ops sur enregistrements, joignez un noeud Flux EA au noeud source de données.
3. Faites un double clic sur le noeud Export EA pour ouvrir la boîte de dialogue correspondante.
4. Cliquez sur la liste **Référentiel d'entité**.
5. Cliquez sur **<Parcourir...>** pour afficher la boîte de dialogue Référentiels d'entité.
6. Dans la boîte de dialogue Référentiels d'entité, 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 mapper. Pour plus d'informations, voir la rubrique «Gestion des types d'entités», à la page 23.
10. Mapper les champs d'entrées de la source de données aux fonctions du référentiel. Pour plus d'informations, voir la rubrique «Mappage de champs d'entrée aux fonctions (noeud Flux EA)».
11. Vous pouvez mettre à jour les enregistrements dans le référentiel en temps réel lors de l'évaluation de vos données. Pour plus d'informations, voir la rubrique «Mappage de champs d'entrée aux fonctions (noeud Flux EA)».
12. Cliquez sur l'onglet **Sorties** 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, voir la rubrique «Affichage des mappages de champs et des sources de données (noeud Flux EA)», à la page 30.
13. Cliquez sur l'onglet **Filtrer** pour afficher les détails des champs de saisie et des fonctions stockées dans le référentiel. Les fonctions qui n'ont pas été mappées dans le noeud sont filtrées par défaut ; vous pouvez toutefois modifier cela si nécessaire.
14. Cliquez sur **OK** lorsque le noeud est correctement configuré.
15. Joignez un noeud Table au noeud Flux EA et exécutez le flux.

La fenêtre des résultats du noeud 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, voir la rubrique «Sortie du noeud Flux EA», à la page 31.

Remarque : Il peut arriver que vous rencontriez une erreur sous la forme **Un nombre incorrect de champs a été détecté dans le modèle de données du serveur** lors de l'exécution du noeud Flux EA. Cela peut se produire si vous avez modifié la configuration du référentiel depuis la création du noeud Flux EA. La modification de la configuration dans ces circonstances peut avoir pour effet de modifier le nombre et les noms des champs issus du noeud. Pour résoudre ce problème, ouvrez le noeud Flux EA et cliquez sur le bouton **Actualiser**. Cela permet de recalculer le nombre et les noms des champs de sortie.

Mappage de champs d'entrée aux fonctions (noeud Flux EA)

L'onglet Entrées contient les options de mappage des champs de l'entrée de ce noeud aux fonctions du référentiel. Configurez les affectations de mappage dans cet onglet ou sélectionnez l'onglet **Vue** pour afficher des détails sur toutes les sources de données du 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 **Importation du mappage**.

Référentiel d'entités. 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, voir la rubrique «Options du référentiel d'entités», à la page 16.

Mapper vers type d'entité. Liste des types d'entité (c'est-à-dire des ensembles de fonctions) définis dans le référentiel. Choisissez-en un dans la liste ou sélectionnez **<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, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.

Conserver les recherches. Sélectionnez cette option si vous voulez mettre à jour les enregistrements dans le référentiel en temps réel lors de l'évaluation de vos données.

Balise source. Uniquement disponible lorsque vous sélectionnez **Conserver les recherches**. Liste des balises indiquant les sources de données actuellement connues du référentiel. Choisissez-en une dans la liste ou sélectionnez **<Ajouter une nouvelle balise source...>** pour créer une balise pour une nouvelle source de données.

Clé unique. Uniquement disponible lorsque vous sélectionnez **Conserver les recherches**. 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 à une fonction correspondante dans le référentiel. S'il n'existe aucune fonction adaptée dans le type d'entité sélectionné, vous pouvez créer une nouvelle fonction ici.

- **Champ.** 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 à une fonction.** Pour mapper un champ à une fonction, faites un double clic sur cette colonne (ou appuyez sur la barre d'espace) sur la ligne du champ et choisissez une fonction dans la liste. Si aucune fonction appropriée n'est disponible, choisissez **<Ajouter une nouvelle fonction...>** pour afficher la fenêtre de configuration du référentiel où vous pouvez définir une nouvelle fonction pour ce type d'entité. Pour plus d'informations, voir la rubrique «Configuration d'un référentiel d'entités», à la page 18.
- **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, voir la rubrique «Gestion des types d'entités», à la page 23.

Importation du mappage. Importe un ensemble précédemment exporté de mappages de champs à des fonctions depuis un fichier XML externe. Cette fonction 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.

Exportation du mappage. Exporte vers un fichier XML externe l'ensemble des mappages de champs à des fonctions affiché dans la table de mappage.

Affichage des mappages de champs et des sources de données (noeud Flux EA)

Dans l'onglet Sortie, 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 noeud est traitée, afin de rechercher et de récupérer des entités correspondantes. Cliquez sur **Actualiser** pour mettre à jour le nombre d'enregistrements.

Inclure les enregistrements provenant des sources de données. Cette table répertorie les différentes sources de données disponibles dans le référentiel, ainsi que le nombre d'enregistrements dans chaque source.

Correspondances. Ces options spécifient à quel niveau les informations de mappage de champs à des fonctions 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.

Remarque : Si plus de 20 correspondances sont trouvées, seules les 20 premières sont renvoyées.

- **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 correspondances possibles.** Utilisez ce paramètre lorsque vous souhaitez renvoyer à la fois des entités correspondantes et des entités qui partagent les mêmes identifiants (ceux ayant des fonctions configurées avec une valeur de fréquence de `Un`, par exemple ceux qui correspondent aux numéros de carte bancaire, aux numéros de TVA, etc.).
- **Inclure toutes les correspondances.** Utilisez cette option lorsque vous souhaitez voir le plus grand nombre possible d'entités partageant des fonctions 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 fonctions

(généralement celles avec une valeur de fréquence de `Un` ou de `Quelques-uns`). Par exemple, deux entités avec le même numéro de TVA et des entités avec des adresses similaires seraient incluses.

Relations. Disponible uniquement si le référentiel est configuré pour l'inclusion des relations. Pour configurer le référentiel afin qu'il inclue les relations, vous devez disposer de la mise à niveau sous licence distincte (appelée IBM SPSS Modeler Entity Analytics Unleashed). Sélectionnez le type de relations à inclure dans la sortie.

- **Pas de relations.** Les détails de relation ne sont pas utilisés.
- **Relations proches.** Sélectionne uniquement les entités étroitement liées. Le caractère proche d'une relation dépend de nombreuses variables, par exemple les propriétés des fonctions mappées, les fonctions qui partagées et la nature de la résolution (classique ou agressive).
- **Toutes les relations.** Sélectionne toutes les entités liées.

Degré maximal de séparation. Disponible uniquement si **Fermer les relations** ou **Toutes les relations** est sélectionné. Sélectionnez le nombre de degrés de séparation à utiliser pour identifier une relation. Par exemple, si Ann et Bob ne se connaissent pas mais que John connaît à la fois Ann et Bob, alors Ann et Bob sont liées par deux degrés de séparation.

Type d'entité de sortie. Par défaut, si le référentiel contient des détails, le premier type d'entité répertorié dans le référentiel est affiché. Si le référentiel en comporte plusieurs, la sélection ici d'un type d'entité modifié les fonctions affichées sous l'onglet Filtrer pour l'affichage des fonctions de ce type. Vous pouvez sélectionner l'un des types d'entité utilisés dans le référentiel.

Sortie du noeud Flux EA

La sortie du noeud Flux EA est composée des champs suivants pour chaque enregistrement récupéré.

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

Si les champs de relation sont activés dans le référentiel et que le degré de séparation est supérieur à zéro sous l'onglet Sortie, la sortie du noeud Mise en flux EA est contient également les champs ci-après pour chaque enregistrement récupéré.

Champ	Description
$EA-DEGREE	Degré de séparation.
$EA-PARENT	Identificateur de l'enregistrement à partir duquel est calculé la séparation.
$EA-CHILD	Identificateur de l'enregistrement dans lequel est calculé la séparation.

Champ	Description
$EA-RULE	

Utilisation d'IBM SPSS Modeler Entity Analytics avec d'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 pour pouvoir utiliser les fonctions d'IBM SPSS Modeler Entity Analytics avec ces produits. Pour plus d'informations, reportez-vous au guide *IBM SPSS Modeler Premium Installation*.

Après l'installation, vous devez utiliser le client IBM SPSS Collaboration and Deployment Services Deployment Manager pour créer une définition de serveur de référentiel Entity Analytics. Cette étape est nécessaire pour l'utilisation d'un flux IBM SPSS Modeler qui contient un noeud Entity Analytics dans un travail IBM SPSS Collaboration and Deployment Services (en d'autres mots, elle est nécessaire pour exécuter les flux Entity Analytics dans IBM SPSS Collaboration and Deployment Services). La définition de serveur doit correspondre au nom de référentiel dans le flux ; cette définition est utilisée pour dire au flux où trouver le réferentiel et pour lui fournir les informations de connexion dont il a besoin.

Tâches administratives

Pour les référentiels qui sont créés dans Entity Analytics, un nouveau service de base de données est créé avec le produit IBM DB2. Il existe peu de tâches administratives associées à DB2. Ces tâches généralement exécutées par l'administrateur de la base de données ou l'administrateur système sont les suivantes :

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

Il existe d'autres tâches administratives pouvant s'avérer nécessaires et s'appliquant à tous les référentiels :

- Déplacement du référentiel vers un autre répertoire de stockage
- Définition des propriétés de flux pour les champs date/heure et horodatage
- Ajustement des paramètres de dépassement de délai
- Exécution d'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'entités
- Suppression d'un référentiel d'entités
- Suppression d'un référentiel quand aucune connexion n'est possible

Configuration des affectations de port

Chaque service de base de données DB2 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 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 IBM SPSS Modeler Server, l'ordinateur qui exécute IBM SPSS Modeler).

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

```
# configuration de la plage de ports pour les analyses d'entités

#

#     cette plage de ports définit les ports que les bases de données DB2

#     (créées pour stocker les référentiels Entity Analytics)

#     peuvent utiliser.  Configurez cela si la plage de ports par défaut

#     introduit un conflit dans le système.

#

# 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 de référentiel

Le nom d'utilisateur et le mot de passe de l'administrateur de la base de données DB2 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 DB2 SQL.

Pour démarrer l'éditeur DB2 SQL

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

   ```
   cd rép_install_modeler\ext\bin\pasw.entityanalytics\DB2\bin
   ```

   où *rép_install_modeler* représente le répertoire dans lequel SPSS Modeler est installé.
3. Entrez :

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

   où *version* est le numéro de version de l'installation SPSS Modeler et *nom_référentiel* 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 afin d'afficher l'invite `solsql>`.

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

1. A l'invite `solsql>`, entrez :

   ```
   alter user nomutil identified by motdepasse;
   commit work;
   ```

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

2. Entrez exit; pour fermer l'éditeur.
3. Redémarrez le client SPSS Modeler.

Pour des informations sur d'autres tâches administratives à exécuter avec la base de données DB2, consultez la documentation pour la version appropriée de IBM DB2 sur http://publib.boulder.ibm.com/.

Déplacement du référentiel vers un autre répertoire de stockage

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)
- *rép_install_modeler*/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 déplacer le référentiel vers un autre répertoire

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 *<chemin d'installation modeler server>/ext/bin/pasw.entityanalytics/ea.cfg* en ajoutant l'option suivante :

   `repository_data_directory, nouvel_emplacement`

   où *nouvel_emplacement* représente le répertoire dans lequel vous avez déplacé le répertoire EA, par exemple *F:\data\EA*.

Définition des propriétés de flux pour les champs date/heure et horodatage

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

Pour définir le format des propriétés de flux

1. Dans le menu principal SPSS Modeler, sélectionnez :

   Outils > **Propriétés du flux** > **Options.**
2. Sélectionnez **Date/heure**.
3. Définissez le **format de date** sur **AAA-MM-JJ**.
4. Définissez le **format d'heure** sur **HH:MM:SS**.
5. Cliquez sur **OK**.

Ajustement des paramètres de dépassement de délai

Sur les systèmes lents ou lourdement chargés, si vous rencontrez des erreurs lors de la création de répertoires ou de l'accès à des répertoires, il peut s'avérer nécessaire d'augmenter les paramètres de dépassement de délai concernant le démarrage et l'arrêt du moteur Entity Analytics ou du serveur de base de données Entity Analytics.

Pour ajuster le paramètre de dépassement de délai pour le moteur Entity Analytics

1. Quittez SPSS Modeler.
2. Modifiez le fichier *<chemin d'installation modeler server>/ext/bin/pasw.entityanalytics/ea.cfg* afin d'augmenter la valeur de l'option suivante :

`timeout, valeur`

où *valeur* représente la valeur du délai en secondes pour le moteur Entity Analytics (la valeur par défaut est 60).

Pour ajuster le paramètre de dépassement de délai pour la base de données Entity Analytics (DB2 uniquement)

1. Quittez SPSS Modeler.
2. Modifiez le fichier *<chemin d'installation modeler server>/ext/bin/pasw.entityanalytics/ea.cfg* afin d'augmenter la valeur de l'option suivante :

`timeout, valeur`

où *valeur* représente la valeur du délai en secondes pour la base de données DB2 Entity Analytics (la valeur par défaut est 100).

Exécution d'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 sur le client SPSS Modeler et sur SPSS Modeler Server sur le même système Windows, par défaut le client et le serveur partagent 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 ports et un dossier de référentiel différents.

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

1. Ouvrez le fichier *<chemin d'installation modeler [server]>/ext/bin/pasw.entityanalytics/ea.cfg* en édition.
2. Mofidiez les paramètres `min_port` et `max_port` afin qu'ils utilisent des ports différents de l'autre système. Pour plus d'informations, voir la rubrique «Configuration des affectations de port», à la page 32.
3. Modifiez le paramètre `repository_data_directory` afin qu'il utilise un répertoire différent de l'autre système.
4. Enregistrez le fichier *ea.cfg*, puis fermez-le.

Purge d'un référentiel d'entités

Si vous souhaitez nettoyer les enregistrements de données d'un référentiel d'entité, tout en conservant les informations de configuration, vous pouvez purger les données du référentiel.

Pour purger toutes les données d'un référentiel :

1. Ouvrez un noeud Entity Analytics (Analyse des entités).
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 **Connecter**.
7. Lorsque le bouton **Purger tout** est activé, cliquez dessus.
8. Dans la boîte de dialogue Purger tout, cliquez sur **Purger** pour confirmer la purge du référentiel.

Suppression des sources de données non utilisées d'un référentiel

Si vous n'utilisez plus une source de données ou si vous n'en avez plus besoin dans un référentiel d'entités, vous pouvez supprimer cette source du référentiel. Vous pouvez sélectionner une ou plusieurs sources à supprimer.

Pour supprimer une source de données sélectionnée dans un référentiel :

1. Ouvrez un noeud Entity Analytics (Analyse des entités).
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 dans lequel vous voulez supprimer une source de données.
6. Si vous n'êtes pas encore connecté, saisissez le nom et le mot de passe administrateur et cliquez sur **Connecter**.
7. Dans la liste **Gestion du référentiel**, sélectionnez la source de données à supprimer. Si nécessaire, utilisez la combinaison de touches Ctrl-clic pour sélectionner d'autres sources de données.
8. Lorsque le bouton **Supprimer les éléments inutilisés** est activé, cliquez dessus.
9. Dans la boîte de dialogue Supprimer les sources de données inutilisées, cliquez sur **Supprimer** pour confirmer la purge du référentiel.

Suppression d'un référentiel d'entités

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

Attention : Il est vraiment supprimé. **Cette opération est irréversible.** Si vous n'êtes pas sûr de vouloir effectuer une suppression totale, utilisez le bouton **Purger** pour supprimer toutes les sources de données. Cela ne supprime pas la configuration du référentiel. Pour plus d'informations, voir la rubrique «Purge d'un référentiel d'entités», à la page 35.

Remarque : La procédure suivante suppose que vous puissiez vous connecter au référentiel à partir 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 permettant de supprimer un référentiel lorsque l'on ne parvient pas à s'y connecter. Pour plus d'informations, voir la rubrique «Suppression d'un référentiel lorsque aucune connexion n'est possible».

Pour supprimer un référentiel

1. Ouvrez un noeud Entity Analytics (Analyse des entités).
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 **Connecter**.
7. Lorsque le bouton **Supprimer tout le référentiel** 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 lorsque aucune connexion n'est possible

Utilisez la procédure suivante si vous souhaitez supprimer une entité de référentiel et que vous ne pouvez pas vous y connecter, soit à cause de problèmes de connectivité avec SPSS Modeler soit parce que vous avez oublié votre nom d'utilisateur ou votre mot de passe.

Effectuez cette procédure sur la machine qui héberge la date de données de référentiel.

Systèmes Windows

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

   ```
   cd rép_install_modeler
   cd ext\bin\pasw.entityanalytics
   delete_repository.bat nom_référentiel
   ```

   où *rép_install_modeler* représente le répertoire d'installation de SPSS Modeler et *nom_référentiel* représente le nom du référentiel.

   Remarque : Le nom du référentiel est sensible à la casse.
3. Passez à la rubrique "Terminer la procédure" de cette section.

Systèmes UNIX

1. Ouvrez un shell (interpréteur de commandes).
2. Entrez :

   ```
   cd rép_install_modeler
   cd ext/bin/pasw.entityanalytics
   ./delete_repository.sh nom_référentiel
   ```

   où *rép_install_modeler* représente le répertoire d'installation de SPSS Modeler Server et *nom_référentiel* représente le nom du référentiel.

   Remarque : Le nom du référentiel est sensible à la casse.

Terminer la procédure (tous les systèmes)

1. A l'invite, confirmez la suppression du référentiel en entrant 0.
2. 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.

Chapitre 4. Fonctionnement des analyses d'entités

A propos de cet exemple

Dans cet exemple, nous verrons en quoi l'ajout des 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 IBM SPSS Modeler qui contient également IBM SPSS Modeler Entity Analytics. Ce répertoire *Demos* est accessible à partir du groupe de programmes IBM SPSS Modeler dans le menu Démarrer de Windows. Le fichier *loan_entity_analytics.str* se trouve dans le répertoire *Entity_Analytics*.

Remarque : Pour pouvoir exécuter cet exemple de flux, vous devez créer un référentiel sur votre système. Faites-le avant de poursuivre cet exemple. Pour plus d'informations, voir la rubrique «Création du référentiel», à la page 14.

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 certain temps 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 été remboursés ou 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 2. Flux d'origine avec noeud de modélisation

En plus des informations sur les prêts passés, le jeu 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 du jeu de données ne sont pas utiles pour les prévisions, par exemple, les champs de nom peuvent être ignorés. Le noeud Type 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 3. Rôles des champs définis dans le noeud Type

Comme 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.

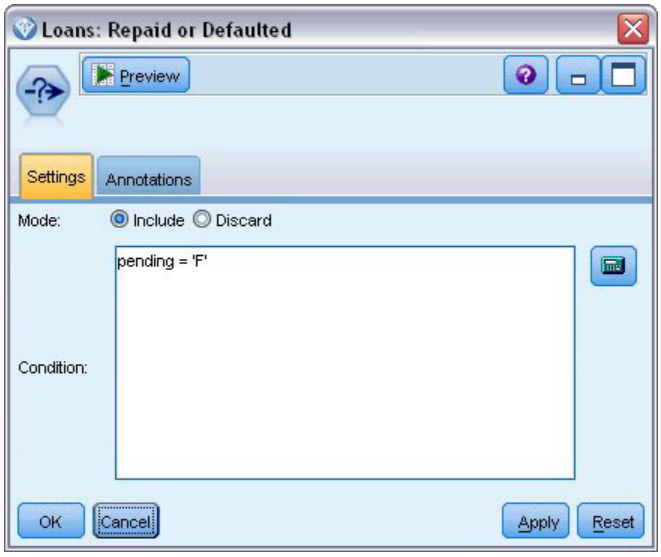


Figure 4. Ignorer les demandes de prêt en attente

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 C&R qui sera utilisé pour créer un modèle permettant de prévoir les mauvais payeurs potentiels en fonction des performances passées des clients de la banque.

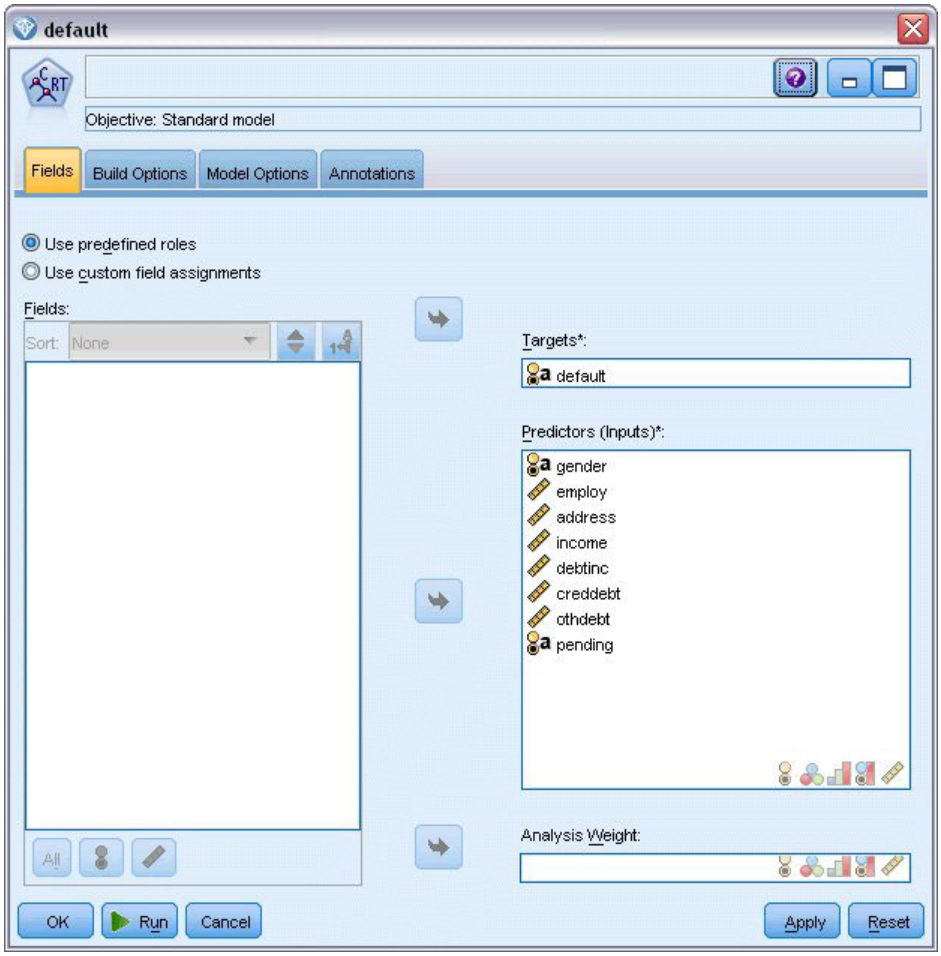


Figure 5. Attribution des champs prédicteur et champs cible

Les champs utilisés pour effectuer la prévision sont créés en tant que champs 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 Type l'a défini auparavant.

L'exécution de ce flux génère un nugget de modèle contenant le modèle qui a été créé à partir des champs prédicteur.

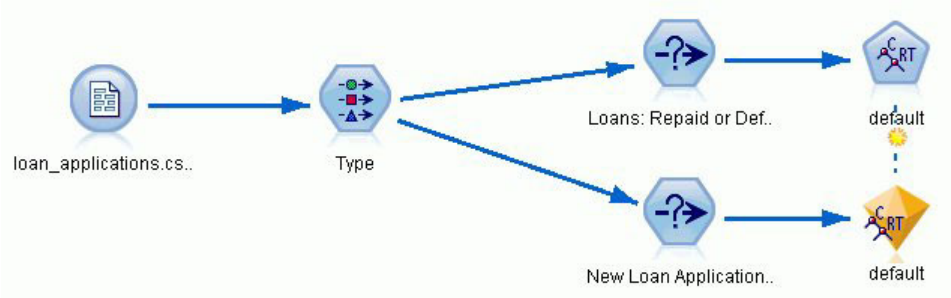


Figure 6. Flux avec nugget de modèle ajouté

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 du jeu de données d'origine, l'analyste insère un noeud Sélectionner qui, cette fois, contient seulement les 150 enregistrements de prêt marqués comme En attente, au lieu de les ignorer. L'analyste transmet ces enregistrements directement au modèle et ajoute un noeud Distribution pour une représentation visuelle des prévisions du modèle.

Figure 7. Flux sélectionnant les nouvelles demandes de prêts avec un noeud Distribution ajouté

Le noeud Distribution affiche la proportion des valeurs du champ *$R-default* dans le modèle. Ce champ est ajouté au modèle de données par le noeud Arbre C&R 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 les effets de l'ajout d'analyses d'entité.

En exécutant cette partie du flux, l'analyste apprend grâce aux résultats du noeud Distribution 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.

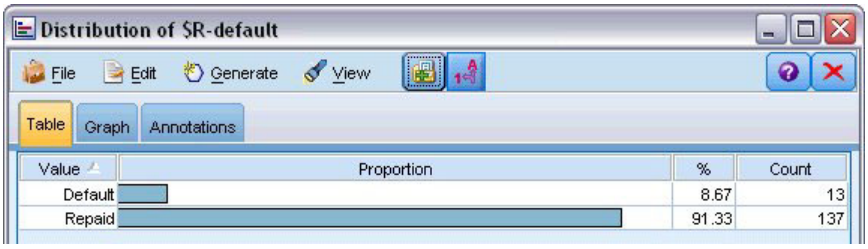


Figure 8. Sortie du noeud Distribution sans analyses d'entités

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 supposerons qu'un référentiel d'entités a déjà été créé. Pour plus d'informations, voir la rubrique «Création du référentiel», à la page 14.

Ajout des données source au référentiel

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

Pour pouvoir exporter les données, vous devez mapper les champs de la source de données aux fonctions du référentiel d'entités. 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'entités propose un ensemble de types d'informations standard (appelé fonctions) pour éviter la duplication.

Dans le noeud Export EA, configurez les informations de connexion du référentiel, 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 une 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 ce qui suit.

Tableau 12. Champs mappés aux fonctions du référentiel.

Champ	**Mappé à la fonction**
fname	NAME.GIVEN_NAME
mname	NAME.MIDDLE_NAME
lname	NAME.SUR_NAME
generation	NAME.NAME_GEN
dob	DOB.DOB
gender	GENDER.GENDER
addr1	ADDRESS.ADDR1
city	ADDRESS.CITY
country	ADDRESS.COUNTRY
postcode	ADDRESS.POSTAL_CODE
phone	PHONE.PHONE_NUM
email	EMAIL_ADDR.ADDR
ssn	SSN.ID_NUM
drlic	DRLIC.ID_NUM
passport	PASSPORT.ID_NUM

Cliquez sur **Exécuter** pour exporter les données dans le référentiel. Cela prend un certain temps. Quand la boîte de dialogue Commentaires d'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 noeud Export EA ; ce champ permet d'identifier les enregistrements de source de données de manière unique.)

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

Ouvrez le noeud source Entity Analytics(EA) et définissez les informations relatives au référentiel d'entités. 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).

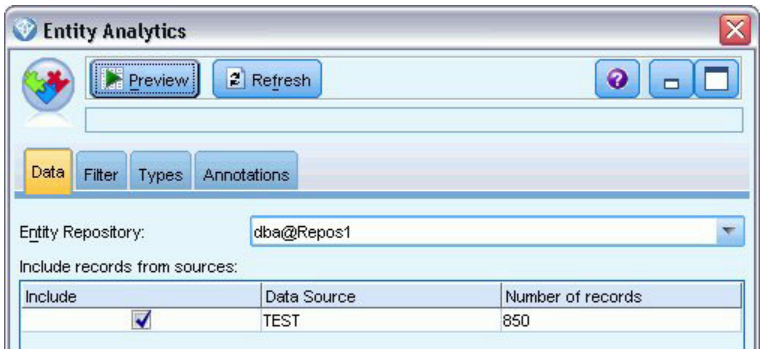


Figure 9. Sélection d'une source de données dans le référentiel

Cochez la case associée à la source de données **TEST** et cliquez sur OK.

Examinons ce que le système Entity Analytics a eu comme effet sur les données. Joignez un noeud Table au noeud source Entity Analytics(EA), ouvrez le noeud Table et cliquez sur **Exécuter** pour afficher la fenêtre de sortie du noeud Table.

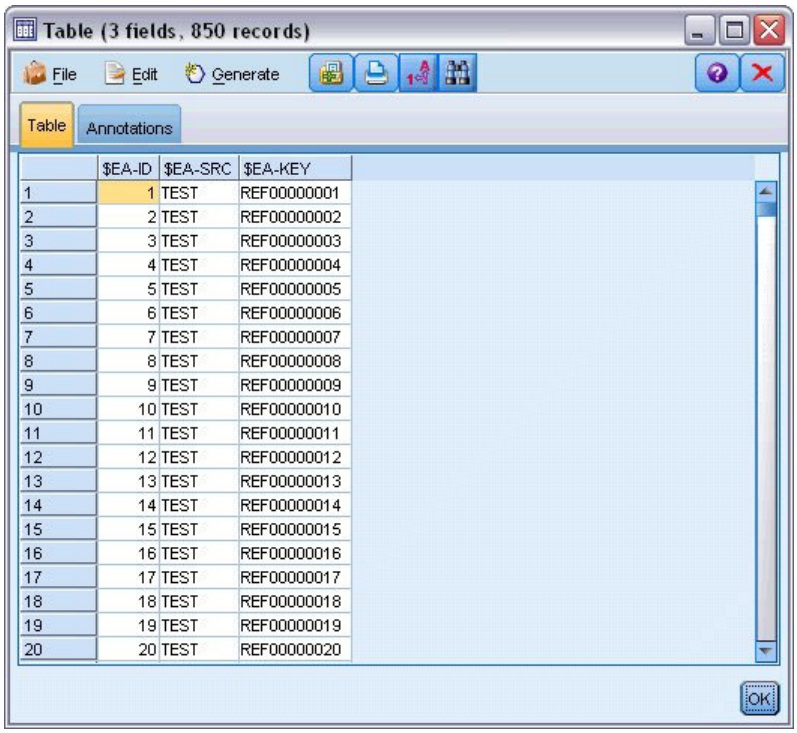


Figure 10. Sortie du noeud Table

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 noeud Export EA.

Cependant, le système a ajouté deux autres champs. Le champ *$EA-ID* est l'identifiant unique, non pas des enregistrements source, mais des identités résolues. Nous étudierons la différence dans un instant. Le

champ *$EA-SRC* identifie l'endroit d'où viennent les données. Ici, il s'appelle **TEST** parce qu'il s'agit de la balise source que vous lui avez attribué dans le noeud Export EA.

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

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

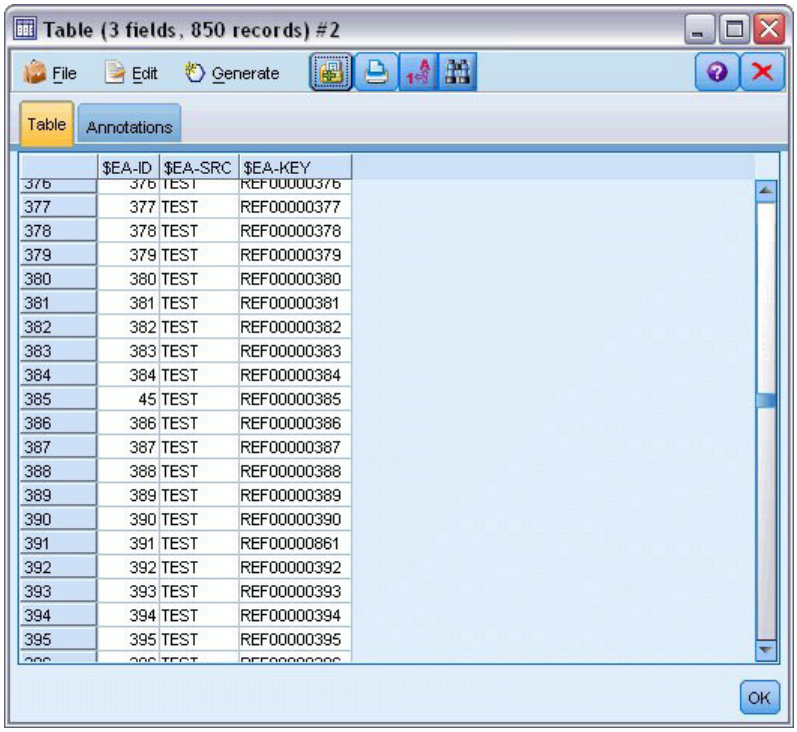


Figure 11. Différences entre les lignes de sortie Table et les nombres $EA-ID

Notez la façon dont le nombre *$EA-ID* semble hors séquence ici. Le système Entity Analytics a déterminé que l'enregistrement REF00000385 faisait référence à la personne identifiée comme l'entité 45 qui a également l'enregistrement REF00000045. Si vous faites défiler jusqu'à la sortie, vous trouverez d'autres nombres hors séquence, par exemple les lignes 485, 517, 520, etc. Examinons cela de plus près.

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

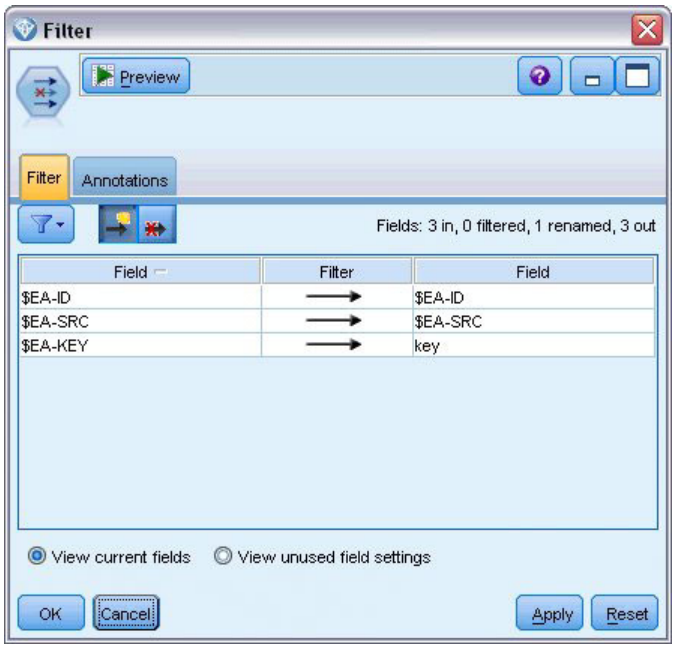


Figure 12. Nouveau nom du champ $EA-KEY

Cliquez sur **OK** pour fermer le noeud Filtre.

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

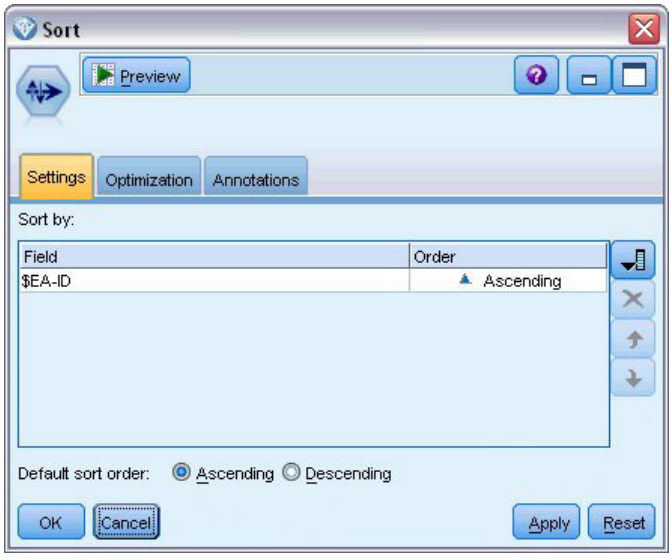


Figure 13. Tri des ID d'entités dans l'ordre croissant

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. Joignez un noeud Dériver au noeud Tri. Ouvrez le noeud Dériver et définissez le nom **Dériver champ** sur `IsDuplicate`. Dans la liste **Dériver en tant que**, choisissez **Indicateur**, ce qui définit également le **Type de champ** sur **Indicateur**. Définissez le champ **Valeur vraie** sur **Doublon** 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))
```

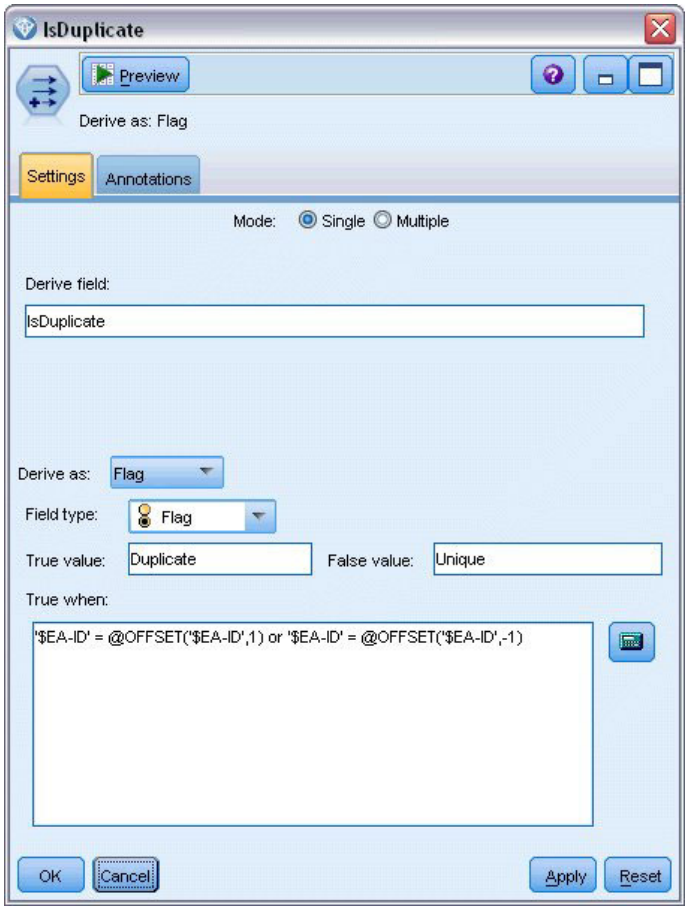
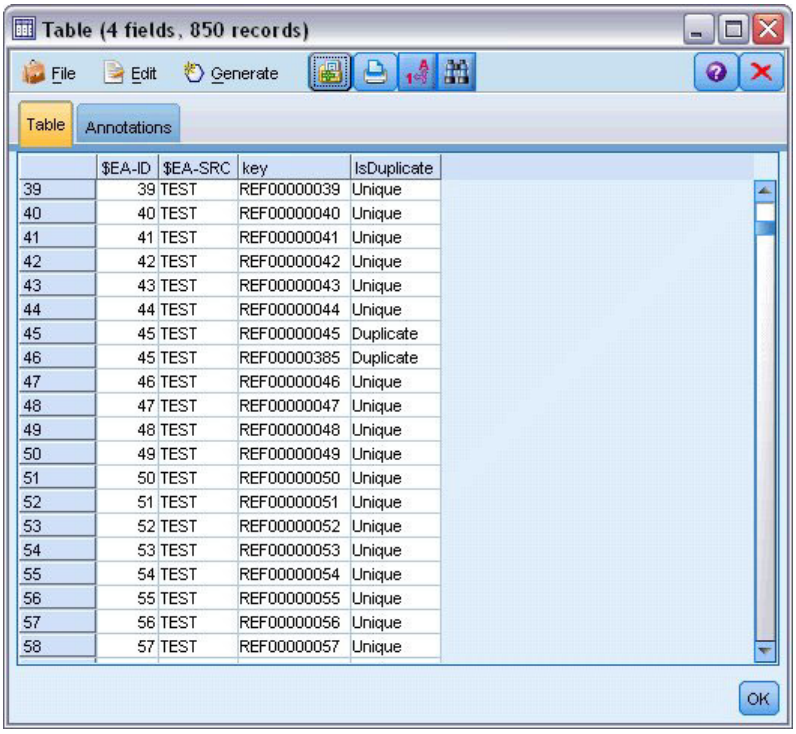


Figure 14. Définition de la condition dans le noeud Dériver

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 *IsDuplicate* est définie sur *Doublon*, sinon, elle est définie sur *Unique*.

Cliquez sur **OK** pour fermer le noeud.

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



	$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

Figure 15. Sortie du noeud Dériver

Souvenez-vous que nous avons examiné la sortie directement depuis le noeud source Entity Analytics(EA). Le système a déjà identifié que l'enregistrement REF00000385 faisait référence au même individu que l'entité 45. Nous avons affiné cette opération et avons signalé le fait que les enregistrements REF00000045 et REF00000385 étaient 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, joignez un noeud Rapport (dans l'onglet Sortie de la palette des noeuds) au noeud Dériver *IsDuplicate*. 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.

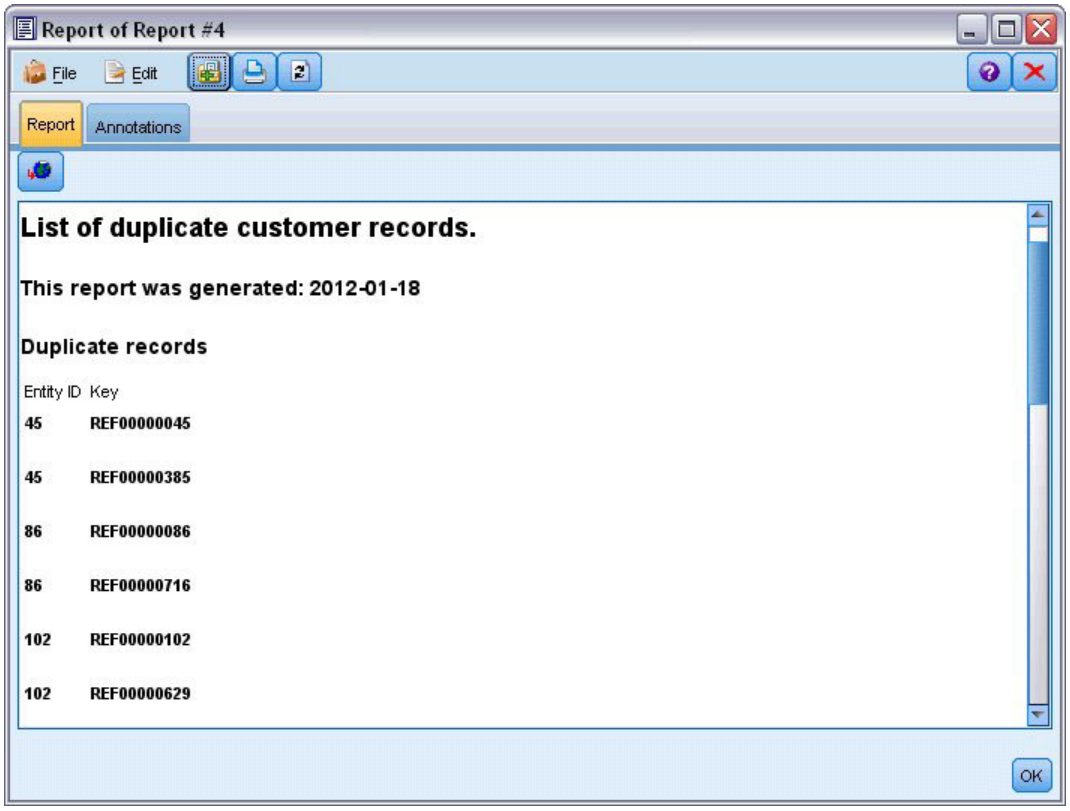


Figure 16. Sortie du noeud Rapport

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 consiste à 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 Fusion pour fusionner la sortie de ce modèle avec des informations sur les enregistrements en double provenant des analyses d'entités pour savoir si cela modifie les prévisions.

D'abord, vous devez vérifier que les nouveaux champs ajoutés par le système Entity Analytics ont des types de données corrects ou des *niveaux de mesure* comme ils sont appelés dans SPSS Modeler. Joignez un noeud Type au noeud Dériver **IsDuplicate**, ouvrez le noeud Type et cliquez sur le bouton **Lire les valeurs**.

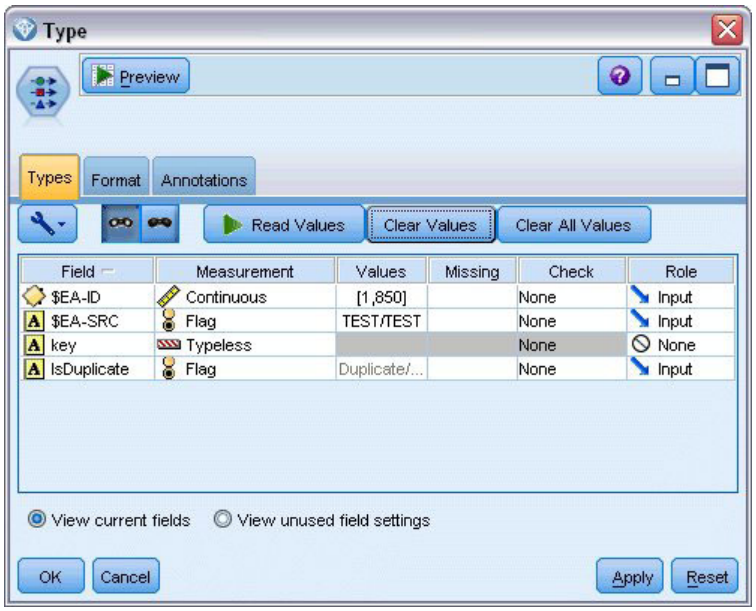


Figure 17. Paramètres du noeud Type

Vous pouvez maintenant ajouter le noeud Fusion. Joignez-le au noeud Type 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 Fusion qui devrait désormais contenir deux flèches d'entrée.

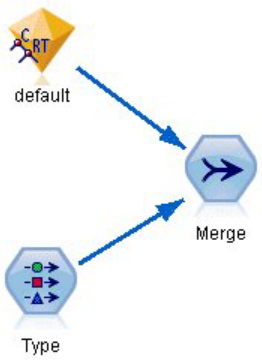


Figure 18. Entrées du noeud Fusion

Ouvrez le noeud Fusion, définissez la **Méthode de fusion** sur **Clés** et cliquez sur la flèche de droite pour déplacer le champ **clé** de **Clés possibles** vers **Clés pour fusion** puis cliquez sur **OK**.

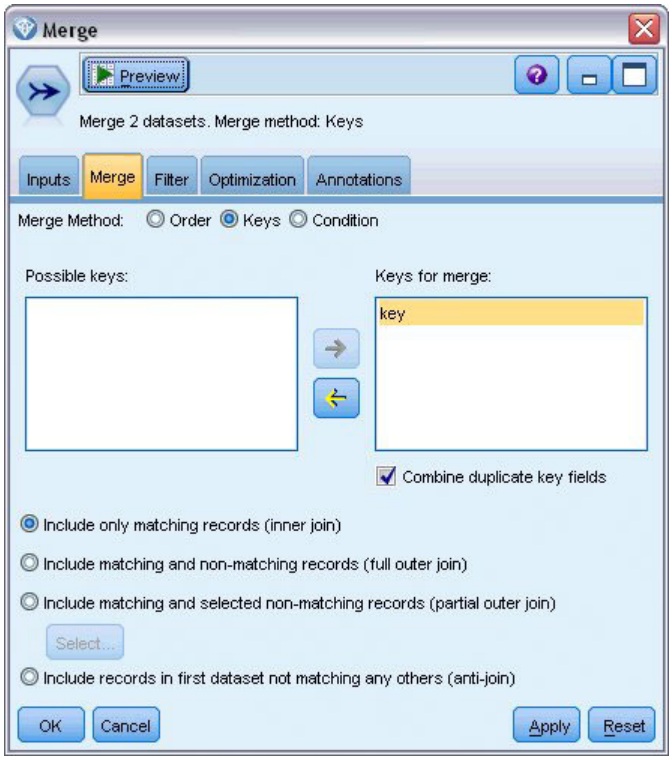


Figure 19. Spécification du champ-clé pour la fusion

Vous êtes presque prêt à effectuer la comparaison. Cependant, si vous deviez joindre un noeud Distribution 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 Entity Analytics, le champ de prévision lui-même (*$R-default*) du modèle de données n'a pas été mis à jour en fonction des nouvelles informations.

Pour ce faire, utilisez un noeud Remplissage qui peut remplacer les valeurs de champ. Joignez un noeud Remplissage au noeud Fusion et ouvrez le noeud Remplissage.

Cliquez sur le bouton en haut à droite de **Renseigner 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 modifiées si la condition spécifiée dans le reste de la boîte de dialogue est remplie.

Pour spécifier la condition, vérifiez que **Remplacer** est défini sur **En fonction de la condition**, puis dans le champ **Condition**, saisissez :

```
default /="default" and IsDuplicate  = "Duplicate"
```

Dans le champ **Remplacer par**, saisissez :

```
"default"
```

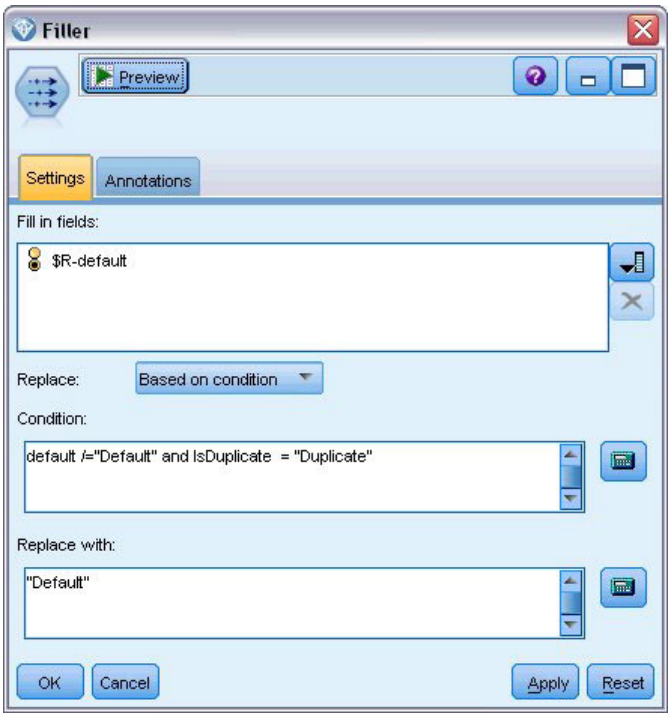


Figure 20. Spécification de la condition de remplacement des valeurs de champs

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

Le champ *$R-default* est le champ du modèle qui contient la prévision concernant la probabilité pour qu'un client ne rembourse pas 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 Remplissage.

Vous êtes maintenant prêt à voir la différence générée par les analyses d'entités. Dans la palette Graphiques, joignez un noeud Distribution au noeud Remplissage et ouvrez le noeud Distribution. Cliquez sur la liste **Champ** et choisissez **$R-default**.

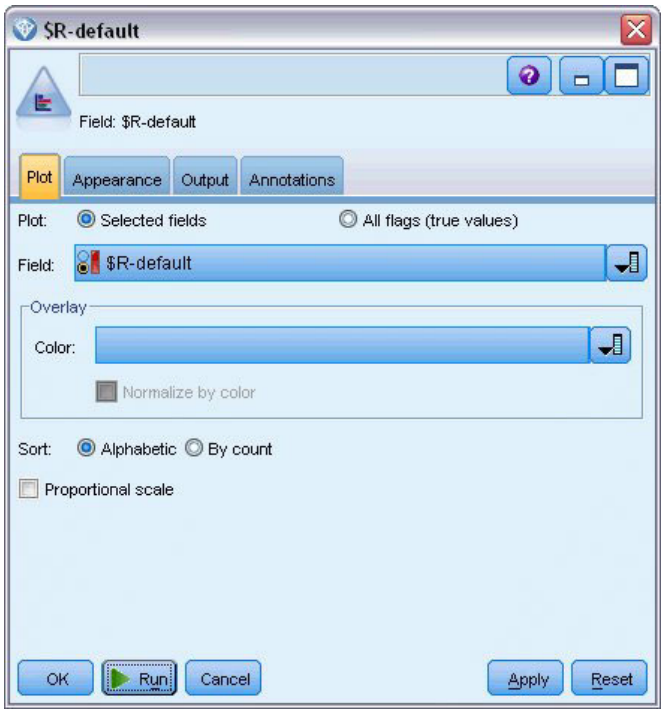


Figure 21. Paramètres du noeud Distribution

Cliquez sur **Exécuter** pour générer le graphique correspondant à la nouvelle prévision.

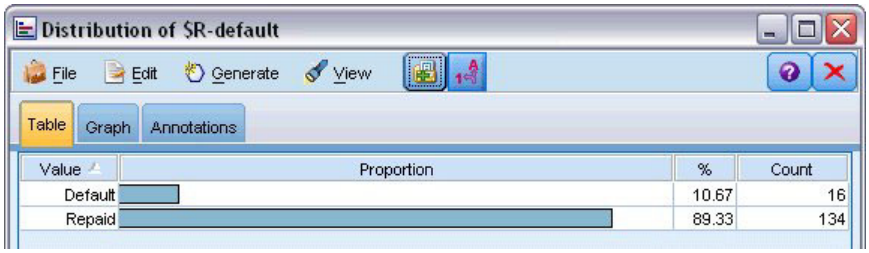


Figure 22. Sortie du noeud Distribution après les analyses d'entités

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 devez éliminer les enregistrements en double avant tout traitement. Vous pourriez poursuivre cette opération en utilisant un noeud Préparation automatique des données (ADP) pour analyser vos données, identifier des corrections, filtrer des champs qui sont problématiques ou peu susceptibles d'être utiles, dériver de nouveaux attributs le cas échéant et améliorer les performances 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.

Annexe. Propriétés de génération de scripts d'IBM SPSS Modeler Entity Analytics

Génération de scripts avec IBM SPSS Modeler Entity Analytics

La génération de scripts dans IBM SPSS Modeler Entity Analytics constitue un outil performant pour automatiser les processus dans l'interface utilisateur. Les scripts permettent d'effectuer les mêmes opérations qu'avec la souris ou le clavier. Vous pouvez les utiliser pour automatiser les tâches dont l'exécution manuelle s'avère très répétitive et très longue. Pour une explication sur la génération de scripts, consultez le guide *ScriptingAutomation.pdf* disponible avec IBM SPSS Modeler.

Propriétés communes

Les propriétés communes aux noeuds IBM SPSS Modeler Entity Analytics sont présentées dans le tableau ci-après. Vous trouverez des informations sur les noeuds spécifiques dans les sections qui suivent.

Tableau 13. Propriétés communes

Nom de la propriété	Type de données	Description de la propriété
`entity_repository`	[*'field','field', ... ,'field'*]	Chaîne de connexion au référentiel. Format : [*'nomréférentiel', 'nomutil', 'motdepasse'*] Exemple : `entity_repository = [ 'repos1', 'dba', 'psw1' ]`
`entity_type`	*string*	Type d'entité (ensemble de fonctions) à utiliser. Exemple : `entity_type = 'PERSON'`

Propriétés entityanalytics_exportnode

Le noeud Export EA est un noeud terminal qui lit les données d'entités d'une source de données et les exporte vers un référentiel afin de résoudre les entités.

Tableau 14. Propriétés entityanalytics_exportnode

Propriétés `entityanalytics_exportnode`	Type de données	Description de la propriété
`mode`	`Add` `PurgeFirst`	Mode Export. `Add` ajoute des enregistrements de fichier source au contenu existant du référentiel ; `PurgeFirst` supprime le contenu existant avant l'exportation.
`source_tag`	*chaîne*	Identifiant de source de données. Exemple : `source_tag = 'CUST'`

Tableau 14. Propriétés entityanalytics_exportnode (suite)

Propriétés entityanalytics_exportnode	Type de données	Description de la propriété
unique_key_field	*chaîne*	Champ d'entrée à utiliser pour les identifiants uniques des enregistrements de données. Exemple : unique_key_field = 'ID'
field_mapping	[[*'field_name'* *'feature.element'* *'usage_type'*]...]	Mappe les champs d'entrée à la fonction correspondante dans le référentiel. Exemple : field_mapping = [['fname' 'NAME.GIVEN_NAME' ''] ['addr1' 'ADDRESS.ADDR1' 'PRIMARY']] *Remarque* : Pour affecter à *usage_type* la valeur équivalente à "(Auto)", utilisez '' comme dans le premier exemple ci-dessus.

Propriétés entityanalytics_sourcenode

Le noeud source Entity Analytics (EA) lit les entités résolues du référentiel et transmet ces données au flux pour un nouveau traitement tel que le formatage sous la forme de rapport.

Tableau 15. Propriétés entityanalytics_sourcenode

Propriétés entityanalytics_sourcenode	Type de données	Description de la propriété
source_tags	*liste*	Liste de balises pour les sources de données à extraire du référentiel. Exemple : source_tags=['LOANS', 'CUSTOMERS']
relations	None Fermer All	Critère d'extraction des relations à partir du référentiel. None ne renvoie aucune relation. Close renvoie des correspondances en fonction de détails tels que le degré de séparation. All renvoie toutes les relations possibles.
max_degree_separation	*integer*	Minimum 0, maximum 3.
output_entity_type	*chaîne*	Liste des types d'entité utilisés dans le référentiel.

Propriétés entityanalytics_processnode

Le noeud Flux EA compare les nouvelles observations aux données d'entités du référentiel.

Tableau 16. Propriétés entityanalytics_processnode

Propriétés **entityanalytics_processnode**	Type de données	Description de la propriété
match	Exact ByIdentifier All	Critère d'extraction des entités à partir du référentiel. Exact renvoie uniquement les correspondances exactes. ByIdentifier renvoie les correspondances exactes et les entités partageant les mêmes identifiants. All renvoie toutes les correspondances possibles.
save_search_records	*boolean*	
relations	None Close All	Critère d'extraction des relations à partir du référentiel. None ne renvoie aucune relation. Close renvoie des correspondances en fonction de détails tels que le degré de séparation. All renvoie toutes les relations possibles.
max_degree_separation	*integer*	Minimum 0, maximum 3.
output_entity_type	*chaîne*	Liste des types d'entité utilisés dans le référentiel.

Remarques

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

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, programme ou service IBM n'implique pas que seul ce produit, programme ou service IBM puisse être utilisé. Tout produit, programme ou service fonctionnellement équivalent peut être utilisé s'il n'enfreint aucun droit de propriété intellectuelle d'IBM. Cependant l'utilisateur doit évaluer et vérifier l'utilisation d'un produit, programme ou service non IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits mentionnés dans le présent 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
U.S.A.

Pour le Canada, veuillez adresser votre courrier à :

IBM Director of Commercial Relations
IBM Canada Ltd
3600 Steeles Avenue East
Markham, Ontario
L3R 9Z7
Canada

Pour toute demande au sujet des licences concernant les jeux de caractères codés sur deux octets (DBCS), contactez le service Propriété intellectuelle IBM de votre pays ou adressez vos questions 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 ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales. LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. 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. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et logiciels décrits dans ce document.

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. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation à votre égard, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci), et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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

Le programme sous licence décrit dans le présent document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions du Livret Contractuel IBM, des Conditions internationales d'utilisation des Logiciels IBM ou de tout autre contrat équivalent.

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 ni leur compatibilité. Aucune réclamation relative à des produits non IBM ne pourra être reçue par 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

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines dans de nombreux pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à l'adresse 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 Etats-Unis et dans d'autres pays.

Linux est une marque déposée de Linus Torvalds aux Etats-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.

Les marques commerciales Java et basées sur Java ainsi que les logos sont des marques commerciales ou déposées de Oracle et/ou de ses filiales.

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

Index