



# Talend Open Studio for Data Quality

Guide de prise en main

## 6.1.2

Convient à la version 6.1.2. Annule et remplace toute version antérieure de ce guide.

Date de publication : 13 septembre 2016

## **Copyleft**

Cette documentation est mise à disposition selon les termes du Contrat Public Creative Commons (CPCC).

Pour plus d'informations concernant votre utilisation de cette documentation en accord avec le Contrat CPCC, consultez : <http://creativecommons.org/licenses/by-nc-sa/2.0/>

## **Mentions légales**

Talend est une marque déposée de Talend, Inc.

Tous les noms de marques, de produits, les noms de sociétés, les marques de commerce et de service sont la propriété de leurs détenteurs respectifs.

## **Licence applicable**

Le logiciel décrit dans cette documentation est soumis à la Licence Apache, Version 2.0 (la "Licence"). Vous ne pouvez utiliser ce logiciel que conformément aux dispositions de la Licence. Vous pouvez obtenir une copie de la Licence sur <http://www.apache.org/licenses/LICENSE-2.0.html> (en anglais). Sauf lorsqu'explicitement prévu par la loi en vigueur ou accepté par écrit, le logiciel distribué sous la Licence est distribué "TEL QUEL", SANS GARANTIE OU CONDITION D'AUCUNE SORTE, expresse ou implicite. Consultez la Licence pour connaître la terminologie spécifique régissant les autorisations et les limites prévues par la Licence.

Ce produit comprend les logiciels développés par ASM, AntLR, Apache ActiveMQ, Apache Ant, Apache Axiom, Apache Axis, Apache Axis 2, Apache Chemistry, Apache Common Http Client, Apache Common Http Core, Apache Commons, Apache Commons Bcel, Apache Commons Lang, Apache Datafu, Apache Derby Database Engine and Embedded JDBC Driver, Apache Geronimo, Apache HCatalog, Apache Hadoop, Apache Hbase, Apache Hive, Apache HttpClient, Apache HttpComponents Client, Apache JAMES, Apache Log4j, Apache Neethi, Apache POI, Apache Pig, Apache Thrift, Apache Tomcat, Apache Xml-RPC, Apache Zookeeper, CSV Tools, DataNucleus, Doug Lea, Ezmorph, Google's phone number handling library, Guava : Google Core Libraries for Java, H2 Embedded Database and JDBC Driver, HighScale Lib, HsqlDB, JSON, JUnit, Jackson Java JSON-processor, Java API for RESTful Services, Java Universal Network Graph, Jaxb, Jaxen, Jetty, Joda-Time, Json Simple, MapDB, MetaStuff, Paracel JDBC Driver, PostgreSQL JDBC Driver, Protocol Buffers - Google's data interchange format, Resty : client simple HTTP REST pour Java, SL4J : Simple Logging Facade for Java, SQLite JDBC Driver, The Castor Project, The Legion of the Bouncy Castle, Woden, Xalan-J, Xerces2, XmlBeans, XmlSchema Core, atinject. Fournis sous leur licence respective.

---

# Table des matières

<b>Préface</b> .....	<b>v</b>
<b>1. Informations générales</b> .....	<b>v</b>
1.1. Objectif .....	v
1.2. Public visé .....	v
1.3. Conventions typographiques .....	v
<b>2. Remarques et Support</b> .....	<b>vi</b>
<b>chapitre 1. Prise en main du Studio Talend</b> .....	<b>1</b>
<b>1.1. Lancement du Studio Talend</b> .....	<b>2</b>
1.1.1. Lancer le Studio pour la première fois .....	2
1.1.2. Se connecter à TalendForge .....	2
<b>1.2. Une approche multi-perspectives</b> .....	<b>4</b>
1.2.1. Passer d'une perspective à une autre .....	4
1.2.2. Sauvegarder la configuration d'une perspective .....	5
<b>chapitre 2. Profiling des données</b> .....	<b>7</b>
<b>2.1. Profiling de données clients</b> .....	<b>8</b>
2.1.1. Identifier des anomalies de données .....	8



---

# Préface

## 1. Informations générales

### 1.1. Objectif

Ce guide a pour objectif de permettre aux utilisateurs de prendre en main rapidement *Talend Open Studio for Data Quality*. Pour plus d'informations détaillées concernant les fonctions et fonctionnalités de *Talend Open Studio for Data Quality*, consultez les autres Guides fournis avec *Talend Open Studio for Data Quality*.

Les informations présentées dans ce Guide s'appliquent à la version **6.1.2** de *Talend Open Studio for Data Quality*.

### 1.2. Public visé



Ce guide s'adresse aux utilisateurs et aux administrateurs du *Talend Open Studio for Data Quality*.



Dans ce document, l'interface présentée par le biais des captures d'écran peut être légèrement différente de la vôtre.

### 1.3. Conventions typographiques

Ce guide utilise les conventions typographiques suivantes :

- texte en **gras** : boutons et champs d'une fenêtre ou d'une boîte de dialogue, touches du clavier, menus et options des menus ;
- texte en **[gras]** : titre des fenêtres, assistants et boîtes de dialogue ;
- texte en *courier* : paramètres système renseignés par l'utilisateur ;
- texte en *italique* : nom des fichiers, schémas, colonnes, lignes ou variables dont il est question dans les exemples utilisés.
- L'icône  introduit un paragraphe apportant une information supplémentaire sur un point important. Elle introduit également les commentaires relatifs à un tableau ou une image.
- L'icône  introduit un message apportant une information relative aux modalités d'exécution, ou des recommandations particulières. Elle est également utilisée pour attirer l'attention de l'utilisateur sur certaines informations ou situations particulièrement importantes.
- Les commandes sont surlignées en gris clair.

## 2. Remarques et Support

Votre opinion nous intéresse, n'hésitez pas à nous faire part de vos remarques, suggestions, requêtes concernant la documentation ou le produit, et obtenez le support de notre équipe **Talend**, sur le Forum **Talend** (en anglais) disponible à l'adresse :

<http://talendforge.org/forum>



# Chapitre 1. Prise en main du Studio Talend

Ce chapitre fournit les informations requises pour prendre en main le *Studio Talend*, notamment des informations relatives au démarrage du *Studio Talend* et à la création de projets.

## 1.1. Lancement du Studio Talend

Ce chapitre vous guide à travers le lancement du *Studio Talend* pour la première fois et l'ouverture de votre premier projet dans le Studio. Il vous apporte également des informations sur la configuration des projets.

### 1.1.1. Lancer le Studio pour la première fois

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

1. Décompressez le fichier zip du *Studio Talend* et, dans le dossier, double-cliquez sur le fichier exécutable correspondant à votre système d'exploitation.



Le fichier archive .zip du studio contient des fichiers binaires pour différentes plateformes, notamment Mac OS X et Linux/Unix.

2. Dans la fenêtre [**User License Agreement**] qui s'ouvre, lisez et acceptez les termes de la licence pour procéder aux étapes suivantes.

3. Selon que vous utilisez, vous pouvez voir :

- Une présentation rapide du *Studio Talend*. Cliquez sur **Next** pour aller à l'étape suivante de la présentation, ou cliquez sur **Close** pour la terminer et afficher la fenêtre principale de votre *Studio Talend*.

Cette présentation démarre automatiquement lors du premier lancement du Studio. Pour l'ouvrir manuellement plus tard, allez dans **Help > Studio Quick Tour** dans la barre de menu du Studio.

- La fenêtre [**Welcome**], qui fournit des liens directs aux projets Démo, à la documentation utilisateur, aux tutoriels, au forum **Talend**, aux formations à la demande **Talend** et aux dernières actualités de **Talend**. Cliquez sur **Start now!** pour ouvrir la fenêtre principale du *Studio Talend*.

Lorsque l'assistant [**Additional Talend Packages**] s'ouvre, installez les packages supplémentaires, comme les langues dont vous avez besoin. Pour plus d'informations, consultez la section concernant l'installation de packages supplémentaires dans le *Guide d'installation et de migration Talend*.

4. Vous pouvez ignorer cette étape d'installation et fermer l'assistant en cliquant sur le bouton **Cancel**.

Cet assistant apparaît 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.

### 1.1.2. Se connecter à TalendForge

Tous les quatre démarrages du *Studio Talend*, jusqu'à ce que vous soyez connecté à la Communauté **Talend**, la boîte de dialogue [**Connect to TalendForge**] s'ouvre, vous invitant à vous connecter à la Communauté **Talend** afin que vous puissiez regarder, télécharger et installer des composants externes, charger vos propres composants afin de les partager avec les autres utilisateurs de **Talend**, le tout directement à partir de la vue **Exchange** de la création de Job dans le Studio.

Pour en apprendre plus sur la Communauté **Talend**, cliquez sur le lien **TalendForge Terms of Use**. Pour plus d'informations sur l'utilisation et le partage des composants avec la Communauté, consultez la section traitant du chargement/téléchargement de composants communautaires **Talend** dans le Guide utilisateur de votre Studio.

Si vous souhaitez vous connecter à la Communauté **Talend** ultérieurement, cliquez sur le bouton **Skip this Step** pour continuer le démarrage du Studio sans avoir à configurer une connexion à la Communauté **Talend**.



1. Par défaut, le Studio collecte automatiquement les données d'utilisation et les envoie périodiquement aux serveurs hébergés par **Talend**, dans un but d'amélioration. Si vous ne souhaitez pas que le Studio effectue ces envois, décochez la case **I want to help to improve Talend by sharing anonymous usage statistics**.

Vous pouvez désactiver la collecte de données dans la fenêtre **[Preferences]** (**Talend > Usage Data Collector**). Pour plus d'informations concernant le collecteur de données d'utilisation, consultez la section relative à la configuration des préférences, dans le Guide utilisateur de votre *Studio Talend*.

2. Renseignez les informations requises, cochez la case **I Agree to the TalendForge Terms of Use** puis cliquez sur **CREATE ACCOUNT** afin de créer votre compte, de vous connecter automatiquement à la Communauté **Talend** et de continuer à lancer le Studio.



Soyez assuré qu'aucune information personnelle que vous communiquez à **Talend** ne sera transmise à des tiers ou utilisée à d'autres fins que celles de vous connecter à la Communauté ou de vous informer sur **Talend** des dernières mises à jour de **Talend**.

Si vous avez déjà un compte sur <http://www.talendforge.org/>, cliquez sur **Connect to Existing Account**, saisissez votre identifiant et votre mot de passe, cliquez sur **CONNECT TO MY ACCOUNT** pour vous connecter à la Communauté **Talend** et continuer le lancement du Studio.

## Connect your Studio to TalendForge, the Talend Online Community.

- Download **new components and connectors** from Talend Exchange.
- Access the most recent **Documentation and Tech articles** from Talend social knowledgebase.
- See the latest messages in the Talend **Discussions Forums**.

A screenshot of a web form for connecting to a Talend account. It features two input fields: the first contains the text "user" and has a green checkmark to its right; the second contains a series of dots representing a password and also has a green checkmark to its right. Below the fields is a checked checkbox with the text "I want to help to improve Talend by sharing anonymous usage statistics". A prominent green button labeled "CONNECT TO MY ACCOUNT" is centered below the checkbox. At the bottom of the form, there are two buttons: "Create a New Account" on the left and "Skip this Step" on the right.

Cette page ne s'affiche plus à l'ouverture du Studio une fois que vous êtes connecté avec succès à la Communauté **Talend**. Pour afficher de nouveau la page, ouvrez la boîte de dialogue [**Préférences**], sélectionnez **Talend > Exchange** puis cliquez sur **Sign In**. Pour plus d'informations, consultez la section concernant les préférences du *Studio Talend* dans le Guide utilisateur de votre Studio.

## 1.2. Une approche multi-perspectives

Le *Studio Talend* offre un ensemble complet d'outils et de fonctions clés, notamment d'intégration de données et d'applications, de profiling de données et de gestion des données maître. Ces outils sont tous accessibles à partir des différentes perspectives du studio.

### 1.2.1. Passer d'une perspective à une autre

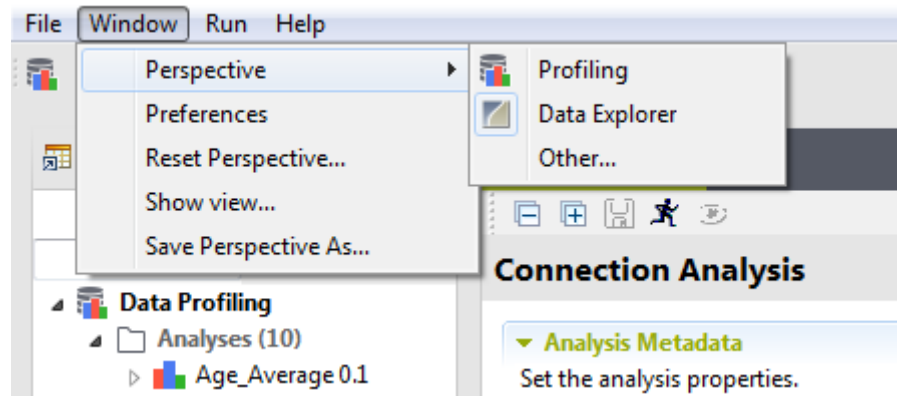
Il y a différentes manières de passer d'une perspective à une autre dans le studio.

Pour passer d'une perspective à une autre à l'aide des icônes d'accès rapide, procédez comme suit :

- Cliquez sur l'icône d'accès rapide dans le coin supérieur gauche du studio pour passer d'une perspective à une autre.

Vous pouvez également passer aux autres perspectives via la barre de menu :

1. Dans la barre de menu, cliquez sur **Window > Perspective**.



2. Sélectionnez dans la liste :

Élément	pour...
<b>Profiling</b>	ouvrir la perspective de profilage de données dans laquelle vous pouvez examiner les données disponibles dans les différents sources de données.
<b>Data Explorer</b>	ouvrir la perspective d'exploration de données dans laquelle vous pouvez parcourir et interroger les données analysées.
<b>Other...</b>	ouvrir une boîte de dialogue dans laquelle vous pouvez choisir d'ouvrir différentes perspectives étendant les fonctionnalités du studio.

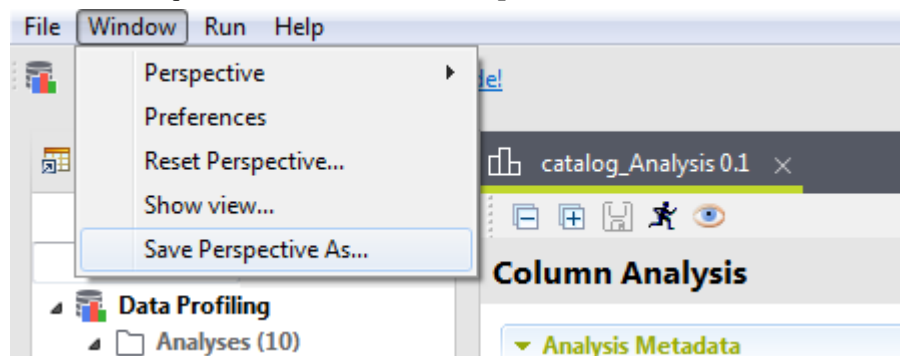
Il est également possible d'afficher des vues de perspectives différentes dans la perspective ouverte, à l'aide du menu **Window > Show view...**

## 1.2.2. Sauvegarder la configuration d'une perspective

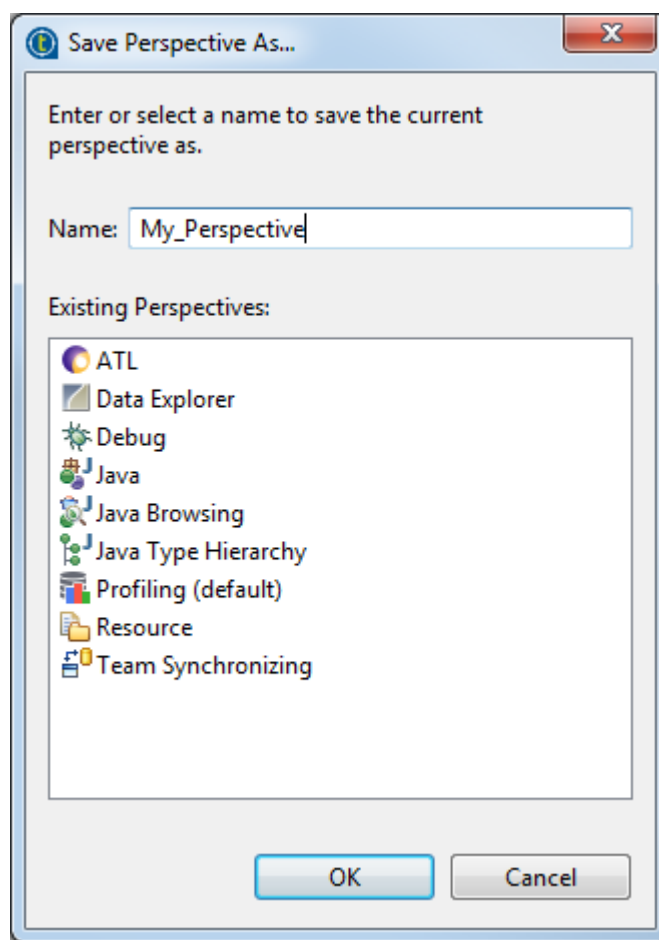
Vous pouvez sauvegarder la configuration de votre perspective active afin de la lister comme perspective, dans la boîte de dialogue des perspectives.

Pour sauvegarder la configuration de la perspective active, procédez comme suit :

1. Dans la barre de menu, cliquez sur **Window > Save Perspective As...**



2. Dans le champ **Name**, saisissez un nom.



3. Cliquez sur **OK**.

La perspective courante est sauvegardée comme nouvelle perspective sous le nouveau nom.

Vous pouvez ouvrir cette perspective à tout moment en cliquant sur l'icône d'accès rapide ou en la sélectionnant dans la boîte de dialogue [**Open Perspective**]. Pour plus d'informations, consultez [Passer d'une perspective à une autre](#).



## Chapitre 2. Profiling des données

Ce chapitre est destiné aux utilisateurs de la solution de qualité de données Talend cherchant des cas d'utilisation réelle pour leur permettre de maîtriser au mieux les produits de qualité de données.

Il décrit comment utiliser la perspective **Profiling** du studio pour profiler des données.

## 2.1. Profiling de données clients

Incorporer des outils de qualité de données appropriés au sein de vos processus métier est vital au commencement de tout projet et durant celui-ci, afin de déterminer le type de qualité de vos données et de décider des données à résoudre.

Imaginez, par exemple, que vous démarrez une campagne dans votre service commercial et votre service marketing, ou que vous devez contacter des clients pour la facturation et le paiement et que votre source principale pour contacter les personnes concernées sont les adresses e-mail et postales. Avoir des données cohérentes et correctes est vital dans des campagnes, afin de pouvoir contacter toutes les personnes que vous souhaitez contacter.

Cette section fournit un exemple de profiling d'adresses e-mail et postales de clients des Etats-Unis.

### 2.1.1. Identifier des anomalies de données

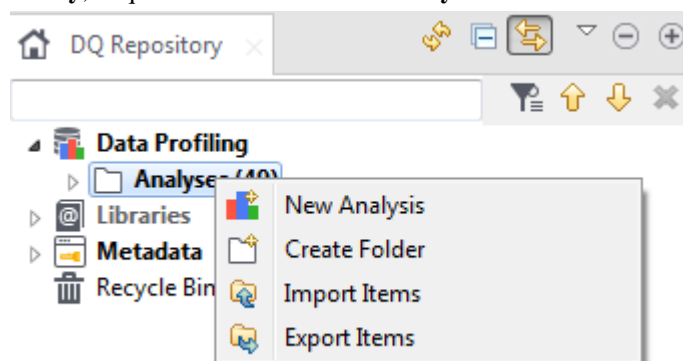
La première étape dans cet exemple est le profiling de vos ressources, ici les informations de contact des clients, dans une base de données MySQL. Les résultats du profiling vous fournissent des statistiques concernant les valeurs dans chaque colonne.

#### 2.1.1.1. Profiler les colonnes d'adresses

Utilisez votre studio pour analyser les colonnes de clients, notamment *email* et *postal*. A l'aide d'indicateurs et de modèles natifs sur ces colonnes, les résultats d'analyse affichent les données d'adresses qui correspondent et ne correspondent pas, le nombre d'enregistrements les plus fréquents pour chaque modèle distinct, ainsi que le nombre de lignes, de doublons et de blancs dans chaque colonne.

#### Définir l'analyse de colonnes

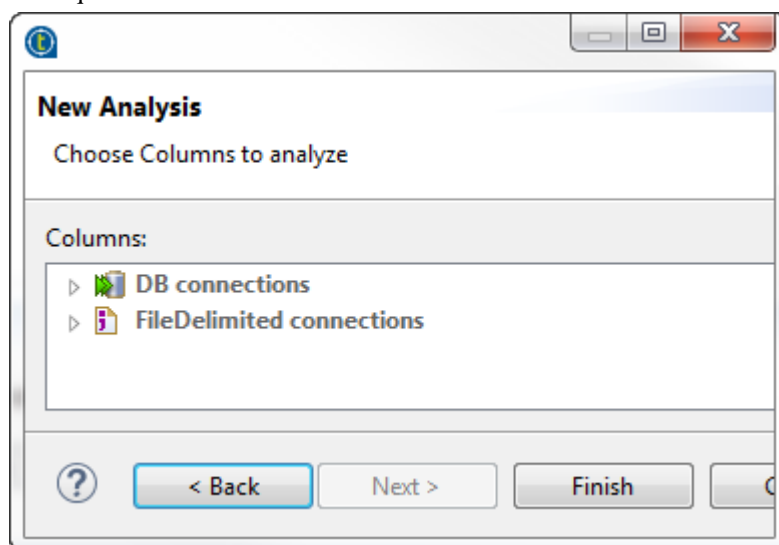
1. Dans la vue **DQ Repository**, cliquez-droit sur le dossier **Analysis** et sélectionnez **New Analysis**.



L'assistant **[Create New Analysis]** s'ouvre.



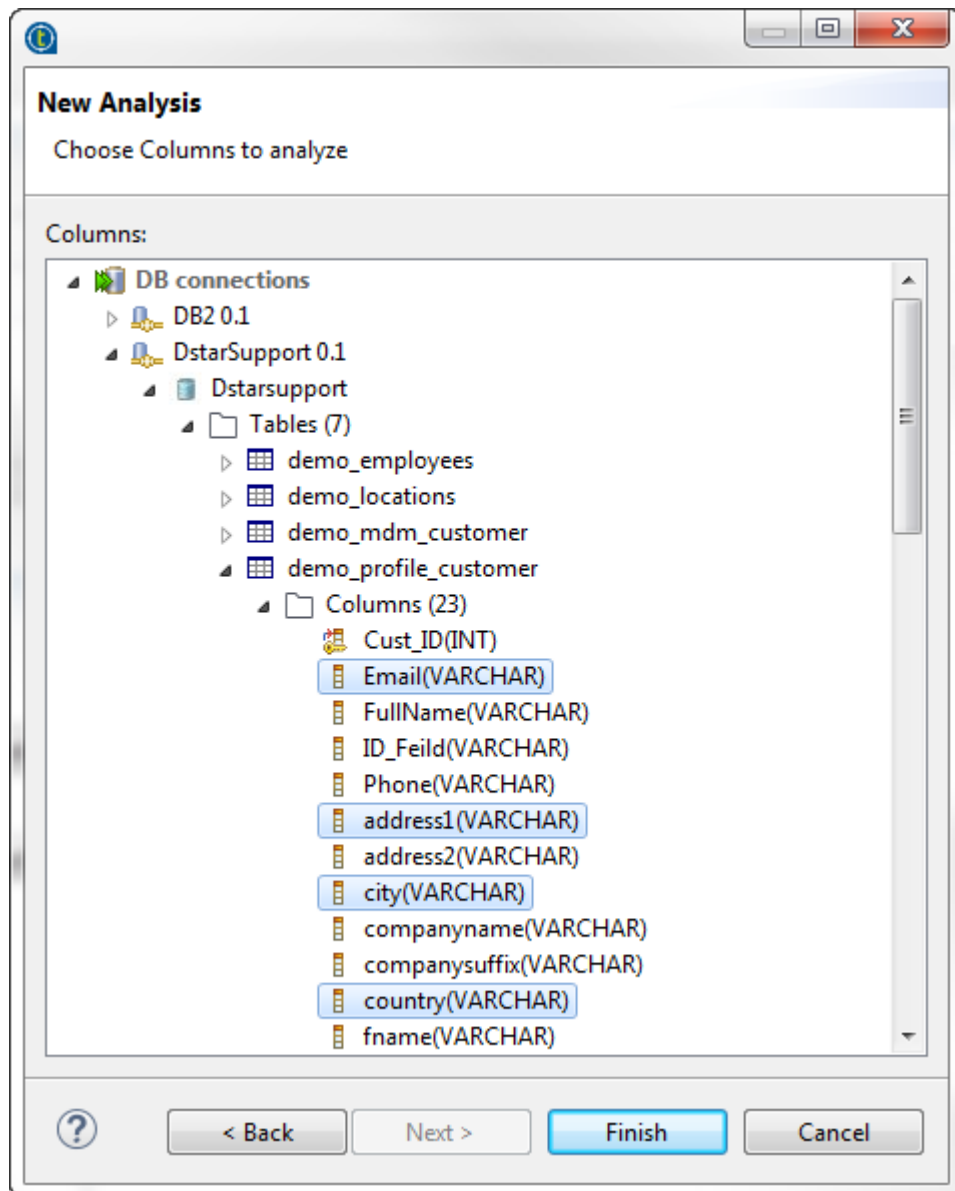
4. Configurez les métadonnées de l'analyse de colonnes (objectif, description et nom de l'auteur) dans les champs correspondants puis cliquez sur **Next**.



### Sélectionner les colonnes d'adresse et configurer les données d'exemple

1. Développez le nœud **DB connections** et parcourez-le jusqu'aux colonnes d'adresses que vous souhaitez analyser.





2. Sélectionnez les colonnes et cliquez sur **Finish** pour fermer l'assistant.

Un fichier pour la nouvelle analyse de colonnes s'affiche sous le nœud **Analysis** de la vue **DQ Repository** et l'éditeur d'analyse s'ouvre sur les métadonnées de cette analyse.

**Column Analysis**

▼ **Analysis Metadata**  
Set the analysis properties.

Name:

Purpose:

Description:

Author:

Status:

▼ **Data preview**

Connection:  Version:0.1

Limit

	Email	postal	city	state	country	address1
1	DebraEvans@fa...	16054	Saint Petersburg	PA	US	5870 E EVANS CT
2	TeresaBailey@fa...	94188	San Francisco	CA	US	5411 S THROOP ST
3	JeanMiller@gma...	5477	Richmond	VT	US	8004 E WASHINGTON S
4	HenryMartin@g...	23642	Virginia Beach	VA	US	6383 NW SCHICK PL
5	SandraMorgan@...	26036	Dallas	WV	US	9849 W ST CLAIR ST
6	EvelynWalker@g...	5841	Greensboro	VT	US	8738 S ACADEMY PL
7	KathleenFoster@...	89199	Las Vegas	NV	US	10900 SW BANKS ST
8	BrendaBaker@h...	17501	Akron	PA	US	4420 W CULLERTON ST
9	ShirleyBrown@fr...	23642	Virginia Beach	VA	US	5282 NW WILSON AV

3. Dans la vue **Data preview**, cliquez sur **Refresh Data**.

Les données des colonnes sélectionnées sont affichées dans la table.

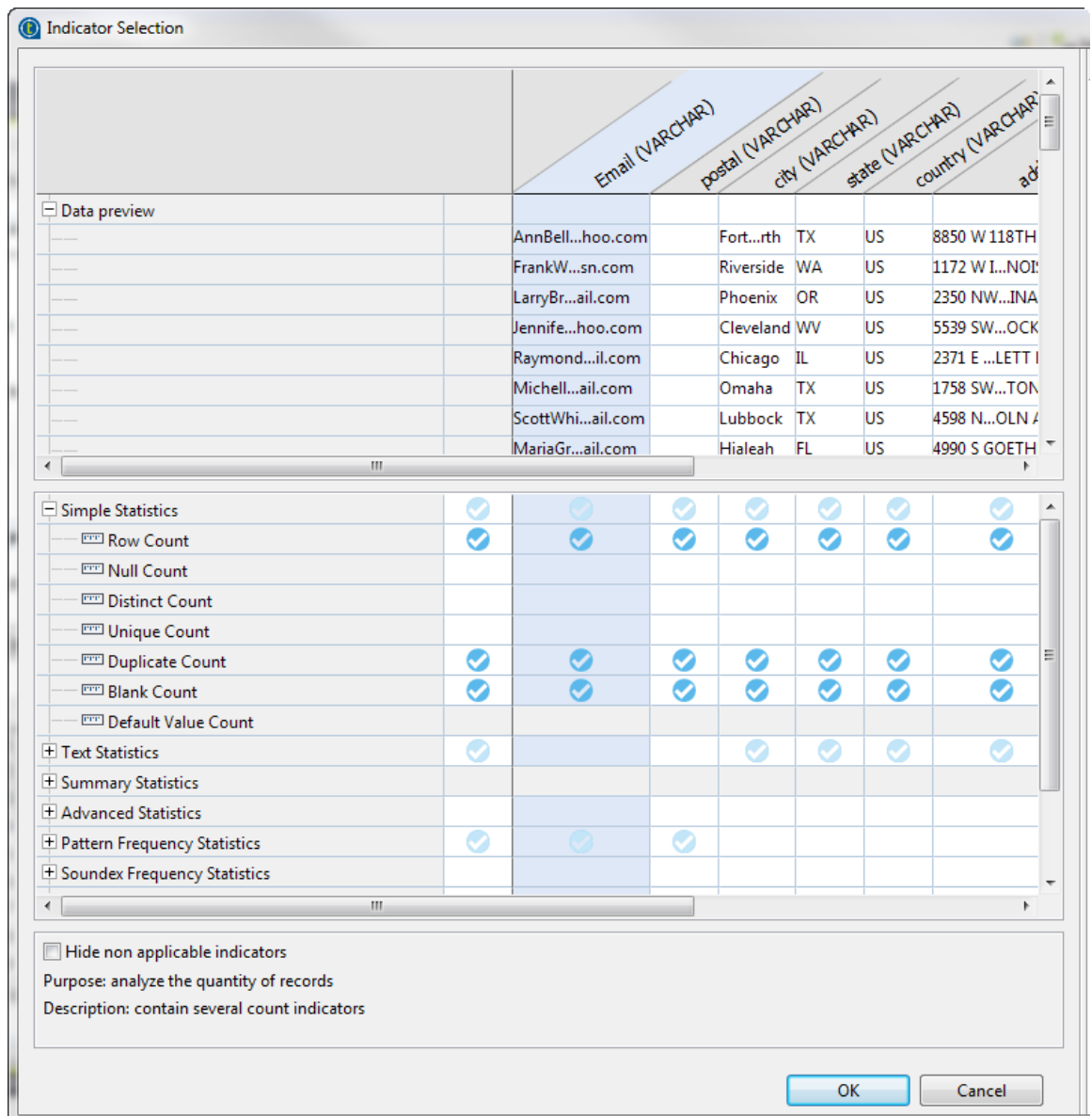
Vous pouvez modifier votre source de données et les colonnes sélectionnées à l'aide des boutons **New Connection** et **Select Data**, respectivement.

4. Dans le champ **Limit**, saisissez *50* pour le nombre d'enregistrement que vous souhaitez afficher dans la table et utiliser comme données d'exemple.
5. Sélectionnez **n random rows** afin de lister *50* enregistrements aléatoires des colonnes sélectionnées.

Pour plus d'informations concernant les analyses de colonnes, consultez le chapitre relatif aux analyses de colonnes dans le *Guide utilisateur du Studio Talend*.

## Configurer les indicateurs système

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



2. Cliquez dans les cellules à côté des noms d'indicateurs afin de les paramétrer pour les colonnes analysées et cliquez sur **OK**.


Dans cet exemple, vous souhaitez consulter le nombre de lignes, de blancs et de doublons dans toutes les colonnes, afin de voir si les données sont cohérentes. L'indicateur **Pattern Frequency Table** est utilisé sur les colonnes *email* et *postal* afin de calculer le nombre des enregistrements les plus fréquents pour chaque modèle ou valeur distinct(e).

Les indicateurs sont ajoutés aux colonnes dans la vue **Analyzed Columns**.

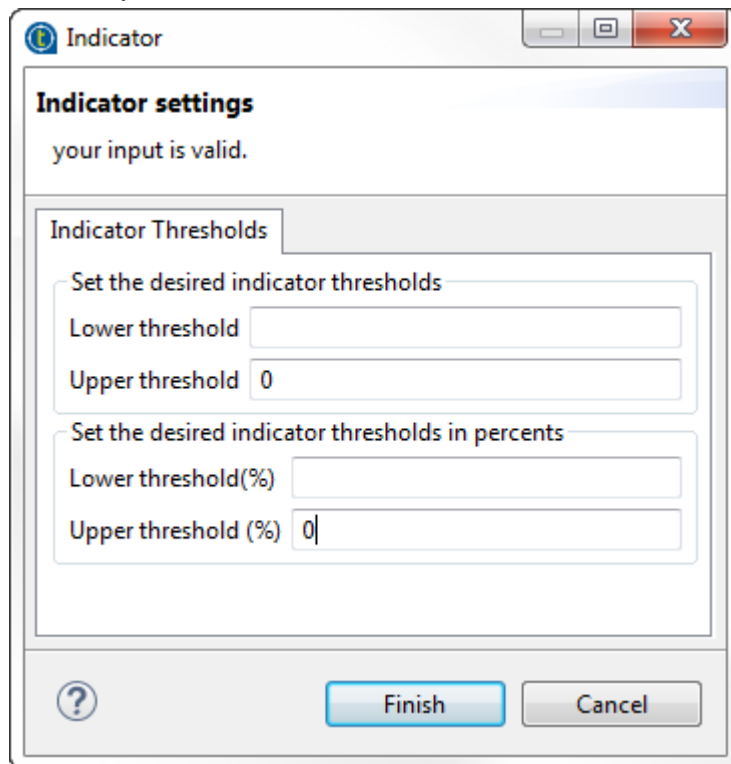
▼ Analyzed Columns

Go  |< < > >| 1/2

Analyzed Columns	Datamining Type	Pattern	UDI	Operation
<ul style="list-style-type: none"> <li>▶ Email (VARCHAR)                             <ul style="list-style-type: none"> <li>▶ Blank Count</li> <li>▶ Duplicate Count</li> <li>▶ Pattern Frequency Table</li> <li>▶ Row Count</li> <li>▶ Email Address</li> </ul> </li> <li>▶ postal (VARCHAR)</li> <li>▶ city (VARCHAR)</li> <li>▶ state (VARCHAR)</li> <li>▶ country (VARCHAR)</li> </ul>	Nominal			<span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span> <span style="color: red; font-size: 2em;">✖</span>

3. Cliquez sur l'icône d'option  à côté de l'indicateur **Blank Count** et saisissez 0 dans le champ **Upper threshold**.


Définir des seuils sur les indicateurs est très utile. Cela permet de marquer en rouge le nombre de valeurs nulles dans les résultats d'analyse.

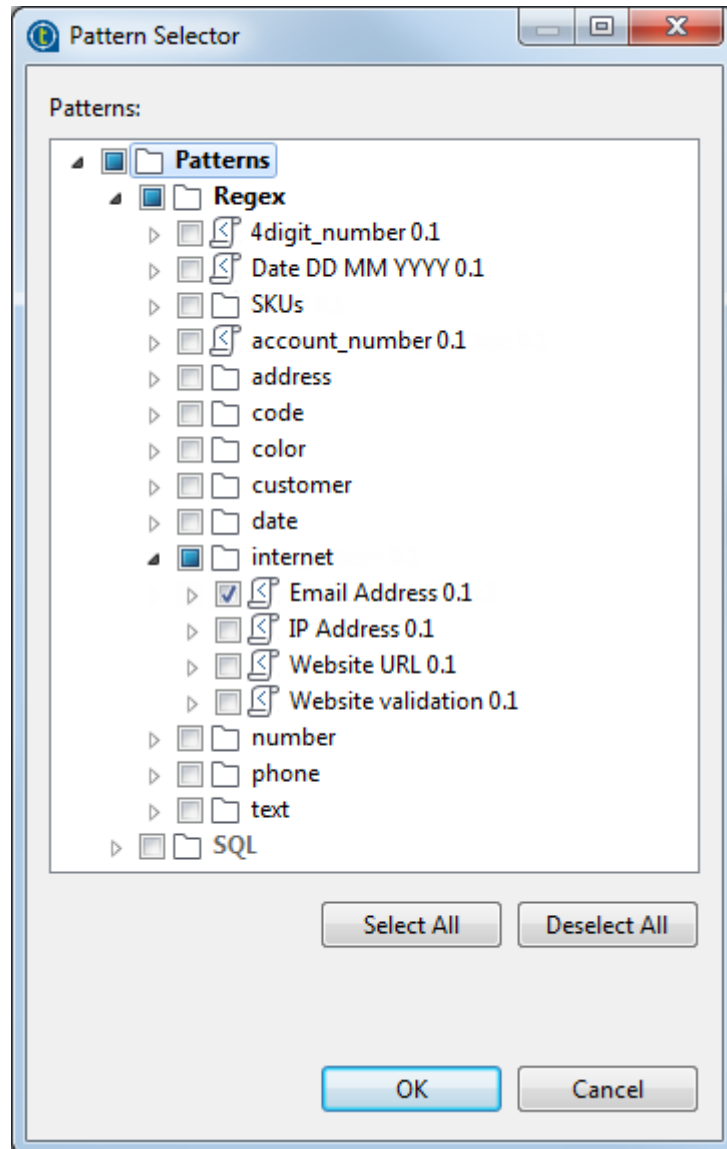



Pour plus d'informations concernant les types d'indicateurs et leur utilisation lors d'analyses de données, consultez le *Guide utilisateur du Studio Talend*.

## Configurer les modèles

Vous allez mettre en correspondance le contenu de la colonne *email* par rapport au format d'e-mail standard et le contenu de la colonne *postal* par rapport au format standard des code postaux des Etats-Unis. Cela permet de définir le contenu, la structure et la qualité des adresses e-mail et des codes postaux, ainsi que donner un pourcentage des données correspondant aux formats standards et des données ne correspondant pas.

1. Dans la vue **Analyzed Columns**, cliquez sur l'icône  à côté de la colonne *email*.



2. Dans la boîte de dialogue **[Pattern Selector]**, développez **Regex** et parcourez l'arborescence jusqu'au nœud **Email Address**, dans le dossier **internet**, puis cliquez sur **OK**.
3. Cliquez sur l'icône d'option  à côté de l'indicateur **Email Address** et saisissez **98.0** dans le champ **Lower threshold (%)**.

Si le nombre d'enregistrements correspondant au modèle est inférieur à 98 %, ils seront marqués en rouge dans les résultats de l'analyse.

4. Répétez l'opération pour ajouter la colonne *postal* au modèle **US Zipcode Validation** depuis le dossier **address**.

Pour plus d'informations concernant les types de modèles et leur utilisation lors d'analyses de données, consultez le *Guide utilisateur du Studio Talend*.

## Exécuter l'analyse et afficher les résultats du profiling

1. Sauvegardez l'analyse de colonnes dans l'éditeur d'analyse puis appuyez sur **F6** pour l'exécuter.

Un groupe de diagrammes s'affiche dans le panneau **Graphics**, à droite de l'éditeur d'analyse, et montre les résultats de l'analyse de colonnes, notamment ceux de la mise en correspondance des modèles.

2. Cliquez sur l'onglet **Analysis Results** au bas de l'éditeur d'analyse pour accéder à une vue plus détaillée des résultats.

Ces résultats affichent les graphiques générés pour les colonnes analysées, ainsi que les tables détaillant les résultats des statistiques et des mises en correspondance des modèles.

Les résultats pour la colonne *email* se présentent comme suit :



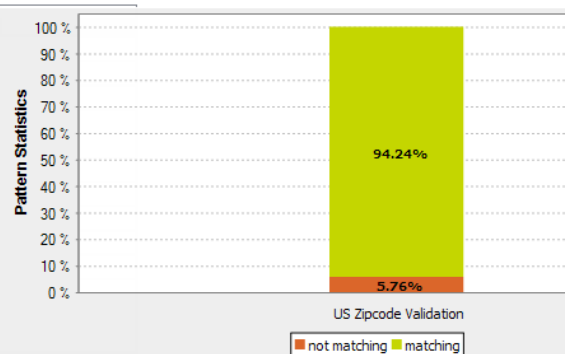
Les résultats de la mise en correspondance des modèles montrent qu'environ 10 % des adresses e-mail enregistrées ne correspondent pas au modèle standard. Les résultats des statistiques simples montrent qu'environ 8 % des adresses e-mail enregistrées sont vides et qu'environ 5% sont des doublons. Les résultats de fréquence des modèles donnent le nombre d'enregistrements les plus fréquents pour chaque modèle distinct. Cela permet de voir si les données sont cohérentes ou si vous devez corriger et nettoyer les données d'e-mails avant de commencer votre campagne.

Les résultats de la colonne *postal* se présentent comme suit :

▼ Column:demo\_profile\_customer.postal

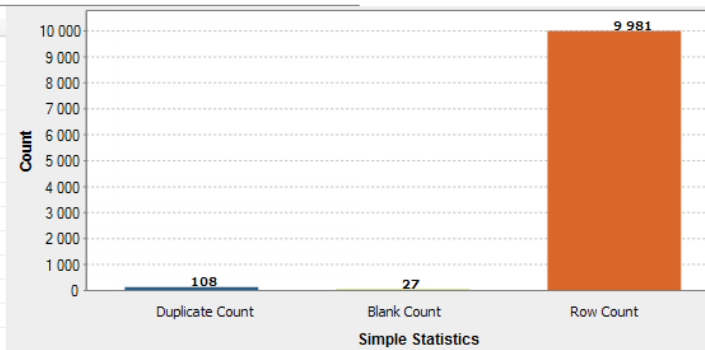
▼ Pattern Matching

Label	%Match	%No Match	#Match	#No Match
US Zipcode Validation	94.24%	5.76%	9406.0	575.0



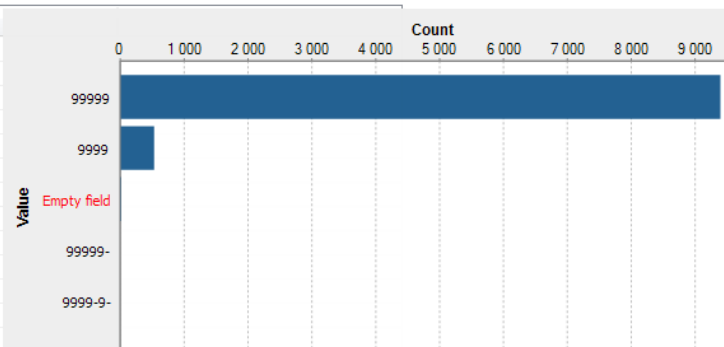
▼ Simple Statistics

Label	Count	%
Duplicate Count	108.00	1.08%
Blank Count	27.00	0.27%
Row Count	9981.00	100.00%



▼ Pattern Frequency Statistics

value	count	%
99999	9406.00	94.24%
9999	540.00	5.41%
Empty field	27.00	0.27%
99999-	7.00	0.07%
9999-9-	1.00	0.01%



Les ensembles de résultats pour la colonne *postal* donnent le nombre d'enregistrements qui correspondent et le nombre d'enregistrements qui ne correspondent pas au format standard des codes postaux des Etats-Unis. Les ensembles de résultats donnent également le nombre de blancs et de doublons, ainsi que le nombre d'enregistrements les plus fréquents pour chaque modèle distinct. Ces résultats montrent que les données ne sont pas vraiment cohérentes.

Un certain pourcentage des clients ne peut être contacté ni par e-mail ni par courrier. Ces résultats montrent clairement que vos données ne sont pas vraiment cohérentes et qu'il faut qu'elles soient corrigées.

### 2.1.1.2. Visualiser les données analysées

Après exécution de l'analyse de colonnes à l'aide du moteur SQL, dans la vue **Analysis Results** de l'éditeur d'analyse, vous pouvez cliquer-droit sur l'une des barres ou lignes des graphiques et ainsi accéder à une vue des données analysées. Cela peut être très utile pour voir les lignes invalides, par exemple, et trouver la manière d'ajuster les données.

Pour visualiser et exporter les données analysées, procédez comme suit :

1. Au bas de l'éditeur d'analyse, cliquez sur l'onglet **Analysis Results** pour ouvrir une vue détaillée des résultats d'analyse.
2. Cliquez-droit sur une ligne de données, dans les résultats des statistiques de la colonne *Email* et sélectionnez **View rows**, par exemple.

La perspective **Data Explorer** s'ouvre et liste les lignes invalides dans la colonne *email*.

Cust_ID	FullName	fname	lname	address1	city	state	postal	country	region	Phone	Email
3	Larry Bryant	Larry	Bryant	2350 NW CHINA PL	Phoenix	OR		US	USNWEST	834-367-3427	LarryBryant@gmail.com
23	Sarah Taylor	Sarah	Taylor	3284 W NEWBURG AV	Irving	TX	75063	US	USSOUTH	815-522-9116	SarahTaylor@yahoo.com
91	Alice Torres	Alice	Torres	3656 S PEARSON ST	Santa ...	CA	92799	US	USWEST	212-856-5683	AliceTorres@yahoo.com
119	Marie Murphy	Marie	Murphy	7687 S TRIPP AV	Charlo...	VT	5445	US	<null>	813-723-6681	MarieMurphy@yahoo.com
193	Kevin Garcia	Kevin	Garcia	2529 N CLARENCE AV	Oakland	TX	78951	US	USSOUTH	604-834-6474	KevinGarcia@yahoo.com
210	Scott Brooks	Scott	Brooks	10714 NE SCOTT ST	Fresno	TX	77545	US	USSOUTH	802-359-3898	ScottBrooks@yahoo.com
238	Joyce Cooper	Joyce	Cooper	10004 W PLEASANT ...	Orlando	WV	26412	US	USEAST	212-963-1775	JoyceCooper@gmail.com
298	Karen Thomas	Karen	Thomas	6651 N PONCHART...	Corpus...	TX	78480	US	USSOUTH	816-779-3247	KarenThomas@gmail.com
310	Betty Watson	Betty	Watson	7983 SW FARRAR DR	Omaha	TX	75571	US	USSOUTH	304-701-7857	BettyWatson@yahoo.com
349	Betty Bailey	Betty	Bailey	8360 NE 120TH ST	San Fra...	CA	94188	US	USWEST	622-835-3474	BettyBailey@yahoo.com
378	Nancy Bryant	Nancy	Bryant	10141 S 41ST ST	Saint P...	PA	16054	US	USEAST	623-965-5728	NancyBryant@gmail.com
379	Karen Morgan	Karen	Morgan	10166 E ABERDEEN ST	Kansas...	MO	64999	US	USMWST	303-342-8553	KarenMorgan@gmail.com
393	Nancy Hughes	Nancy	Hughes	10418 SE 45TH ST	Durham	PA	18039	US	USEAST	709-724-5816	NancyHughes@yahoo.com
394	Joyce Parker	Joyce	Parker	5123 W 32ND ST	Aurora	WV	26705	US	USEAST	508-717-4278	JoyceParker@yahoo.com
400	Henry Rogers	Henry	Rogers	6358 NE CHELTENH...	Wichita	KS	67278	US	USMWST	812-873-8641	HenryRogers@yahoo.com
430	Frank Foster	Frank	Foster	8198 W KIRKLAND AV	Phoenix	OR	97535	US	USNWEST	603-876-5331	FrankFoster@gmail.com
446	Betty Brooks	Betty	Brooks	2568 SW CABRINI ST	Cincin...	OH	45999	US	USMWST	307-392-4335	BettyBrooks@yahoo.com
453	Marie Torres	Marie	Torres	3631 SE 47TH PL	Minne...	NC	28652	US	USEAST	715-402-7674	MarieTorres@gmail.com
495	Frank Howard	Frank	Howard	4898 S FARRELL ST	Spokane	WA	99299	US	USNWEST	738-201-3245	FrankHoward@yahoo.com
496	Scott Powell	Scott	Powell	11065 W TORRENCE...	Clevena...	WV	26215	US	USEAST	218-791-8511	ScottPowell@yahoo.com
506	David Morris	David	Morris	7872 SE MAGNET AV	Akron	PA	17501	US	USEAST	734-791-2962	DavidMorris@gmail.com
539	Linda Taylor	Linda	Taylor	7498 S CASTLE ISLA...	Saint L...	OK	74866	US	USWEST	403-861-9549	LindaTaylor@gmail.com
586	Carol Harris	Carol	Harris	10191 SE EDWARD B...	Portland	TX	78374	US	USSOUTH	808-611-4555	CarolHarris@yahoo.com
647	Brian Powell	Brian	Powell	2562 NW CORBETT ...	Durham	PA	18039	US	USEAST	406-889-9681	BrianPowell@yahoo.com
650	Carol Barnes	Carol	Barnes	9642 E HICKORY AV	Detroit	TX	75436	US	USSOUTH	627-747-1229	CarolBarnes@yahoo.com