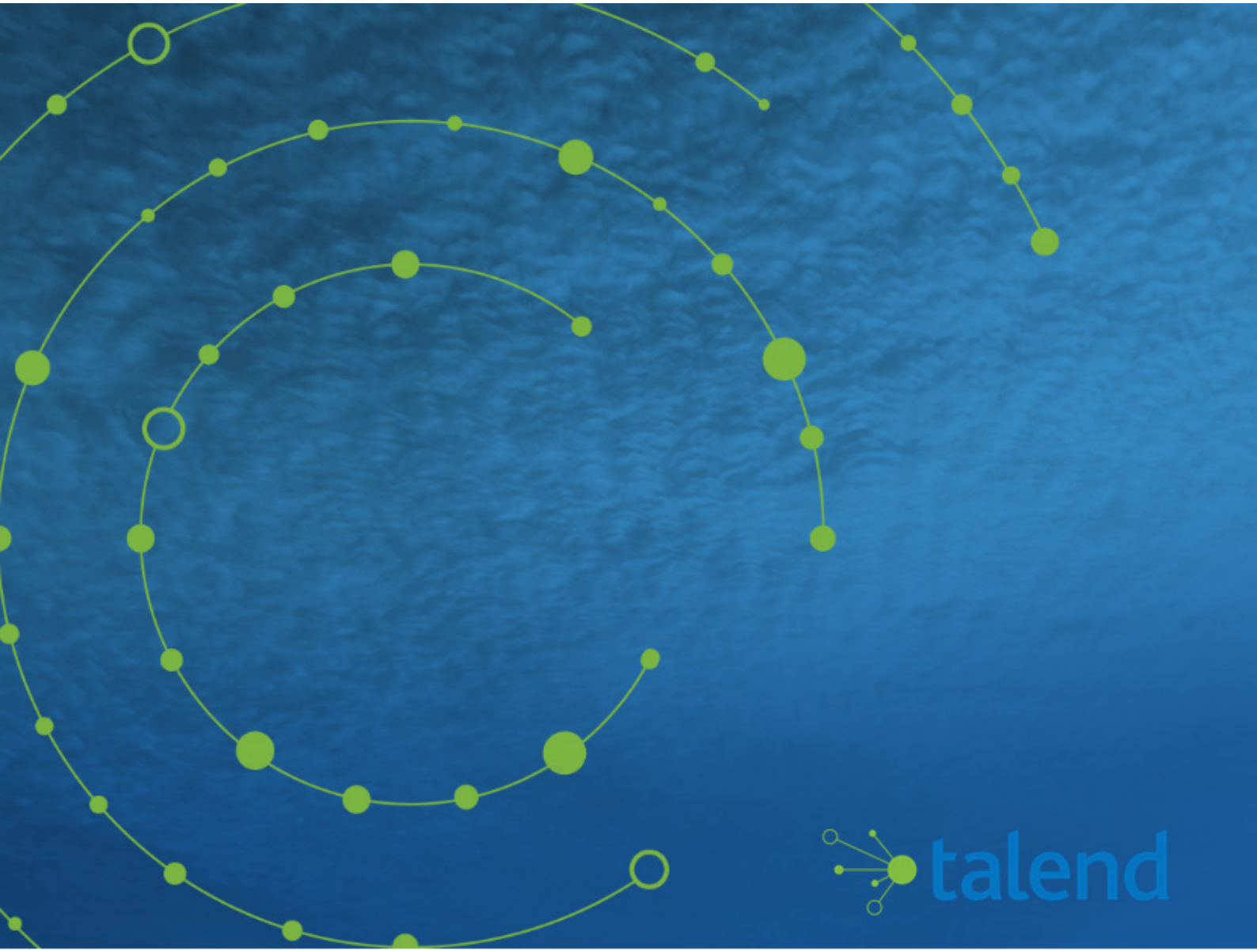


Guide de prise en main de Talend Open Studio for Data Quality

7.0.1

Table des matières

Copyleft

Convient à la version 7.0.1. Annule et remplace toute version antérieure de ce guide.

Date de publication : 13 avril 2018

Introduction à Talend Open Studio for Data Quality

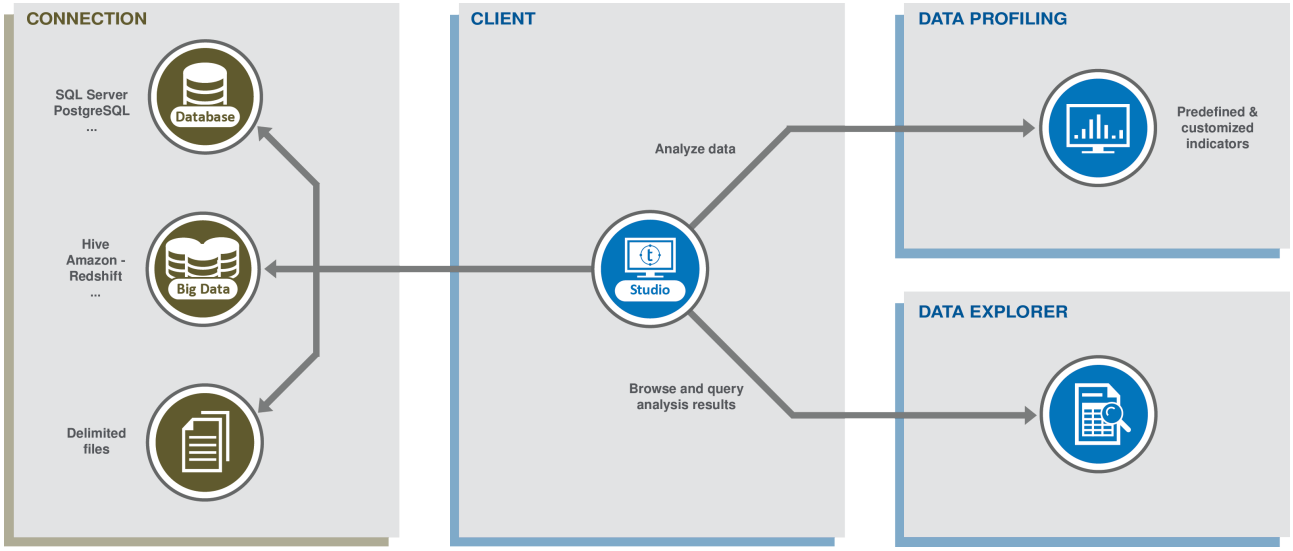
Talend fournit des outils de développement et de gestion unifiés pour intégrer et traiter toutes vos données dans un environnement graphique simple à utiliser.

Dans le Talend Open Studio for Data Quality, les utilisateurs peuvent accéder aux et examiner les données disponibles dans différentes sources de données et collecter des statistiques et des informations concernant ces données.

Architecture fonctionnelle de Talend Open Studio for Data Quality

L'architecture fonctionnelle de Talend Open Studio for Data Quality est un modèle architectural qui identifie les fonctions, les interactions et les besoins informatiques correspondants de Talend Open Studio for Data Quality. L'architecture d'ensemble a été décrite en isolant les fonctionnalités spécifiques en blocs fonctionnels.

Le graphique suivant illustre les blocs fonctionnels de l'architecture principale.



Plusieurs de ces blocs fonctionnels sont définis :

- Une perspective **Profiling** dans laquelle vous pouvez utiliser des modèles et des indicateurs prédéfinis ou personnalisés pour analyser les données stockées dans différentes sources de données.
- Une perspective **Data Explorer** dans laquelle vous pouvez explorer et interroger les résultats des analyses de profiling effectuées sur les données.

Prérequis à l'utilisation de Talend Open Studio for Data Quality

Ce chapitre vous fournit des informations simples concernant le logiciel et le matériel requis et recommandés pour prendre en main votre Talend Open Studio for Data Quality.

Il vous guide également à travers les étapes d'installation et de configuration des outils tiers requis et recommandés :

Recommandations relatives à la mémoire

Pour optimiser l'utilisation des produits Talend, référez-vous aux recommandations de mémoire et espace disque ci-dessous :

Utilisation de la mémoire	3 Go minimum, 4 Go recommandés
Utilisation du disque	3 Go

Recommandations logicielles

Pour optimiser l'utilisation des produits Talend, référez-vous aux recommandations système et logicielles ci-dessous :

Logiciels requis

- Systèmes d'exploitation pour le Studio Talend :

Type de support	**Système d'exploitation (64 bits seulement)**
Recommandé	Ubuntu 16.04 LTS
Recommandé	Microsoft Windows 10
Supporté	Apple macOS 10.13/High Sierra
	Apple macOS 10.12/Sierra
	Apple macOS 10.13/High Sierra
	Apple OS X 10.11/El Capitan

- Java 8 JRE Oracle. Consultez Installation de Java à la page 6.
- Une base de données MySQL installée et configurée, avec une base de données `gettingstarted`.

Logiciel facultatif

- 7-Zip. Consultez Installation de 7-Zip (Windows) à la page 7.

Installation de Java

Pour utiliser votre produit Talend, vous avez besoin d'une JRE Oracle (Oracle Java Runtime Environment) installée sur votre ordinateur.

Procédure

1. Dans la page Java SE Downloads (en anglais), sous **Java Platform, Standard Edition**, cliquez sur **JRE Download**.
2. Dans la page **Java SE Runtime Environment 8 Downloads**, sélectionnez le bouton radio **Accept License Agreement**.
3. Sélectionnez le téléchargement correspondant à votre système d'exploitation.
4. Suivez les étapes d'installation de Java proposées par l'assistant Oracle.

Résultats

Lorsque Java est installé sur votre ordinateur, vous devez configurer la variable d'environnement `JAVA_HOME`. Pour plus d'informations, consultez :

- Configuration des variables d'environnement Java sous Windows à la page 6.
- Configuration des variables d'environnement Java sous Linux à la page 7.

Configuration des variables d'environnement Java sous Windows

Avant d'installer votre produit Talend, vous devez configurer les variables d'environnement JAVA_HOME et Path :

Procédure

1. Dans le menu **Démarrer** de votre ordinateur, cliquez-droit sur **Ordinateur** et sélectionnez **Properties**.
2. Dans la fenêtre **Control Panel Home**, cliquez sur **Advanced system settings**.
3. Dans la fenêtre **System Properties**, cliquez sur **Environment Variables...**.
4. Sous **System Variables**, cliquez sur **New...** pour créer une variable. Nommez la variable `JAVA_HOME`, saisissez le chemin d'accès à votre JRE 8 Java, puis cliquez sur **OK**.

   Exemple de chemin vers la JRE par défaut : `C:\Program Files\Java\jre1.8.0_77`.
5. Sous **System Variables**, sélectionnez la variable **Path** et cliquez sur **Edit...** pour ajouter la variable `JAVA_HOME` précédemment définie à la fin de la variable d'environnement `Path`, en les séparant par un point-virgule.

   Exemple : `<PathVariable>;%JAVA_HOME%\bin`.

Configuration des variables d'environnement Java sous Linux

Avant d'installer votre produit Talend, vous devez configurer les variables d'environnement JAVA_HOME et Path.

Procédure

1. Trouvez le répertoire d'installation de la JRE.

   Exemple : `/usr/lib/jvm/jre1.8.0_65`
2. Spécifiez-le dans la variable d'environnement `JAVA_HOME`.

   Exemple :

   ```
   export JAVA_HOME=/usr/lib/jvm/jre1.8.0_65
   export PATH=$JAVA_HOME/bin:$PATH
   ```
3. Ajoutez ces lignes à la fin des profils utilisateurs dans le fichier `~/.profile` ou, en tant que super-utilisateur, à la fin des profils globaux dans le fichier `/etc/profile`.
4. Connectez-vous à nouveau.

Installation de 7-Zip (Windows)

Talend recommande d'installer 7-Zip et de l'utiliser pour extraire les fichiers d'installation : http://www.spiroo.be/7zip/.

Procédure

1. Téléchargez l'installeur de 7-Zip correspondant à votre système d'exploitation.
2. Naviguez dans vos dossiers locaux, trouvez le fichier .exe de 7-Zip et double-cliquez dessus pour l'installer.

Résultats

Le téléchargement démarre automatiquement.

Téléchargement et installation de Talend Open Studio for Data Quality

Talend Open Studio for Data Quality est simple à installer. Après l'avoir téléchargé depuis le site Web de Talend, un simple dézippage permet de l'installer sur votre ordinateur.

Ce chapitre vous fournit les informations de base relatives au téléchargement et à l'installation.

Téléchargement de Talend Open Studio for Data Quality

Talend Open Studio for Data Quality est un produit open source libre que vous pouvez télécharger directement depuis le site Web de Talend.

Procédure

1. Allez à la page de téléchargement de Talend Open Studio for Data Quality.
2. Cliquez sur **TÉLÉCHARGER L'OUTIL LIBRE**.

Résultats

Le téléchargement démarre automatiquement.

Installation de Talend Open Studio for Data Quality

L'installation s'effectue en dézippant le fichier .zip précédemment téléchargé.

Vous pouvez faire ceci en utilisant :

- 7-Zip (recommandé sous Windows) : Extraire via 7-Zip (recommandé pour Windows) à la page 8.
- le dézippeur par défaut de Windows : Extraire via l'outil de dézippage Windows par défaut à la page 9.
- le dézippeur par défaut de Linux (pour un système d'exploitation basé Linux) : Extraire via l'outil de dézippage Windows par défaut à la page 9.

Extraire via 7-Zip (recommandé pour Windows)

Sous Windows, Talend vous recommande d'installer 7-Zip et de l'utiliser pour extraire des fichiers. Pour plus d'informations, consultez Installation de 7-Zip (Windows) à la page 7.

Pour installer le Studio, suivez les étapes suivantes :

Procédure

1. Naviguez dans vos dossiers locaux, trouvez le fichier .zip et déplacez-le à un autre emplacement, avec un chemin d'accès aussi court que possible et sans caractère d'espace.

   Exemple : `C:/Talend/`
2. Dézippez-le en cliquant-droit sur le fichier compressé et sélectionnez **7-Zip** > **Extract Here**.

Extraire via l'outil de dézippage Windows par défaut

Si vous ne souhaitez pas utiliser 7-Zip, vous pouvez utiliser l'outil de dézippage par défaut de Windows :

Procédure

1. Dézippez-le en cliquant-droit sur le fichier compressé et sélectionnez **Extract All**.
2. Cliquez sur **Browse** et naviguez jusqu'au disque C:.
3. Sélectionnez **Make new folder** et nommez le dossier `Talend`. Cliquez sur **OK**.
4. Cliquez sur **Extract** pour commencer l'installation.

Extraire via l'outil de dézippage Linux

Pour installer le Studio, suivez les étapes ci-dessous :

Procédure

1. Naviguez dans vos dossiers locaux, trouvez le fichier .zip et déplacez-le à un autre emplacement, avec un chemin d'accès aussi court que possible, sans caractère d'espace.

   Exemple : `home/user/talend/`
2. Dézippez-le en cliquant-droit sur le fichier compressé et sélectionnez **Extract Here**.

Configuration de votre produit Talend

Ce chapitre prend l'exemple d'une entreprise fournissant des services de locations de films et de streaming de vidéos. Il vous explique comment une telle entreprise peut tirer parti de Talend Open Studio for Data Quality.

Démarrage du Studio pour la première fois

Le répertoire d'installation du Studio contient des fichiers binaires pour différentes plateformes, notamment Mac OS X et Linux/Unix.

Pour ouvrir le Studio Talend pour la première fois, procédez comme suit :

Procédure

1. Double-cliquez sur le fichier exécutable correspondant à votre système d'exploitation, par exemple :
   - `TOS_*-win-x86_64.exe`, sous Windows.
   - `TOS_*-linux-gtk-x86_64`, sous Linux.
   - `TOS_*-macosx-cocoa.app`, sous Mac.
2. Dans la fenêtre **User License Agreement** qui s'ouvre, lisez et acceptez les termes de la licence pour procéder aux étapes suivantes.

Résultats

Le Studio Talend s'ouvre rapidement, puis l'assistant **Connect to TalendForge** s'affiche. Vous pouvez vous connecter à **TalendForge** pour bénéficier de la Communauté Talend ou cliquer sur le bouton **Skip this step** pour continuer.

Installation des packages supplémentaires

Talend vous recommande d'installer des packages supplémentaires, y compris des bibliothèques tierces et les pilotes de bases de données, dès que vous vous connectez à votre Studio Talend, afin de tirer pleinement parti de toutes les fonctionnalités du Studio.

Procédure

1. Lorsque l'assistant **Additional Talend Packages** s'ouvre, installez les packages supplémentaires, en cochant les cases **Required** et **Optional third-party libraries**. Cliquez sur **Finish**.

   Cet assistant s'affiche à chaque fois que vous lancez le studio si des packages supplémentaires sont disponibles à l'installation à moins que vous ne cochiez la case **Do not show this again**. Vous pouvez également afficher cet assistant en sélectionnant **Help** > **Install Additional Packages** dans la barre de menu.

   Pour plus d'informations, consultez la section concernant l'installation de packages supplémentaire s dans le Guide d'installation et de migration de Talend Open Studio for Data Quality.
2. Dans la fenêtre **Download external modules**, cliquez sur le bouton **Accept all** au bas de l'assistant pour accepter toutes les licences des modules externes dans le studio.

Attendez que toutes les bibliothèques soient installées avant de commencer à utiliser le studio.

3. Si nécessaire, redémarrez votre Studio Talend pour que certains packages supplémentaires soient pris en compte.

Profiling des données

e chapitre prend l'exemple d'une entreprise fournissant des services de locations de films et de streaming de vidéos. Il vous explique comment une telle entreprise peut tirer parti de Studio Talend.

Vous allez utiliser des données relatives à vos clients, tout en apprenant à valider les adresses e-mail des clients et standardiser les numéros de téléphone avant de les envoyer au logiciel de support client.

Configurer les données d'entrée

Dans ce document, l'exemple suppose que les données clients à profiler sont stockées dans une base de données MySQL.

Si vous souhaitez reproduire l'exemple et utiliser les données d'entrée exactes, téléchargez le fichier `gettingstarted.sql` des données clients et importez-le dans une base de données MySQL.

Avant de commencer

- Vous disposez d'un accès à une base de données MySQL.
- Vous avez téléchargé tos_dq_gettingstarted_source_files.zip depuis l'onglet **Downloads** de la version en ligne de cette page à l'adresse https://help.talend.com, et stocké le fichier source `gettingstarted.sql` en local.

Procédure

1. Ouvrez MySQL Workbench pour démarrer une instance de la base de données.
2. Dans la barre de menu, sélectionnez **Server** > **Data Import** pour ouvrir l'assistant d'import.
3. Sélectionnez l'option **Import from Self-Contained File** et naviguez jusqu'à l'emplacement où est stocké le fichier .sql `gettingstarted`.
4. Sélectionnez le schéma dans lequel vous souhaitez importer les données ou cliquez sur **New...** pour définir un nouveau schéma.
5. Cliquez sur **Start Import** en bas à droite.

Résultats

La base de données gettingstarted est importée dans la base de données MySQL.

Identifier les anomalies dans les données

Ce cas d'utilisation explique comment utiliser la perspective **Profiling** du studio pour analyser les adresses e-mail et les numéros de téléphone des clients. Il utilise des indicateurs et des modèles prêts à l'emploi sur les colonnes et montre les données d'adresse correspondantes et ne correspondant pas.

Vous pouvez utiliser la perspective **Data Explorer** pour parcourir les données ne correspondant pas.

La séquence de profiling des données clients comprend les étapes suivantes :

Procédure

1. Création d'une analyse de colonnes sur les adresses e-mail et les numéros de téléphone des clients. Pour plus d'informations, consultez Définir une analyse de colonnes à la page 13.
2. Connexion à la base de données comprenant les données clients dans l'éditeur d'analyse. Pour plus d'informations, consultez Créer une connexion à la base de données à la page 14.
3. Ajout d'indicateurs fournissant des statistiques simples sur les données comme le nombre de lignes, de valeurs blanches et de valeurs en doublon. Pour plus d'informations, consultez Configurer des indicateurs système à la page 16.
4. Ajout de modèles standard par rapport auxquels les adresses e-mail et les numéros de téléphone correspondent. Pour plus d'informations, consultez Configurer des modèles à la page 18.
5. Exécution de l'analyse afin que les résultats s'affichent dans les tables et les graphiques. Pour plus d'informations, consultez Afficher les résultats d'analyse à la page 19.
6. Accès à une vue des données analysées pour consulter les enregistrements invalides. Pour plus d'informations, consultez Explorer les données ne correspondant pas à la page 21.

Définir une analyse de colonnes

Vous voulez créer une analyse de colonnes dans la perspective **Profiling** du studio pour examiner les colonnes Email et Phone dans une base de données MySQL et collecter des statistiques sur celles-ci. L'analyse fonctionne sur plusieurs colonnes mais chaque colonne est analysée de manière séparée et indépendante.

Procédure

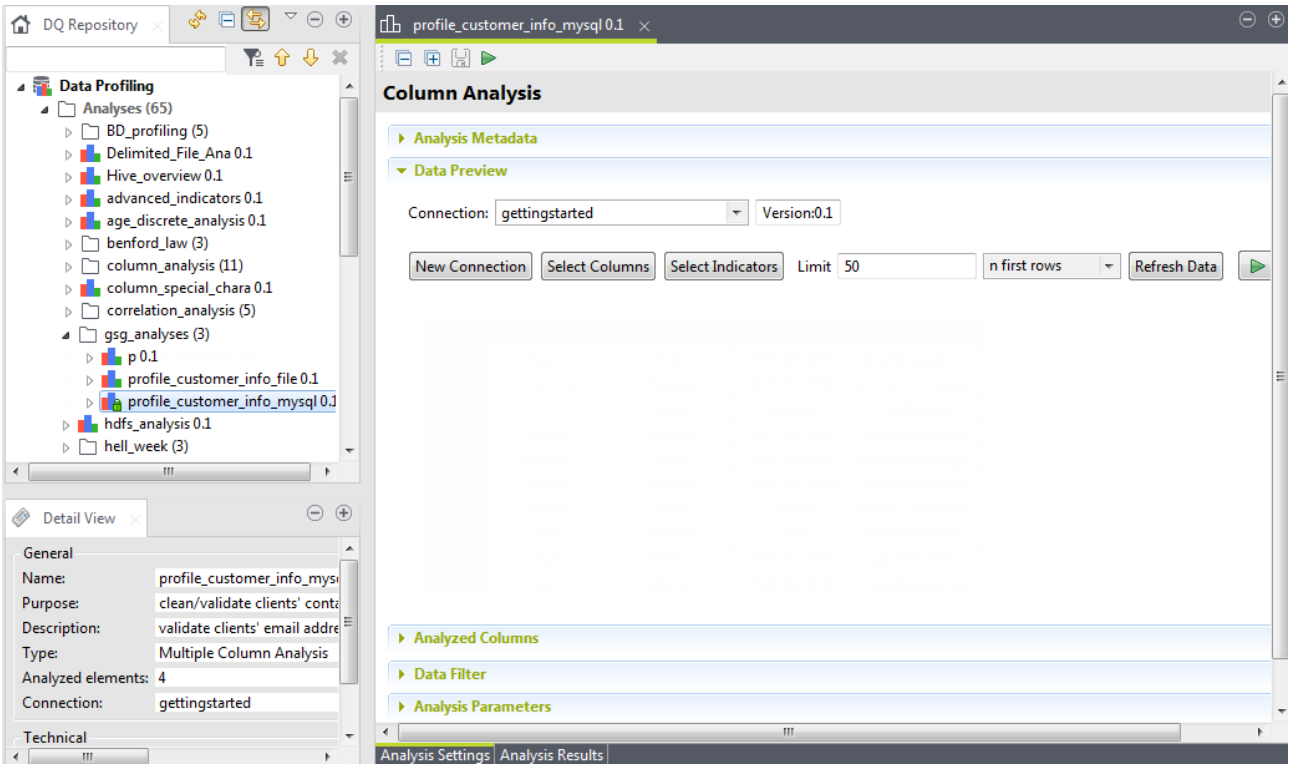
1. Dans l'arborescence **DQ Repository**, cliquez-droit sur **Analyses** et sélectionnez **New Analysis**.

   L'assistant **[Create New Analysis]** s'ouvre.
2. Commencez à saisir `Basic column analysis` dans le champ de recherche, sélectionnez **Basic Column Analysis** dans la liste et cliquez sur **Next**.
3. Dans le champ **Name**, nommez l'analyse.

   Le champ **Name** est obligatoire. N'utilisez aucun espace ou caractère spécial dans le nom de l'analyse.
4. Définissez un objectif et une description pour l'analyse et cliquez sur **Finish** pour ouvrir l'éditeur d'analyse.

   Les champs **Purpose** et **Description** sont facultatifs, mais il est conseillé de renseigner ces informations s'affichant dans **Detail View** lorsque vous sélectionnez l'analyse.

Résultats



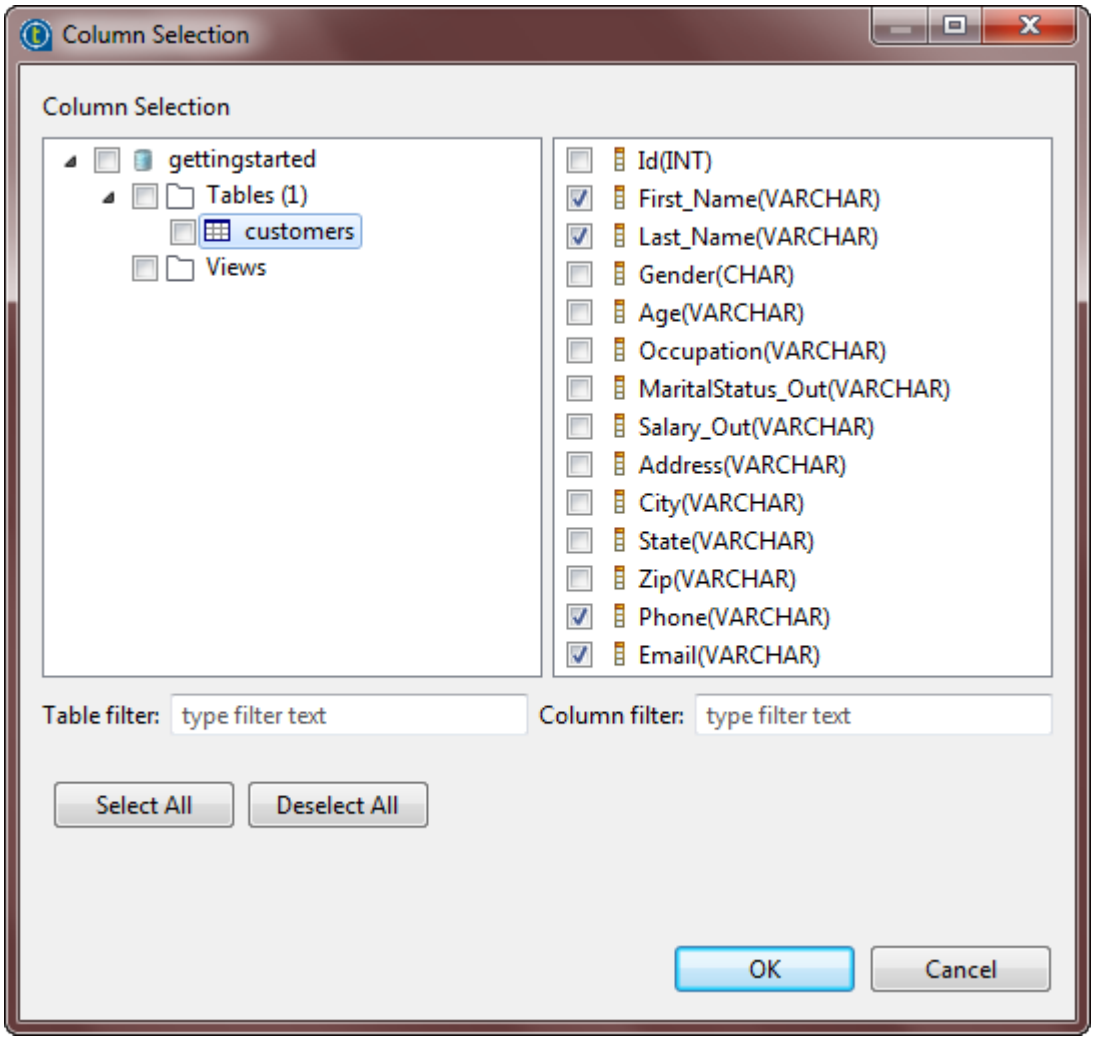
La nouvelle analyse est listée dans le dossier **Analysis** dans l'arborescence **DQ Repository**.

Créer une connexion à la base de données

Avant d'effectuer l'analyse des données clients, enregistrées dans la base de données MySQL dans cet exemple, vous devez d'abord créer la connexion à la base de données.

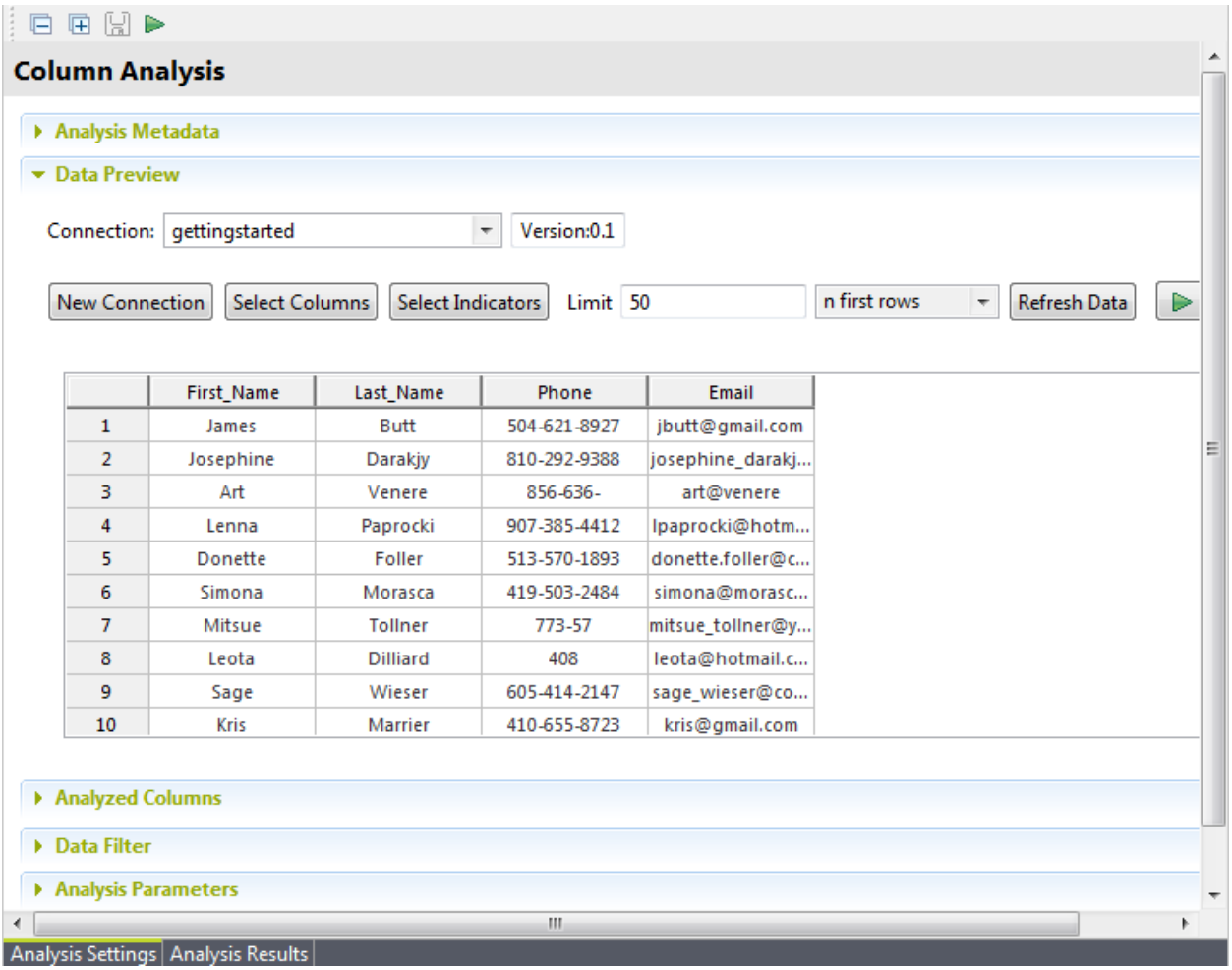
Procédure

1. Dans l'éditeur d'analyse, cliquez sur l'onglet **New Connection** pour ouvrir l'assistant **[Create New Connection]**.
2. Dans la liste **Connection Type**, sélectionnez **DB connections** et cliquez sur **Next**.
3. Clquez sur **Finish** pour créer la connexion à la base de données. Elle s'affiche sous le nœud **Metadata** et une nouvelle étape de l'assistant s'ouvre.
4. éveloppez la connexion à la base de données, cliquez sur le nom de la table et cochez les cases des colonnes sur lesquelles vous désirez créer l'analyse.



5. Cliquez sur **OK** pour fermer l'assistant et lister les colonnes dans l'éditeur d'analyse.

   Vous pouvez cliquer sur **Refresh Data** pour ouvrir les données actuelles dans l'éditeur d'analyse.



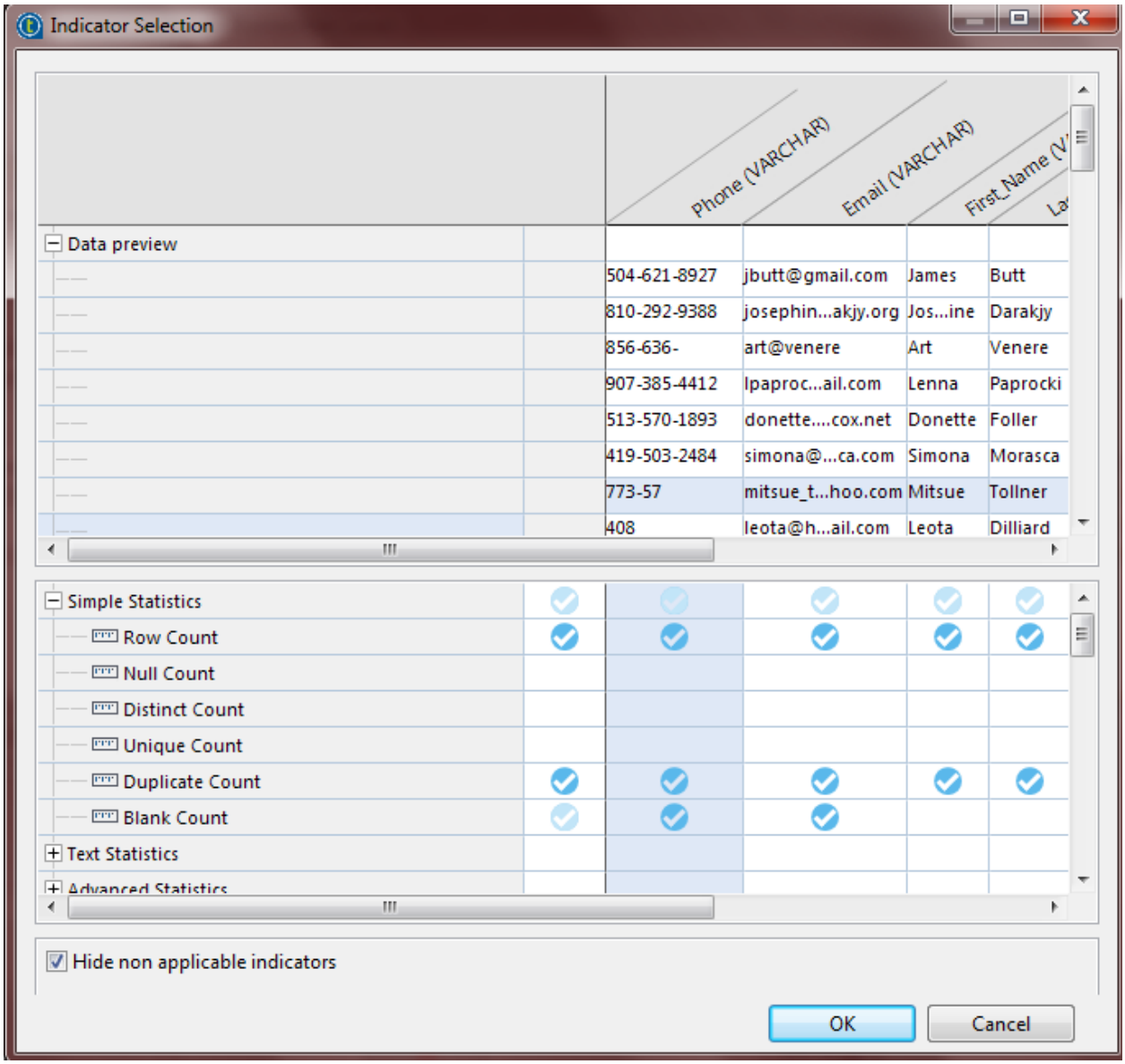
	First_Name	Last_Name	Phone	Email
1	James	Butt	504-621-8927	jbutt@gmail.com
2	Josephine	Darakjy	810-292-9388	josephine_darakj...
3	Art	Venere	856-636-	art@venere
4	Lenna	Paprocki	907-385-4412	lpaprocki@hotm...
5	Donette	Foller	513-570-1893	donette.foller@c...
6	Simona	Morasca	419-503-2484	simona@morasc...
7	Mitsue	Tollner	773-57	mitsue_tollner@y...
8	Leota	Dilliard	408	leota@hotmail.c...
9	Sage	Wieser	605-414-2147	sage_wieser@co...
10	Kris	Marrier	410-655-8723	kris@gmail.com

Configurer des indicateurs système

Cette analyse de colonnes utilise des indicateurs prêts à l'emploi pour fournir des statistiques simples comme le nombre de lignes, de valeurs blanches et de valeurs en doublon dans les colonnes Email et Phone.

Procédure

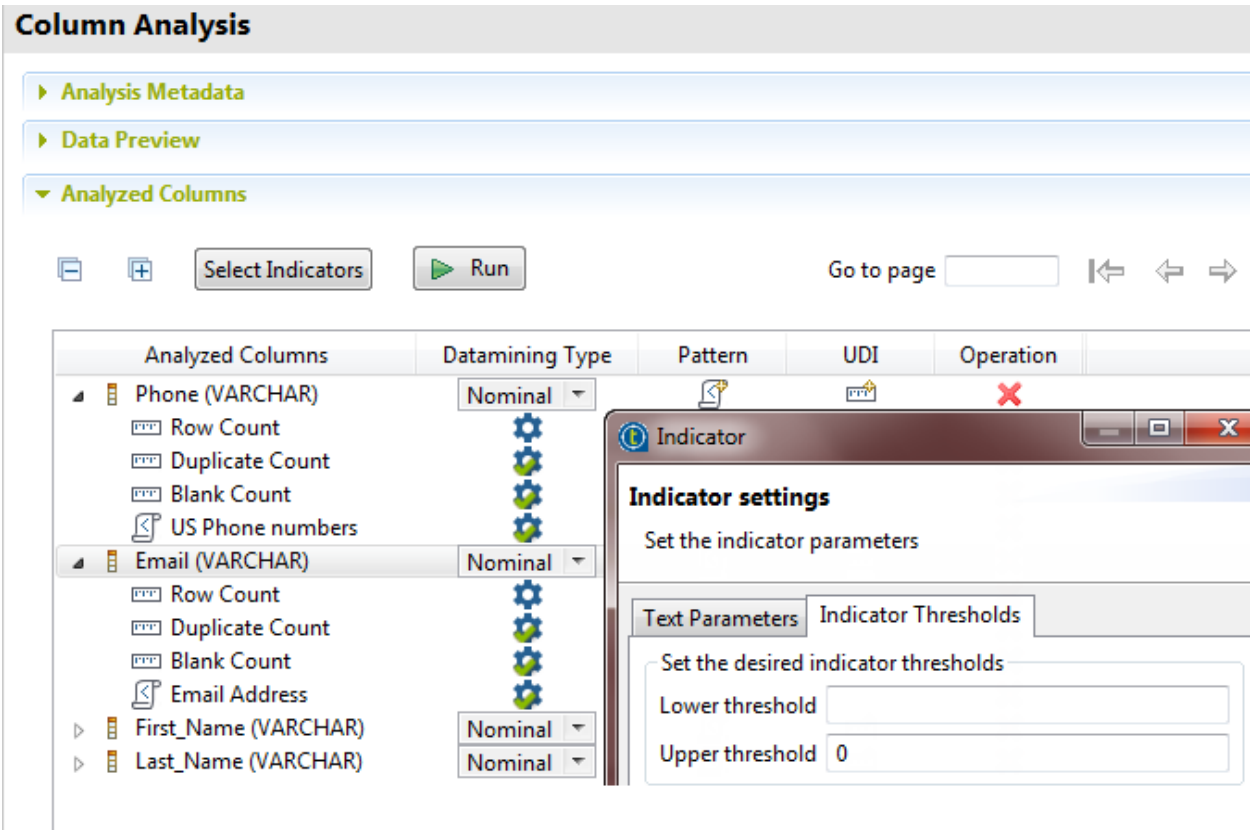
1. Dans la zone **Data Preview** dans l'éditeur d'analyse, cliquez sur **Select indicators** pour ouvrir la boîte de dialogue **[Indicator Selection]**.



2. Développez **Simple Statistics** sélectionnez les indicateurs **Row Count**, **Blank Count** et **Duplicate Count**. Cliquez sur **OK** pour fermer l'assistant.

   Vous voulez consulter le nombre de lignes, de valeurs blanches et de valeurs en doublon dans les colonnes Email et Phone pour contrôler la cohérence des données.

   Des indicateurs sont ainsi ajoutés dans les colonnes qui se trouvent dans la zone **Analyzed Columns**.



3. Cliquez sur l'icône à côté des indicateurs **Duplicate Count** et **Blank Count** et dans le champ **Upper threshold**, saisissez la valeur `0`.

   Définir des limites dans les colonnes `Email` and `Phone` est très utile car le nombre des valeurs blanches et des valeurs en doublon est indiqué en rouge dans les résultats d'analyse.

Configurer des modèles

Cette analyse de colonnes utilise des modèles prédéfinis pour que le contenu des colonnes Email et Phone corresponde aux modèles standard d'e-mails et de numéros de téléphone basés aux États-Unis, respectivement. Cette analyse définit le contenu, la structure et la qualité des adresses e-mail et des numéros de téléphone, et donne un pourcentage des données qui correspondent aux formats standard ainsi que des données qui ne correspondent pas.

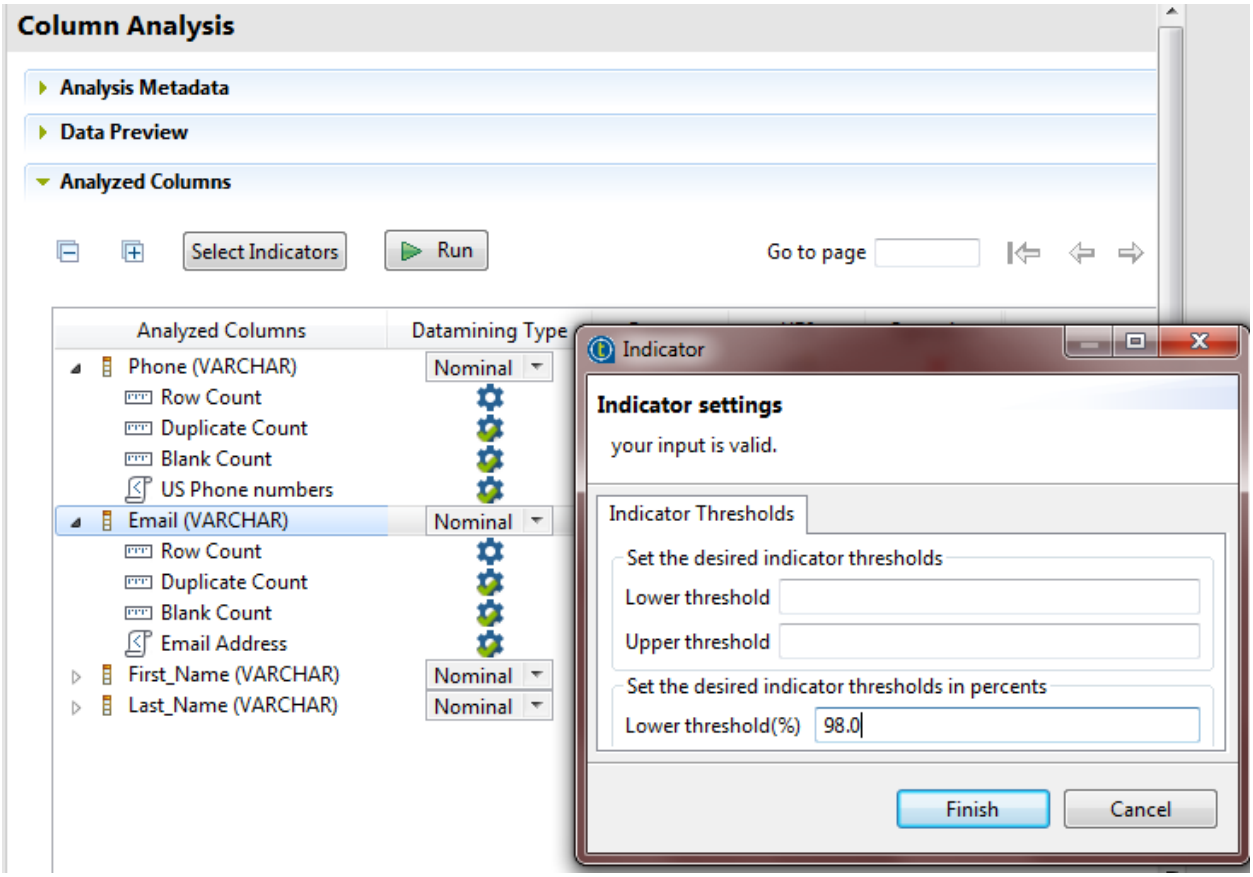
Procédure

1. Dans la zone **Data Preview** dans l'éditeur d'analyse, cliquez sur l'icône à côté de la colonne `Email` pour ouvrir la boîte de dialogue **[Pattern Selector]**.
2. Développez **Regex** > **internet**, cochez la case **Email Address** et cliquez sur **OK** pour fermer la boîte de dialogue.

   Le modèle est ajouté à la colonne dans la zone **Analyzed Columns**.
3. Cliquez sur l'icône à côté de la colonne `Phone` pour ouvrir la boîte de dialogue **[Pattern Selector]**.
4. Développez **Regex** > **phone**,cochez la case **US phone numbers** et cliquez sur **OK** pour fermer la boîte de dialogue

   Le modèle est ajouté à la colonne dans la zone **Analyzed Columns**.

5. Cliquez sur l'icône à côté des modèles **Email Address** et **US phone numbers**, et dans les champs **Lower threshold (%)**, saisissez `98.0`.



Si le nombre d'enregistrements correspondants aux modèles est inférieur à 98 %, il est indiqué en rouge dans les résultats d'analyse.

Afficher les résultats d'analyse

Une fois la création de l'analyse de colonnes et la définition des indicateurs et des modèles terminées, vous pouvez exécuter l'analyse et afficher ses résultats dans des tables et des graphiques.

Procédure

1. Dans **Analysis Parameters**, sélectionnez **java** dans la liste **Execution engine** pour exécuter l'analyse à l'aide du moteur Java.
2. Dans l'éditeur d'analyse, appuyez sur la touche **F6** pour exécuter l'analyse ou cliquez sur le bouton **Run**.

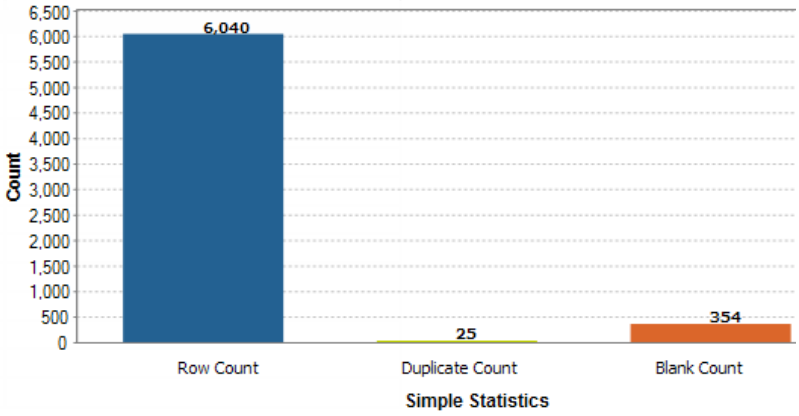
   L'éditeur passe à la vue **Analysis Results**. Les résultats d'analyse comportent des graphiques générés pour les colonnes analysées ainsi que des tables détaillant les résultats correspondants aux statistiques et aux modèles.

   Les résultats pour la colonne Email se présentent comme suit:

Column: customers.Email

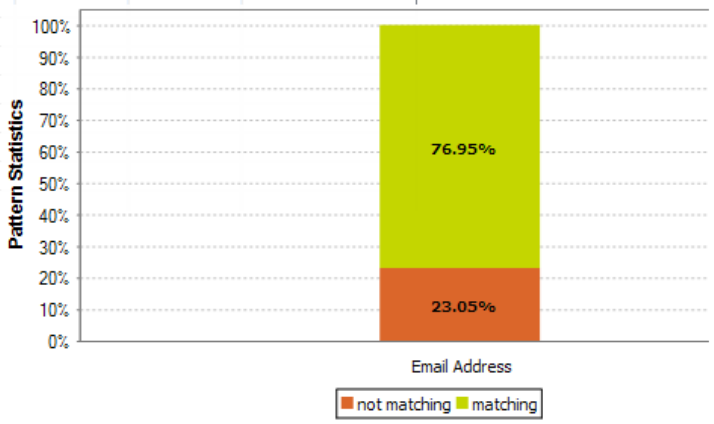
Simple Statistics

Label	Count	%
Row Count	6040.00	100.00%
Duplicate Count	25.00	0.41%
Blank Count	354.00	5.86%

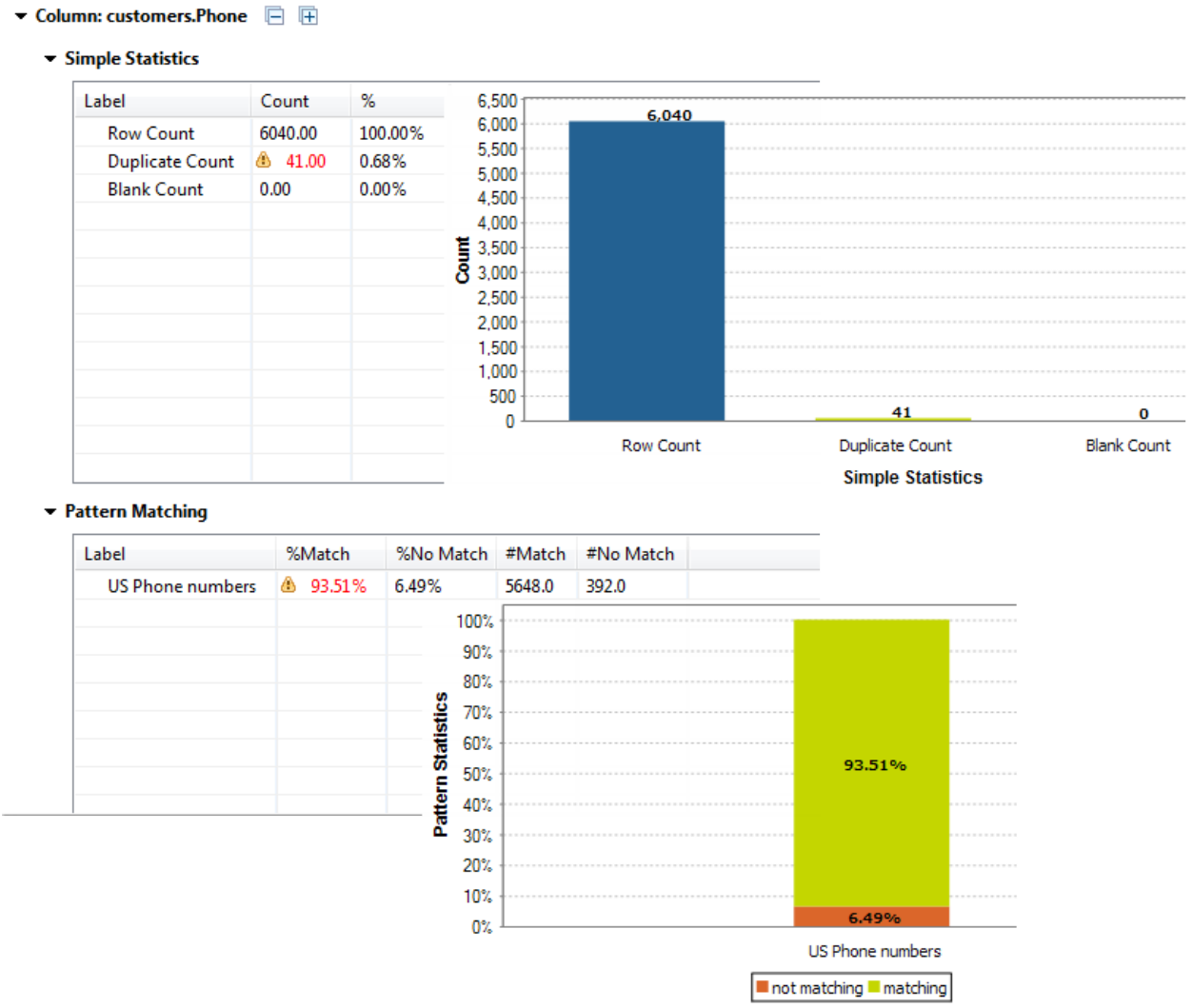


Pattern Matching

Label	%Match	%No Match	#Match	#No Match
Email Address	76.95%	23.05%	4648.0	1392.0



Les résultats pour la colonne Phone se présentent comme suit:



Résultats

Les résultats concernant les colonnes Email et Phone donnent le nombre d'enregistrements correspondant et ne correspondant pas, respectivement, aux modèles d'e-mail standard et aux numéros de téléphone standard basés aux États-Unis. Les résultats donnent également le nombre de valeurs blanches et de valeurs en doublon. Ils montrent que les données ne sont pas vraiment cohérentes et qu'elles doivent être corrigées.

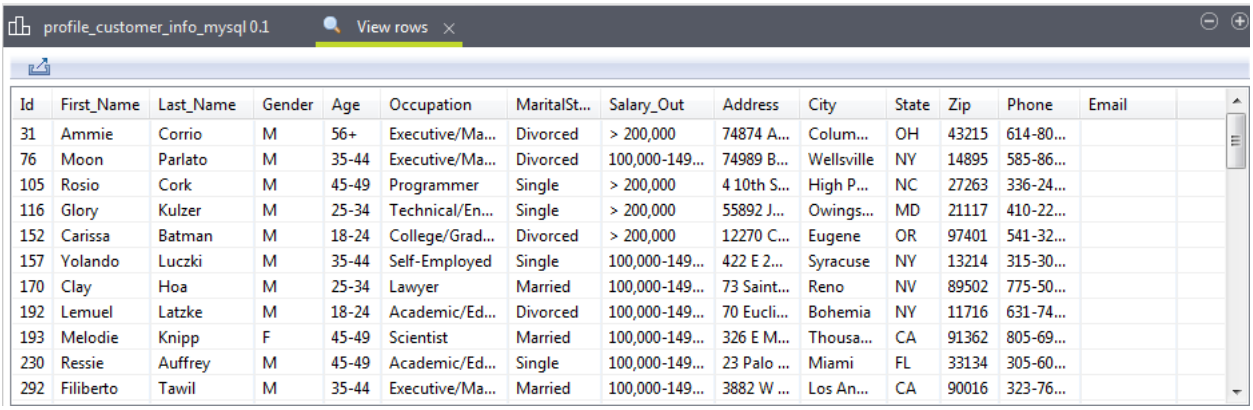
Explorer les données ne correspondant pas

Après le démarrage de l'analyse de colonnes, vous pouvez accéder à une vue des données correspondant et ne correspondant pas. Cela est très utile pour consulter les lignes invalides par exemple et pour commencer l'analyse de ce qui doit être effectué pour valider et nettoyer ces données.

Procédure

1. Dans la vue **Analysis Results**, cliquez-droit sur **Blank Count** dans les résultats statistiques de la colonne Email et sélectionnez **View rows** par exemple.

   Une vue s'ouvre listant toutes les lignes blanches dans la colonne Email.

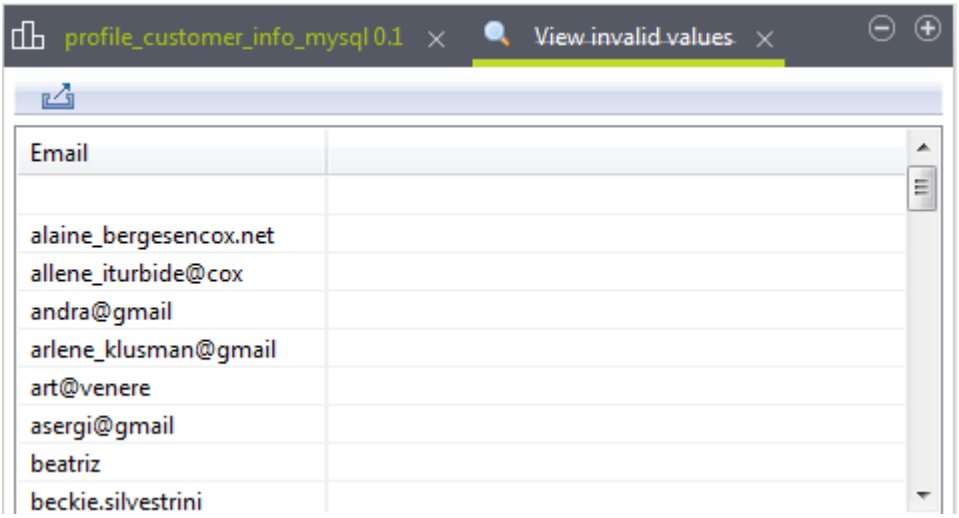


Id	First_Name	Last_Name	Gender	Age	Occupation	MaritalSt...	Salary_Out	Address	City	State	Zip	Phone	Email
31	Ammie	Corrio	M	56+	Executive/Ma...	Divorced	> 200,000	74874 A...	Colum...	OH	43215	614-80...	
76	Moon	Parlato	M	35-44	Executive/Ma...	Divorced	100,000-149...	74989 B...	Wellsville	NY	14895	585-86...	
105	Rosio	Cork	M	45-49	Programmer	Single	> 200,000	4 10th S...	High P...	NC	27263	336-24...	
116	Glory	Kulzer	M	25-34	Technical/En...	Single	> 200,000	55892 J...	Owings...	MD	21117	410-22...	
152	Carissa	Batman	M	18-24	College/Grad...	Divorced	> 200,000	12270 C...	Eugene	OR	97401	541-32...	
157	Yolando	Luczki	M	35-44	Self-Employed	Single	100,000-149...	422 E 2...	Syracuse	NY	13214	315-30...	
170	Clay	Hoa	M	25-34	Lawyer	Married	100,000-149...	73 Saint...	Reno	NV	89502	775-50...	
192	Lemuel	Latzke	M	18-24	Academic/Ed...	Divorced	100,000-149...	70 Eucli...	Bohemia	NY	11716	631-74...	
193	Melodie	Knipp	F	45-49	Scientist	Married	100,000-149...	326 E M...	Thousa...	CA	91362	805-69...	
230	Ressie	Auffrey	M	45-49	Academic/Ed...	Single	100,000-149...	23 Palo ...	Miami	FL	33134	305-60...	
292	Filiberto	Tawil	M	35-44	Executive/Ma...	Married	100,000-149...	3882 W ...	Los An...	CA	90016	323-76...	

2. Dans la vue **Analysis Results**, cliquez-droit sur le résultat dans le champ **Pattern Matching** de la colonne Email et sélectionnez **View invalid values** par exemple.

Résultats

Une vue s'ouvre listant toutes les adresses e-mail invalides.



Que faire ensuite ?

Vous avez découvert comment Studio Talend vous aide à profiler vos données et à collecter les statistiques et les informations relatives à ces données afin d'évaluer le niveau de qualité des données selon des objectifs définis.

Vous avez vu :

- Comment utiliser la perspective **Profiling** du studio pour analyser les adresses e-mail et les numéros de téléphone des clients à l'aide d'indicateurs et de modèles prêts à l'emploi.
- Comment les résultats d'analyse montrent les enregistrements des adresses correspondant et ne correspondants pas et comment explorer ces données.

Une fois que vous avez terminé les procédures simples décrites dans Identifier les anomalies dans les données à la page 12, vous pouvez commencer à voir plus en détail toutes les fonctionnalités de profiling de Studio Talend.

Pour plus d'informations concernant le Studio Talend, consultez le Guide utilisateur du Studio Talend.

Pour plus d'informations concernant les produits et les solutions Talend, consultez www.talend.com.