



Pentaho Data Integration

Spoon Manuel de l'utilisateur

traduction française *en cours*
réalisée du début au paragraphe 10.4.18
(16 mars 2007)
non relu

en cas de doute se référer au document original en anglais

Copyright © 2007 Pentaho Corporation. Diffusion autorisée. Toutes les marques sont la propriété de leurs propriétaires respectifs. Pour les informations récentes, visitez notre site internet www.pentaho.org

Dernière modification le 16 mars 2007

Table des matières

| | |
|--|----|
| 1.A propos de ce document..... | 5 |
| 1.1.Ce qu'il est :..... | 5 |
| 1.2.Ce qu'il n'est pas :..... | 5 |
| 1.3.Autres documentations..... | 5 |
| 2.Introduction à Spoon..... | 6 |
| 2.1.Qu'est Spoon ?..... | 6 |
| 2.2.Installation..... | 6 |
| 2.3.Lancer spoon..... | 7 |
| 2.4.Plates-formes supportées..... | 7 |
| 2.5.Problèmes connus..... | 7 |
| 2.6.Captures d'écran..... | 8 |
| 2.7.Options de la ligne de commande..... | 9 |
| 2.8.Référentiel..... | 11 |
| 2.9.Connexion automatique au référentiel..... | 12 |
| 2.10.License..... | 13 |
| 2.11.Notes..... | 13 |
| 2.12.Définitions..... | 13 |
| 2.13.Options..... | 15 |
| 2.14.Rechercher..... | 22 |
| 2.15.Variables d'environnement..... | 22 |
| 2.16.Historique du journal d'exécution..... | 22 |
| 2.17.Rejouer..... | 22 |
| 2.18.Réaliser une correspondance de champs..... | 23 |
| 2.19.Mode sécurisé dit de « contrôle renforcé »..... | 24 |
| 2.20.Onglet de bienvenue..... | 24 |
| 3.Créer une transformation ou une tâche..... | 25 |
| 4.Connexion aux bases de données..... | 26 |
| 4.1.Description..... | 26 |
| 4.2.Capture d'écran..... | 26 |
| 4.3.Options..... | 26 |
| 4.4.Liste des bases de données supportées..... | 28 |
| 4.5.Utilisation..... | 31 |
| 4.6.Bases de données non supportées..... | 33 |
| 5.Éditeur SQL..... | 35 |

| | |
|--|-----|
| 5.1.Description..... | 35 |
| 5.2.Capture d'écrans..... | 35 |
| 6.Explorateur de base de données..... | 36 |
| 6.1.Capture d'écran..... | 36 |
| 6.2.Description..... | 36 |
| 7.Liens..... | 37 |
| 7.1.Description..... | 37 |
| 7.2.Capture d'écrans..... | 37 |
| 7.3.Créer un lien..... | 38 |
| 7.4.Séparer un lien..... | 38 |
| 7.5.Boucles..... | 38 |
| 8.Variables..... | 39 |
| 8.1.Utilisation des variables..... | 39 |
| 8.2.Portée des variables..... | 39 |
| 9.Paramétrage des transformations..... | 42 |
| 9.1.Description..... | 42 |
| 9.2.Captures d'écran..... | 42 |
| 9.3.Paramètres..... | 43 |
| 9.4.Extra..... | 45 |
| 10.Les étapes au sein d'une transformation..... | 46 |
| 10.1.Description..... | 46 |
| 10.2.Lancer plusieurs copies de la même étape..... | 46 |
| 10.3.Distribuer ou copier ?..... | 47 |
| 10.4.Types d'étapes au sein d'une transformation..... | 48 |
| 11.Paramétrage des tâches..... | 208 |
| 11.1.Description..... | 208 |
| 11.2.Captures d'écran..... | 208 |
| 11.3.Options..... | 208 |
| 11.4. Extra..... | 209 |
| 12. Job Entries..... | 210 |
| 12.1. Description..... | 210 |
| 12.2. Job Entry Types..... | 210 |
| 13. Représentation graphique..... | 237 |
| 13.1. Description..... | 237 |
| 13.2. Adding steps or job entries..... | 238 |
| 13.3. Hiding a step..... | 239 |
| 13.4. Transformation Step options (right click popup menu)..... | 239 |

| | |
|---|-----|
| 13.5. Job entry options (right click popup menu)..... | 240 |
| 13.6. Adding hops..... | 240 |
| 14. Running a Transformation..... | 242 |
| 14.1. Running a Transformation Overview..... | 242 |
| 14.2. Captures d'écran..... | 242 |
| 14.3. Execution Options..... | 242 |
| 14.4. Setting up Remote and Slave Servers..... | 243 |
| 14.5. Clustering..... | 246 |
| 15. Logging..... | 252 |
| 15.1. Logging Description..... | 252 |
| 15.2. Captures d'écrans..... | 252 |
| 15.3. Log Grid..... | 253 |
| 15.4. Buttons..... | 254 |
| 16. Grids..... | 258 |
| 16.1. Description..... | 258 |
| 16.2. Functions..... | 258 |
| 16.3. Navigating..... | 259 |
| 17. Repository Explorer..... | 260 |
| 17.1. Description..... | 260 |
| 17.2. Captures d'écran..... | 260 |
| 17.3. Right click functions..... | 260 |
| 17.4. Backup / Recovery..... | 261 |
| 18. Shared objects..... | 262 |
| 19. APENDIX A..... | 264 |

1. A propos de ce document

1.1. Ce qu'il est :

Ce document est la documentation technique de Spoon,, l'interface graphique de modélisation des transformations et tâches (regroupant depuis la version Spoon et Chef) de Pentaho Data Integration (anciennement Kettle).

1.2. Ce qu'il n'est pas :

Nous ne décrivons pas précisément dans ce document comment créer une transformation ou une tâche dans toutes les situations. Bien évidemment, chaque développeur prendra un chemin différent pour l'intégration de ses données. Et c'est comme cela que cela doit être ! Spoon essaie de donner aux utilisateurs de l'outil une grande liberté sur la façon d'implémenter leur solution d'intégration.

1.3. Autres documentations

Vous trouverez ci-dessous des liens vers d'autres documentations :

- démonstration Flash, captures d'écrans et une introduction pour réaliser votre première transformation :
<http://kettle.pentaho.org/Captures d'écrans/>
- Site internet de la communauté « Pentaho Data Integration » – actualité, études de cas, news, astuces de la semaine et plus :
<http://kettle.pentaho.org>
- Le forum de la communauté « Pentaho Data Integration » – discussions sur la modélisation, on design, caractéristiques, bogues et ajouts :
<http://forums.pentaho.org/forumdisplay.php?f=69>
- Exécuter des transformation en temps différé (batch) en utilisant Pan :
Pan-4-fr_FR.pdf
- Exécuter des tâches en temps différé (batch) en utilisant Kitchen :
Kitchen-4-fr_FR.pdf
- Une introduction à la plate-forme « Pentaho Data Integration »
 - en anglais sur le blog de Roland Bouman :
<http://rpbouman.blogspot.com/2006/06/pentaho-data-integration-kettle-turns.html>
 - en français avec le document réalisé par Samatar Hassan
Introduction_PDI.pdf
 - en anglais sur le blog de Nicholas Goodman (Directeur Business Intelligence de Pentaho) :
<http://www.nicholasgoodman.com>

2. Introduction à Spoon

2.1. Qu'est Spoon ?

Kettle est l'acronyme récursif de "Kettle E.T.T.L. Environment". Il a été réalisé pour répondre aux besoins Extraxtion Transport Transformation and Load : l'extraction, le transport, la transformation et le chargement de données.

Spoon est une interface graphique orientée utilisateur qui permet de modéliser des transformations (ndt en anglais transofrmation) et des tâches (ndt en anglais job) qui peuvent être exécutées par les outils Kettle Pan (pour les transformations), Kitchen (pour les tâches). Pan est un moteur de transformation de données qui est doté d'une multitude de fonctionnalités sur les données : lecture, manipulation, écriture de données depuis et vers de nombreuses sources de données. Kitchen est un programme qui exécute des tâches modélisées par Spoon (ndt anciennement Chef intégré à Spoon depuis la version 0) enregistrées dans un fichier XML ou un référentiel (en anglais repository). Les tâches sont habituellement planifiées en temps différé (batch mode) afin d'être exécutées à intervalles de temps réguliers.

NOTE: Pour une description détaillée de Pan ou Kitchen, se référer aux documentations de ces produits.

Les transformations et les tâches sont décrites dans un fichier XML ou enregistrées au sein d'un référentiel. Ces informations sont lues par Pan et Kitchen afin d'exécuter les étapes décrites au sein d'une transformation ou d'exécuter une tâche.

Bref: **Pentaho Data Integration rend facile la réalisation, la mise à jour et la maintenance de votre entrepôt de données (ndt data warehouses)!**

2.2. Installation

La première étape est l'installation du « Java Runtime Environnement » (JRE) publié par Sun Microsystems, dans sa version 4 ou supérieure. Vous pouvez le télécharger gratuitement à l'adresse suivante

<http://www.java.com/>.

Ensuite, vous décompressez simplement l'archive Kettle-zip dans le répertoire de votre choix.

Dans le répertoire dans lequel vous avez décompressé l'archive, vous trouverez un certain nombre de fichiers, dont des lanceurs.

Dans l'environnement UN*X (Solaris, Linux, MacOSX...), vous devez rendre les lanceurs exécutables. Lancez ces commandes dans le répertoire de Kettle pour rendre tous les lanceurs exécutables.

```
cd Kettle

chmod +x *.sh
```

2.3. Lancer spoon

Pour exécuter Spoon, il vous est fourni des lanceurs différents selon la plate-forme :

Spoon.bat: exécute Spoon en environnement Windows.

spoon.sh: exécute Spoon en environnement UNIX/Linux/MacOSX, Solaris...

SI vous désirez créer un raccourci en environnement Windows, une icône vous est fournie : "spoon.ico". Pour utiliser la bonne icône, pointez votre raccourci vers la bonne icône en éditant ce raccourci.

2.4. Plates-formes supportées

L'interface graphique orientée utilisateur (GUI) Spoon est supportée sur les plates-formes suivantes :

- Microsoft Windows: toutes plates-formes depuis Windows 95, Vista incluse
- Linux GTK: sur architecture i386 et x86_64, fonctionne mieux sur l'environnement Gnome
- MacOS X: fonctionne sur les deux architectures :PowerPC et Intel.
- Solaris: utilisant l'interface Motif (GTK facultatif)
- AIX: utilisant l'interface Motif
- HP-UX: utilisant l'interface Motif (GTK facultatif)
- FreeBSD: support de l'architecture i386, non encore supporté sur architecture x86_64

2.5. Problèmes connus

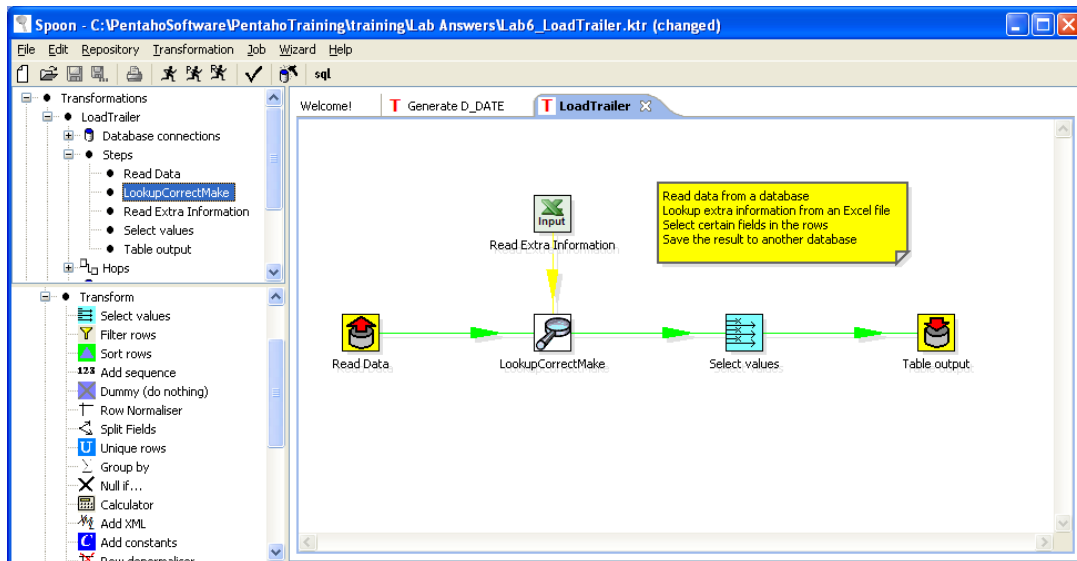
Linux

Des crash occasionnels avec la machine virtuelle Java sous SuSE et KDE. Aucun problème rencontré avec Gnome (détecté sur SUSE Linux 1, les versions plus anciennes connaissent le même problème)

FreeBSD

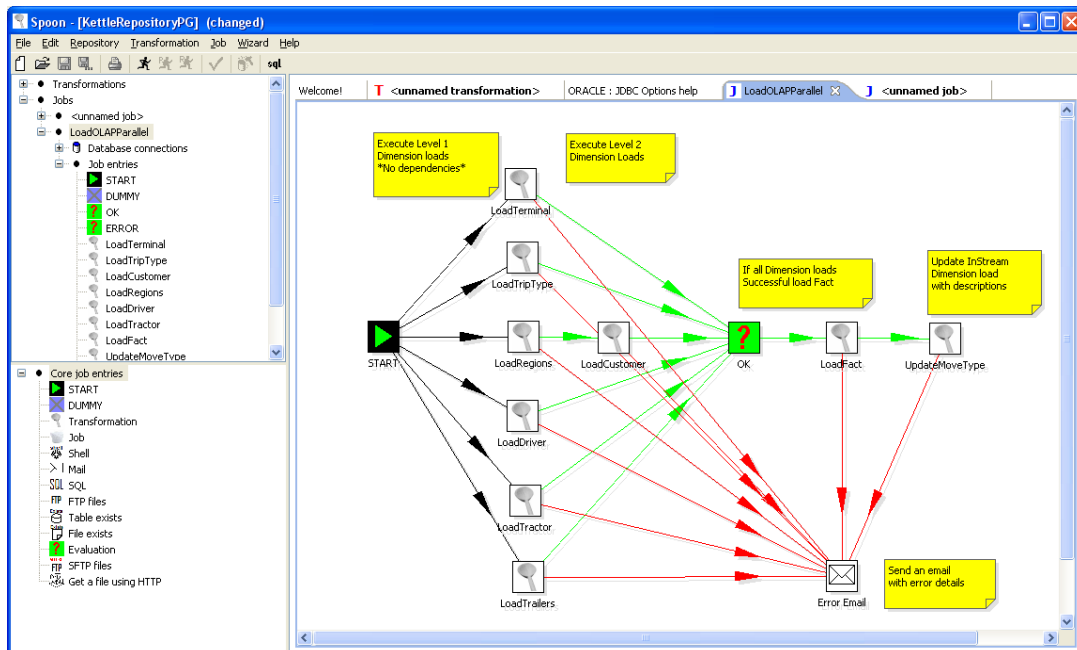
Des Problèmes avec le « Glisser-Déposer » (en anglais drag and drop). Pour se sortir d'affaire, utilisez le clic droit sur la grille afin d'insérer une nouvelle étape ou tâche.

Merci de regarder la liste de diffusion <http://kettle.javaforge.com> pour les informations les plus récentes sur la résolution des problèmes.



2.6. Captures d'écran

Modéliser une transformation



Modéliser une tâche

Le volet supérieur gauche de Spoon permet d'explorer les connexions, les transformations et les tâches que vous avez ouvertes. Quand vous modélisez une transformation, le volet inférieur gauche contient toutes les étapes que vous pouvez utiliser pour vos transformations : extraction, alimentation, recherche de valeurs, transformations, écriture de script, etc... Quand vous modélisez une tâche, le volet inférieur gauche contient les noeuds possibles pour une tâche. Ces étapes et noeuds sont décrits en détail dans le chapitre :

[Database Connections](#), [Hops](#), [Transformation Steps](#), [Job Entries](#), [Graphical View](#).

2.7. Options de la ligne de commande

Notes importantes :

- *Sur les plates-formes Windows, l'utilisation du signe moins ("-") dans les options peut poser problème, de même que le signe d'égalité ("="). A cause de cela, depuis la version 2, vous pouvez utiliser ce format ou n'importe quelle combinaison de /,- and ;,=*
- *Les champs en italique décrivent les valeurs que les options peuvent prendre*
- *Il est important, quand il y a des des espaces dans les valeurs des options, d'entourer de quotes (simples ou doubles) ces valeurs., comme c'est fait dans l'exemple ci-dessous.*
 - *Sur les plates-formes Windows, l'utilisation du signe moins ("-") dans les options peut poser problème, de même que le signe d'égalité ("="). A cause de cela, depuis la version 2, vous pouvez utiliser ce format ou n'importe quelle combinaison de /,- and ;,=*
 - *Les champs en italique décrivent les valeurs que les options peuvent prendre*
 - *Il est important, quand il y a des des espaces dans les valeurs des options, d'entourer de quotes (simples ou doubles) ces valeurs., comme c'est fait dans l'exemple ci-dessous.*

Voici les options en ligne de commande que vous pouvez utiliser :

```
-file=chemin du fichier XML
```

Cette option exécute la transformation définie dans le fichier XML, d'extension ktr. (.ktr : Kettle Transformation, kjb Kettle Job)

```
-logfile=chemin du fichier journal
```

Cette option permet de définir le chemin et le nom du fichier journal (Log), qui, par défaut est la sortie courante (standard output).

```
-level=Niveau de journalisation
```

Cette option permet de définir le niveau de journalisation de Spoon. Les valeurs possibles de ce niveau sont :

| | |
|------------------|---|
| Nothing: | ne journalise rien |
| Error: | ne journalise que les erreurs |
| Minimal: | journalise le minimum |
| Basic: | c'est le niveau par défaut de journalisation |
| Detailed: | c'est le niveau détaillé de journalisation |
| Debug: | à des fins de correction, la sortie est très détaillée |
| Rowlevel: | c'est le niveau le plus détaillé, au niveau de la ligne. Attention, ce niveau peut générer une grande quantité de données |

```
-rep=nom du référentiel
```

Définit le référentiel auquel se connecter. Vous avez besoin alors également de spécifier les options `-user`, `-pass` and `-trans`. La configuration du référentiel est chargée depuis un fichier xml « repositories.xml » qui se trouve dans le répertoire local ou le répertoire de Kettle : `$HOME/.kettle/` sous UN*X ou `C:\Documents and Settings\<username>\.kettle` sous Windows. (ndt vous pouvez aussi définir cette option dans une variable d'environnement `KETTLE_REPOSITORY`).

```
-user=Utilisateur
```

Définit le nom d'utilisateur pour se connecter au référentiel. (ndt vous pouvez aussi définir cette option dans une variable d'environnement `KETTLE_USER`).

```
-pass=Mot de passe
```

Définit le mot de passe pour se connecter au référentiel. (ndt vous pouvez aussi définir cette option dans une variable d'environnement `KETTLE_PASSWORD`).

```
-trans=nom de la transformation
```

Utilisez cette option pour définir la transformation à exécuter dans le référentiel.

```
-job=nom de la tâche
```

Utilisez cette option pour définir la tâche à exécuter dans le référentiel.

NOTE: Il est important, si vous avez utilisé des espaces dans le nom de vos transformations ou tâches de bien entourer ce nom d'apostrophe (quotes) ou de guillemet (*double quotes*).

2.8. Référentiel

Un référentiel Kettle peut contenir entre autres des transformations, des tâches. Pour charger une transformation ou une tâche depuis un référentiel, il vous faut tout d'abord vous connecter à ce référentiel.

Pour réaliser cela, vous devez définir une connexion à la base de données contenant le référentiel. Vous pouvez faire cela dans le dialogue au départ de Spoon (ndt ou en utilisant le menu Référentiel -Se connecter au référentiel, raccourci CTRL-R).



Les informations concernant les référentiels sont enregistrées dans un fichier XML "repositories.xml". Le fichier se trouve dans un répertoire (caché sous Un*x) ".kettle" dans votre répertoire personnel. En environnement Windows, c'est généralement C:\Documents and Settings\<username>\.kettle.

Suggestion: le chemin complet de ce fichier est paramétrable dans la console de Spoon.

Si vous ne désirez pas avoir cette boîte de dialogue, vous pouvez désactiver cette option dans le menu Éditer - Options. Voir aussi [Options](#).

IMPORTANT: Si vous utilisez la création d'un référentiel par Spoon, le mot de passe par défaut pour l'administrateur est « admin » (ndt il existe également un compte « guest » avec mot de passe à « guest » aux droits très restreints). Vous devriez changer ce mot de passe juste après la création de votre référentiel en utilisant l'explorateur de référentiel intégré à Spoon.

2.9. Connexion automatique au référentiel

Spoon peut se connecter automatiquement à un référentiel en renseignant les variables d'environnement suivantes : KETTLE_REPOSITORY, KETTLE_USER et KETTLE_PASSWORD.

Cela vous assure de toujours vous connecter au même référentiel. **Veillez noter que cela induit un risque en terme de sécurité et que vous devriez notamment toujours vous prémunir d'un accès non autorisé au référentiel depuis votre ordinateur.**

2.10. License

Depuis la version 0, Kettle est publiée sous la licence LGPL. Veuillez vous référer à l'annexe A pour le texte complet de cette licence (en anglais)

Note: *Pentaho Data Integration est nommé "Kettle" ci-dessous.*

Copyright (C) 2006 Pentaho Corporation

Kettle is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 1 of the License, or (at your option) any later version.

Kettle is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public License along with the Kettle distribution; if not, write to the Free Software Foundation, Inc., 51 Franklin St, Fifth Floor, Boston, MA 02110-1301 USA

2.11. Notes

Vous pouvez insérer une note où vous le désirez sur la vue graphique simplement en appelant le menu contextuel de la grille (généralement clic droit) et en choisissant l'item "Ajouter une note". La note ainsi créée peut être éditée en double-cliquant dessus, déplacée naturellement avec la souris. Pour supprimer une note, vous avez dans le menu contextuel de la note un item "Supprimer la note".

2.12. Définitions

2.12.1. Définitions autour de la notion de transformation

Valeur (en anglais Value): Les valeurs sont les parties distinctes d'une ligne et peuvent contenir n'importe quel type de données : chaînes de caractères, nombre à virgule flottante, grand entier à précision illimitée, dates, booléens (vrai/faux).

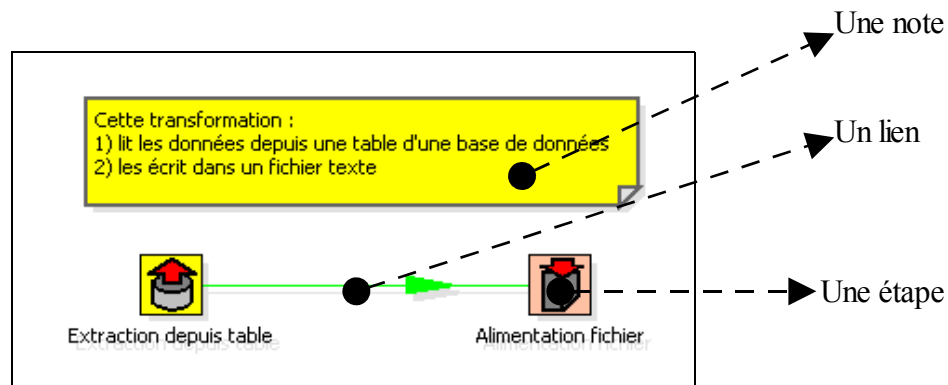
Ligne (en anglais Row): une ligne est une liste ordonnée de 0 à plusieurs valeurs

Flux de sortie (en anglais Output stream): un flux de sortie est une pile de lignes qui quitte une étape.

Flux d'entrée (en anglais Input stream): un flux d'entrée est une pile de lignes qui entre dans une étape.

Lien (en anglais Hop): un lien (ndt la traduction littérale serait « saut ») est une représentation graphique de un ou plusieurs flux de données entre deux étapes. Un lien représente toujours un flux de sortie pour une étape et un flux d'entrée pour une autre étape. **The number of streams is equal to the copies of the destination step.** (1 or more)

Note (en anglais note): une note est une information que l'on peut ajouter à une transformation.

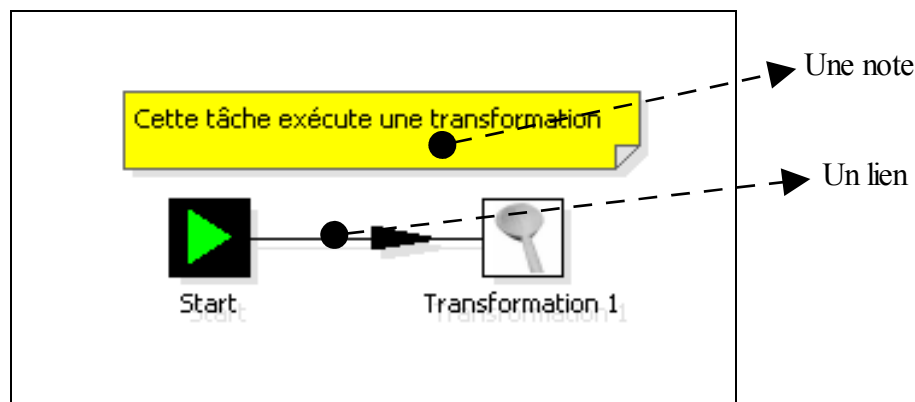


2.12.2. Définitions autour du concept de tâche

(en anglais Job Entry): A job entry is one part of a job and performs a certain










(en anglais Hop): un lien (ndt la traduction littérale serait « saut ») est une représentation graphique de un ou plusieurs flux entre deux étapes. Un lien représente toujours un lien entre deux job entries et flux de sortie pour une étape et un flux d'entrée pour une autre étape. A hop is a graphical representation of one or more data streams between 2 steps. A hop always represents the link between two job entries and can be set (depending on the type of originating job entry) to execute the next job entry unconditionally, after successful execution or failed execution.

(en anglais Note): une note est une information que l'on peut ajouter à une tâche.



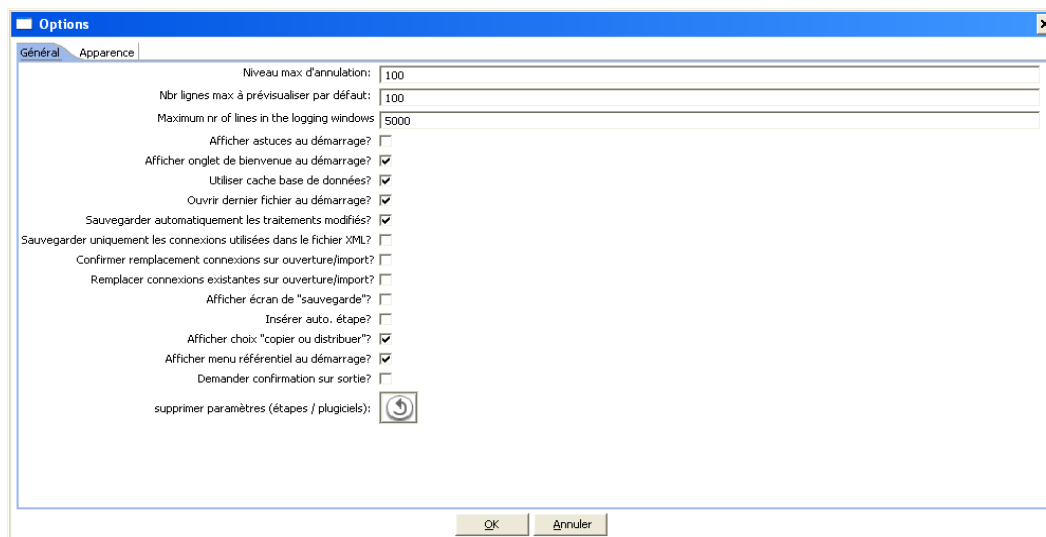
Barre d'outils

Les icônes de la barre d'outils de l'écran principal sont, de gauche à droite :

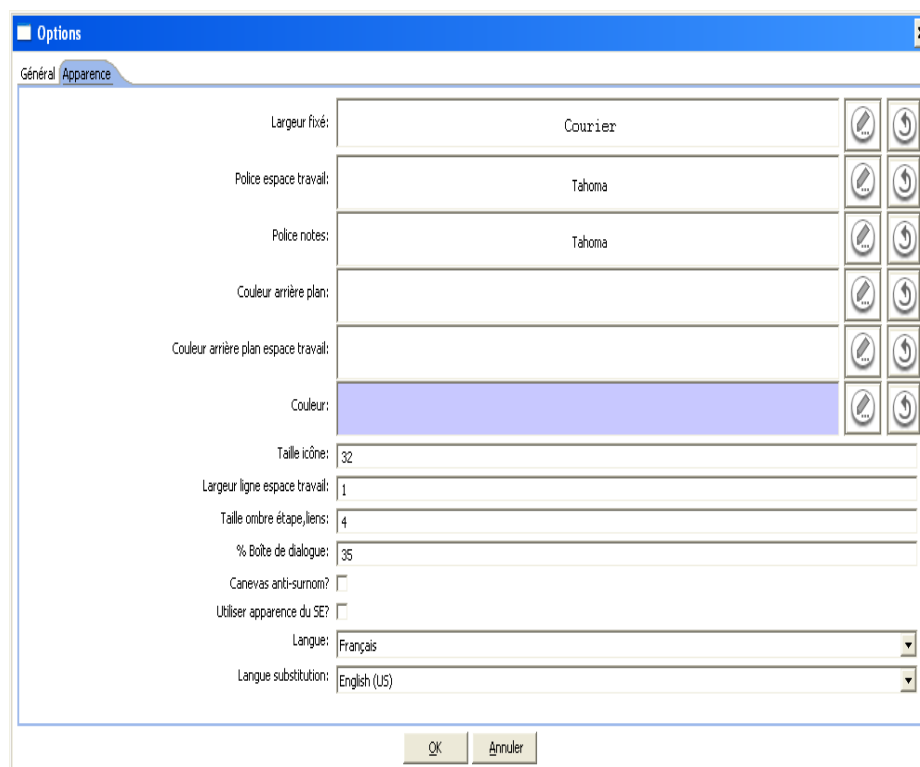
| Icône | Signification |
|---|---|
|  | Crée une nouvelle transformation ou une nouvelle tâche |
|  | Ouvre une transformation ou une tâche depuis le système de fichier si vous n'êtes pas connecté à un référentiel ou depuis le référentiel si vous êtes connecté. |
|  | Sauve la transformation ou la tâche dans un fichier ou le référentiel. |
|  | Sauve la transformation ou la tâche sous un nom différent dans le référentiel ou sous un nom de fichier différent. |
|  | Imprime : un dialogue vous permet de préciser le nombre de pages, l'imprimante, les marges etc. |
|  | Exécute la transformation ou tâche courante depuis le fichier source ou le référentiel, sans tenir compte des modifications non enregistrées. |
|  | Prévisualise la transformation ou tâche courante en mémoire en prenant en compte les éventuelles modifications non enregistrées. Vous pouvez prévisualiser les lignes produites dans les étapes sélectionnées. |
|  | Exécute de nouveau le déroulement de la transformation pour une certaine date et heure. Cette manoeuvre permet de rejouer certaines étapes (fichier texte ou excel en entrée) sur les seules lignes qui ont échouées à être interprétées pour une date et une heure particulière. |
|  | Vérifie une transformation: Spoon exécute une série de vérifications à chacune des étapes afin de voir si tout se passe comme cela est demandé. |
|  | Exécute une analyse d'impact : quel impact la transformation aura dans les bases utilisées. |
|  | Produit le code SQL nécessaire pour exécuter la transformation chargée. |

2.13. Options

Vous pouvez changer un certain nombre de préférences afin de personnaliser l'interface graphique, comme les polices de caractères, la couleur des écrans. Pour changer ces préférences, veuillez aller dans le menu « Édition » puis « Options... ». Vous trouvez ci-dessous quelques captures d'écrans afin de décrire les différentes préférences.



onglet "général"



onglet "apparence"

2.13.1. Options : l'onglet « général »

2.13.1.1. Nombre maximal d'annulations possibles

Ce paramètre permet de définir le nombre maximal d'annulations possibles. Le même nombre est utilisé pour la restauration.

2.13.1.2. Taille de prévisualisation par défaut

Ce paramètre permet de définir le nombre de lignes par défaut qui sera utilisée pour chaque étape lors d'une prévisualisation.

2.13.1.3. Montrer les astuces au démarrage ?

Ce paramètre permet de définir si oui ou non vous voulez que Spoon vous montre une astuce au démarrage de l'interface.

2.13.1.4. Montrer l'écran de bienvenue au démarrage ?

Ce paramètre permet de définir si oui ou non vous voulez que Spoon vous montre l'écran de bienvenue au démarrage de l'interface.

2.13.1.5. Utiliser le cache des données ?

Spoon conserve l'information qui est enregistrée dans une base de données source ou cible. Dans certains cas, cela peut conduire à des résultats incorrects, surtout si celles-ci sont changées souvent au cours du processus. Dans ces cas là, il est possible de désactiver cette fonctionnalité de conservation (cache) au lieu d'effacer le cache à chaque opération.

NOTE: *Spoon efface automatiquement le cache de la base de données quand vous lancez un état DDL. Cependant, si vous utilisez un outils tiers, vous aurez peut-être besoin d'effacer le cache manuellement.*

2.13.1.6. Ouvrir le dernier fichier au démarrage ?

Activer ce paramètre permet de charger automatiquement (ou essayer de le faire) la dernière transformation ou tâche utilisée (ouverte ou enregistrée) depuis un fichier XML ou un référentiel.

2.13.1.7. Enregistrer automatiquement les changements avant l'exécution.

Ce paramètre permet de définir si oui ou non Spoon enregistre automatiquement la transformation ou la tâche avant de l'exécuter.

2.13.1.8. Sauvegarder seulement les connexions utilisées dans le fichier XML.

Ce paramètre permet de limiter l'enregistrement des seules connexions utilisée au sein du XML. Vous devez désactiver ce paramètres si vous souhaitez conserver toutes les connexions définies et non seulement celles utilisées.

2.13.1.9. Demander si spoon remplace les connexions existantes lors de l'ouverture/import.

Ce paramètre permet de définir si spoon demande à l'utilisateur avant de remplacer les connexions existantes lors de l'import.

2.13.1.10. Remplacer les connexions existantes lors de l'ouverture/import

Ce paramètre permet de définir l'action qui sera réalisée lorsqu'aucun dialogue ne sera affiché (voire l'option précédente).

2.13.1.11. Montrer le dialogue de confirmation avant la sauvegarde

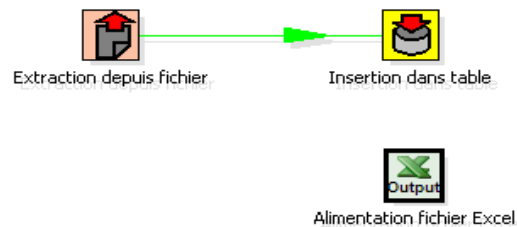
Ce paramètre permet de définir si oui ou non spoon affiche un dialogue de confirmation avant la sauvegarde.

2.13.1.12. Séparation automatique des liens.

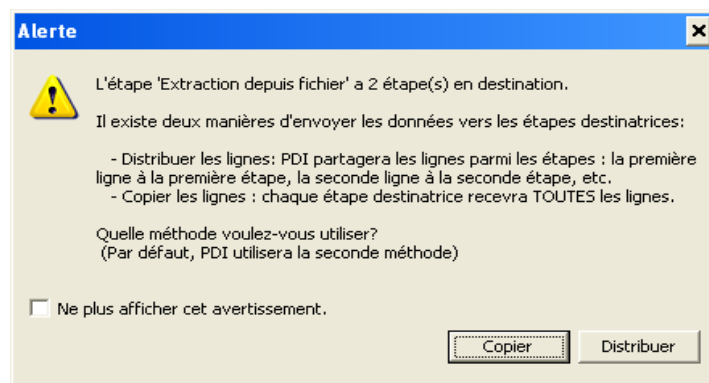
Ce paramètre définit si oui ou non spoon affiche un dialogue de confirmation avant de séparer un lien. (voir aussi [Séparer un lien](#))

2.13.1.13. Montrer l'avertissement « Copie ou répartition) ?

Ce paramètre permet de définir si spoon affiche ou non le message d'avertissement qui apparaît quand vous liez une étapes avec plusieurs sorties. Par exemple, si vous disposez d'une transformation comme celle ci-dessous :



Lorsque vous liez l'étape « Extraction depuis fichier » à « Alimentation fichier Excel », spoon affiche l'avertissement suivant :



2.13.1.14. Montrer au démarrage la boîte de dialogue de choix du référentiel

Ce paramètre définit si oui ou non spoon affiche la boîte de dialogue de choix du référentiel au démarrage.

2.13.1.15. Effacer les paramètres personnalisés ?

Ce bouton permet d'effacer tous les paramètres personnalisés qui ont été modifiés dans les boîtes de dialogues « plugin » ou « étapes ».

2.13.2. Onglet « apparence »

2.13.2.1. Police de caractères par défaut

Vous pouvez définir ici la police de caractères utilisée par défaut dans les boîtes de dialogues, les arbres, les champs etc.

2.13.2.2. Police de caractères de la représentation graphique

Vous pouvez définir ici la police de caractères utilisée pour la représentation graphique.

2.13.2.3. Police de caractères des tables

Vous pouvez définir ici la police de caractères utilisée pour les tables.

2.13.2.4. Police de caractères des notes

Vous pouvez définir ici la police de caractères utilisée pour les notes affichées dans la représentation graphique.

2.13.2.5. Couleur de fond

Vous pouvez définir ici la couleur de fond dans spoon. Cela affecte aussi les boîtes de dialogue.

2.13.2.6. Couleur de fond de la représentation graphique

Vous pouvez définir ici la couleur de fond de la représentation graphique de spoon.

2.13.2.7. Couleur des onglets sélectionnés

Vous pouvez définir ici la couleur des onglets sélectionnés ou activés.

2.13.2.8. Taille des icônes

Vous pouvez définir ici la taille des icône dans la représentation graphique. La taille par défaut d'une icône est de 32x32 pixels. Les meilleurs résultats (graphiques) seront probablement obtenus pour des tailles de 16,24,32,48,64 ou tout autres multiples de

2.13.2.9. Largeur de ligne

Vous pouvez définir ici la largeur des lignes représentant les liens et la bordure des cadres représentant les étapes.

2.13.2.10. Taille de l'ombre

Vous pouvez définir ici la distance entre une représentation graphique et son ombre . Si cette mesure est supérieure à 0, une ombre est dessinée sous les étapes, les liens, les notes, rendant une impression de relief.

2.13.2.11. Largeur de la colonnes des paramètres

Par défaut, au sein de la boîte de dialogue des préférences, la colonne de gauche contenant les paramètres est taillée à 35% de la largeur totale de la boîte de dialogue. Vous pouvez changer ce paramètre. Peut-être cela vous sera-t-il utile si vous utilisez des grandes polices de caractères.

2.13.2.12. Anticrênelage (en anglais anti-aliasing) activé ?

Quelques plates-formes comme Windows, OSX et Linux supporte l'anticrênelage grâce à des outils comme GDI, Carbon ou Cairo. Vous pouvez utiliser ce paramètre pour adoucir vos icônes et vos images.

2.13.2.13. Récupérer l'apparence du système d'exploitation ?

Dans l'environnement windows, vous pouvez activer ce paramètre si vous désirez que les fenêtres récupèrent les paramètres par défaut du système de fenêtrage. Pour les autres plates-formes, c'est toujours le cas.

2.13.2.14. Langage de l'interface

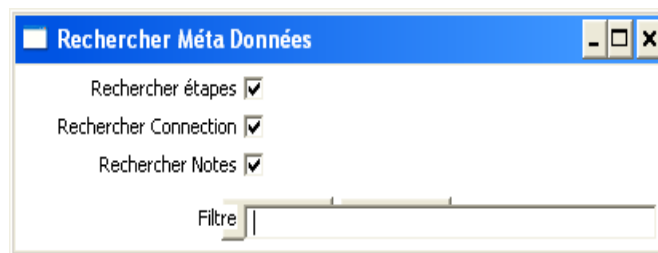
Vous pouvez définir ici le langage de l'interface. Si un libellé n'a pas été traduit dans le langage demandé, spoon utilise le libellé du langage défini par défaut.

2.13.2.15. Langage par défaut

Parce que spoon est conçu en anglais, ce paramètre devrait toujours être défini à « English ».

2.14. Rechercher

Cette fonctionnalité se trouve dans le menu « Édition » ou en utilisant le raccourci CTRL+F.



boîte de dialogue "Rechercher méta-données"

Vous pouvez rechercher une chaîne de caractères dans tous les champs, les connexions, les notes, les transformations et tâches chargées. Le résultat est détaillé au sein d'une table.

2.15. Variables d'environnement

La possibilité de modifier une variable d'environnement a été ajoutée car c'était le moyen le plus facile de tester les transformations qui utilisent des variables dynamiques. Normalement, ces variables sont modifiées par une transformation dans une tâche. Cependant, durant le développement et le test, vous pouvez vouloir les modifier manuellement.

Cette fonctionnalité est accessible depuis le menu « Édition » ainsi qu'avec le raccourci CTRL-ALT-J. L'écran affiché est le même que celui que vous aurez si vous utilisez une transformation avec des variable non définies. Vous pourrez ainsi les définir avant l'exécution.

2.16. Historique du journal d'exécution

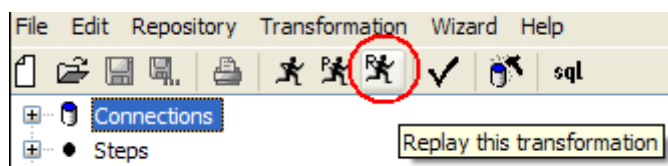
Vous pouvez voir toutes les informations d'exécution journalisées sur une transformation ou une tâche. Pour ce faire, sélectionnez la transformation ou l'étape voulue, et en faisant un clic droit sur celle-ci, choisissez « Ouvrir l'onglet historique ». Un onglet apparaît alors dans la vue de droite.

| | Batch ID | Status | Read | Written | Updated | Input | Output | Errors | Start date | End date |
|---|----------|--------|------|---------|---------|-------|--------|--------|-------------------------|-------------------------|
| 1 | 2 | end | 0 | 0 | 0 | 0 | 0 | 0 | 1899/12/31 18:00:00.000 | 2007/01/18 15:24:24.046 |
| 2 | 2 | start | 0 | 0 | 0 | 0 | 0 | 0 | 1899/12/31 18:00:00.000 | 2007/01/18 15:24:24.046 |

Onglet historique de la transformation

2.17. Rejouer

La fonctionnalité « Rejouer » vous permet d'exécuter de nouveau une transformation qui a échouée. Elle a été implémentée pour l'extraction depuis un fichier texte ou excel afin de permettre le renvoi du fichier s'il provoque des erreurs à l'intégration.



« rejouer »

Seules les lignes qui ont échouées seront intégrées dans la nouvelle exécution. Si une ligne est présente dans le fichier, spoon utilise la date du fichier pour comparer les dates lors de la nouvelle exécution.

2.18. Réaliser une correspondance de champs

Si vous disposez d'une table fixe dans l'étape, vous voulez peut-être faire correspondre des champs dans le flux d'entrée avec les champs de cette table. Pour ce faire, un clic droit sur la table vous ouvre un menu contextuel dans lequel vous trouverez l'item « Faire une sélection des champs de sortie ». Une boîte de dialogue vous aide alors à déterminer quels sont les champs en entrée que vous voulez faire correspondre aux champs du flux en sortie :

La sélection des champs en sortie est réalisée de façon semi-automatique, en se fondant en partir sur le nom du champ. Comme résultat de ce dialogue, une nouvelle étape

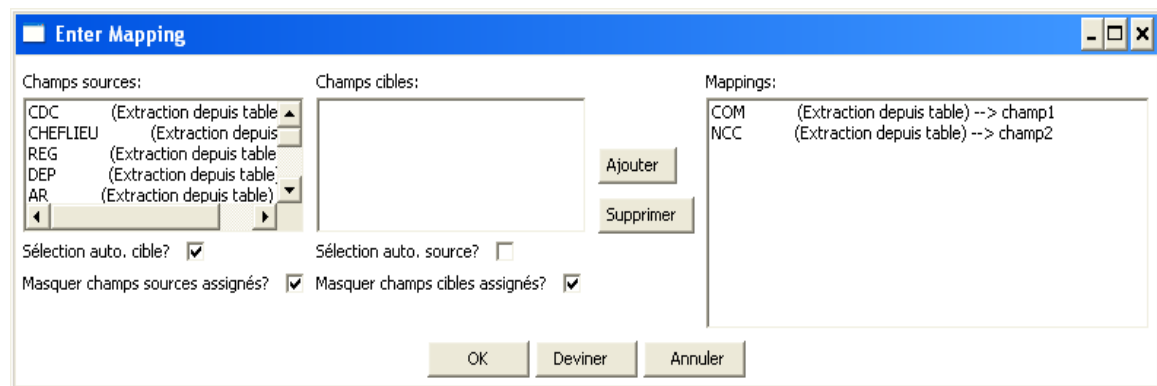
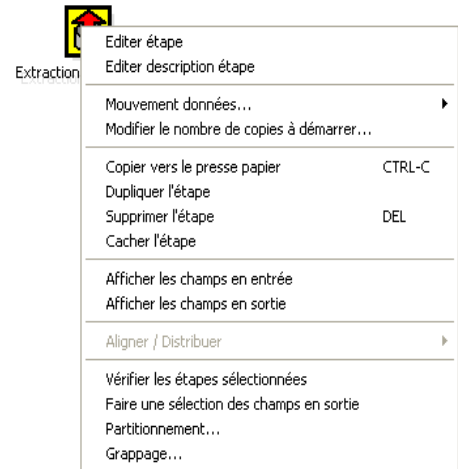
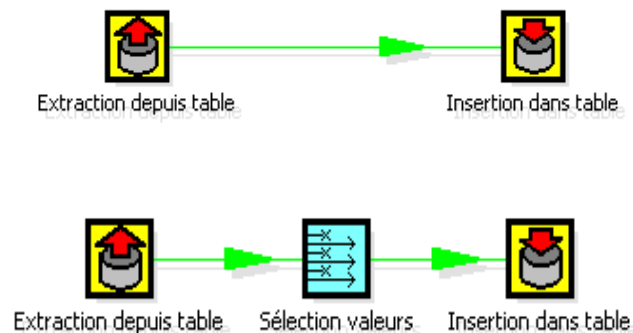


Illustration 1: boîte de dialogue de correspondance des champs en entrée et sortie

« Sélection de valeurs » est générée. : Spoon place alors simplement cette étape entre l'entrée et la sortie et la correspondance est ainsi réalisée.



2.19. Mode sécurisé dit de « contrôle renforcé »

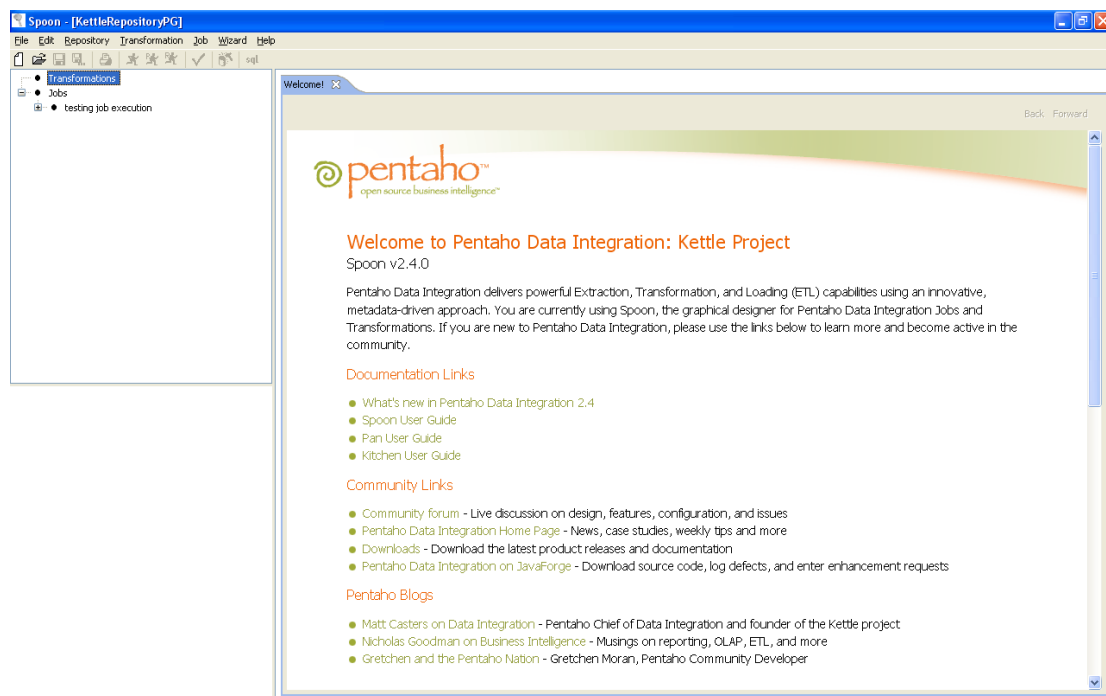
Lorsque vous mélangez des lignes de plusieurs sources, vous devez vous assurer que ces lignes ont le même gabarit dans toutes les conditions. Pour cela, nous avons ajouté une « mode de contrôle renforcé » que vous pouvez activer (case à cocher en bas à droite) dans la fenêtre « Journal » (en anglais Log) avant de démarrer votre transformation. A l'exécution, la transformation vérifiera toutes les lignes unes à unes afin de déterminer si le gabarit de chacune est identique.

Dans le cas où une ligne est trouvée avec un gabarit différent de celui de la première ligne, une erreur est générée avec la ligne incriminée.

Note: Cette possibilité est aussi possible dans Pan avec l'option « *safemode* ».

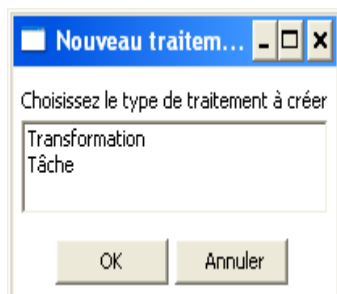
2.20. Onglet de bienvenue

L'onglet de bienvenue est affiché la première fois que vous lancez Spoon. Il vous fournit des liens supplémentaires vers le produit Pentaho Data Integration. Vous pouvez désactiver le lancement de cet écran en utilisant les options (Edition|Options). Pour l'afficher de nouveau si vous le fermez, utilisez le menu



(Aide|Afficher l'onglet de bienvenue)



3. Créer une transformation ou une tâche



Pour créer une nouvelle transformation ou une tâche, vous pouvez :

- soit cliquer sur le bouton « Nouveau » dans la barre d'outils,
- soit sélectionner l'item Fichier|Nouveau dans le menu

Après avoir sélectionné le type de fichier que vous désirez créer, un nouvel onglet s'ouvre montrant une représentation graphique de votre transformation ou votre tâche (Voir [#13 Représentation graphique](#) pour plus de détails).

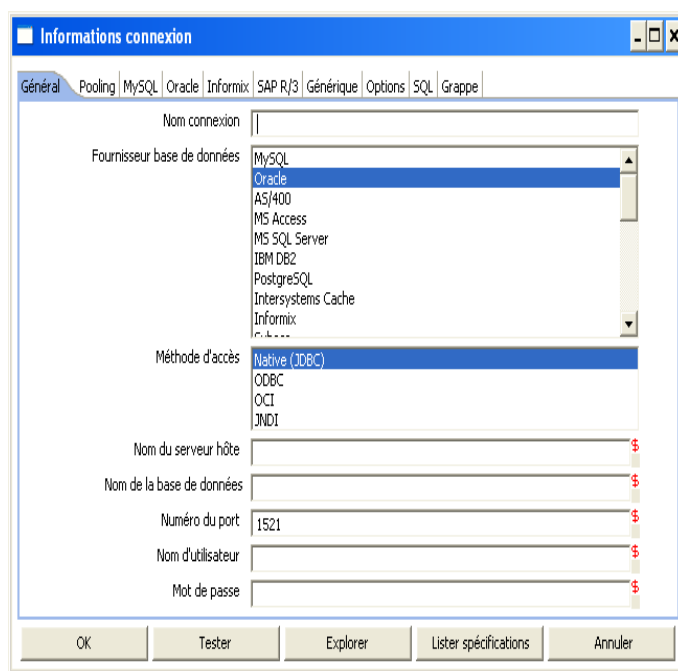
| Icône | Description |
|---|---|
|  | Représentation graphique d'une tâche (ndt J comme Job en anglais) |
|  | Représentation graphique d'une transformation (ndt T comme Transformation en anglais) |

4. Connexion aux bases de données

4.1. Description

Une connexion décrit la méthode utilisée par Kettle pour se connecter à une base de données. La première entrée du menu à gauche recense les connexions possibles.

4.2. Capture d'écran



4.3. Options

Nom connexion : vous devez donner un nom unique pour définir chacune des connexions entre la transformation ou la tâche et la base de données.

Fournisseur base de données : vous devez renseigner le type de base de données à laquelle vous vous connectez.

Méthode d'accès : elle peut être native (JDBC), ODBC, OCI ou définie par une JNDI (Java Naming and Directory Interface).

Nom du serveur hôte : vous spécifiez ici le nom du serveur de base de données ou son adresse IP.

Nom de la base de données : ce paramètre est le nom de la base de données à laquelle vous désirez vous connecter. Dans le cas d'ODBC, spécifiez ici le nom DSN (voir aussi [#4 4 Liste des bases de données supportées](#)).

Numéro du port: mettez ici le port TCP/IP écouté par votre serveur de base de données.

Utilisateur/mot de passe: Vous pouvez utiliser ces deux paramètres pour renseigner le nom de l'utilisateur et le mot de passe pour se connecter à la base de données. Ces paramètres sont optionnels.

EXTRA:

- Pour Oracle, vous pouvez spécifier l'unité logique par défaut (tablespaces) dans laquelle Kettle placera les objets SQL pour les tables et index.
- Pour Informix, vous aurez besoin de spécifier le nom du serveur dans l'onglet « Informix » pour pouvoir utiliser la connexion.
- Pour SAP R/3, Les paramètres de langues, de nombre système et de client SAP peuvent être précisé dans l'onglet SAP R/3 tab.
- Liste des spécifications: liste les paramètres de la connexion à la base de données comme l'url JDBC URL, les classes utilisées, la liste des mots réservés.
- Onglet options: cet onglet vous permet d'ajouter des paramètres spécifiques pour votre connexion. (ndt Un bouton d'aide affiche l'aide en ligne depuis le site de l'éditeur de la base de données).

4.4. Liste des bases de données supportées

| Base de données | Méthode d'accès | Nom d'hôte /IP | Nom de la base de données | Numéro de port (défaut) | Utilisateur et mot de passe |
|--------------------|-----------------|----------------|---------------------------|-------------------------|-----------------------------|
| Oracle | Native | requis | SID Oracle | requis (1521) | requis |
| | ODBC | | DSN ODBC | | requis |
| | OCI | | Nom TNS | | requis |
| MySQL | Native | requis | Nom MySQL | requis (3306) | requis |
| | ODBC | | DSN ODBC | | requis |
| AS/400 | Native | requis | Nom librairie AS/400 | optionnel | requis |
| | ODBC | | DSN ODBC | | requis |
| MS Access | ODBC | | DSN ODBC | | optionnel |
| MS SQL Server | Native | requis | Nom | requis (1433) | requis |
| | ODBC | | DSN ODBC | | requis |
| IBM DB2 | Native | requis | Nom | requis (50000) | requis |
| | ODBC | | DSN ODBC | | requis |
| PostgreSQL | Native | requis | Nom | requis (5432) | requis |
| | ODBC | | DSN ODBC | | requis |
| InterSystems Caché | Native | requis | Nom | requis (1972) | requis |
| | ODBC | | DSN ODBC | | requis |
| Sybase | Native | requis | Nom | requis (5001) | requis |

| Base de données | Méthode d'accès | Nom d'hôte / IP | Nom de la base de données | Numéro de port (défaut) | Utilisateur et mot de passe |
|-------------------|-----------------|-----------------|---------------------------|-------------------------|-----------------------------|
| | ODBC | | DSN ODBC | | requis |
| Gupta SQL Base | Native | requis | Nom | requis (2155) | requis |
| | ODBC | | DSN ODBC | | requis |
| Dbase III,IV or 0 | ODBC | | DSN ODBC | | optionnel |
| Firebird SQL | Native | requis | Nom | requis (3050) | requis |
| | ODBC | | DSN ODBC | | requis |
| Hypersonic | Native | requis | Nom | requis (9001) | requis |
| MaxDB (SAP DB) | Native | requis | Database name | | requis |
| | ODBC | | DSN ODBC | | requis |
| CA Ingres | Native | requis | Nom | | requis |
| | ODBC | | DSN ODBC | | requis |
| Borland Interbase | Native | requis | Nom | requis (3050) | requis |
| | ODBC | | DSN ODBC | | requis |
| ExtenDB | Native | requis | Nom | requis (6453) | requis |
| | ODBC | | DSN ODBC | | requis |
| Générique* | Native | requis | Nom | requis (Tous) | requis |
| | ODBC | | DSN ODBC | | optionnel |

(*) La connexion générique nécessite également de spécifier l'URL et le pilote de données dans l'onglet « Générique ».

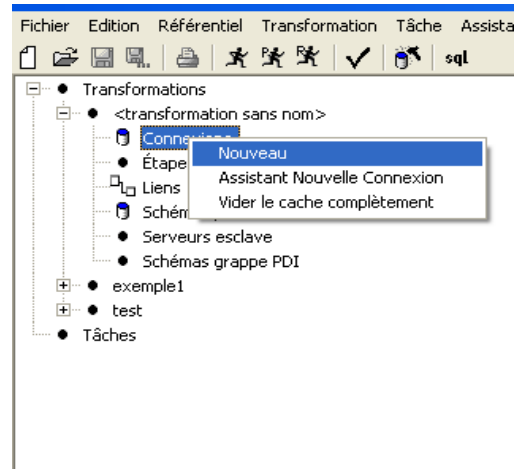
4.5. Utilisation

4.5.1. Créer une nouvelle connexion

Vous pouvez créer une nouvelle connexion avec un clic droit sur l'item « Connexion » dans la vue en haut à gauche et en sélectionnant « Nouveau ».

Ou : simplement double-cliquez sur l'entrée « Connexions » du menu

Ou : pressez la touche F3 pour démarrer l'assistant « Nouvelle connexion ».



4.5.1.1. Onglet Général

L'onglet général est l'endroit où vous renseignez les informations générales de votre connexion comme le nom de la connexion, le type, la méthode d'accès, le nom du serveur de bases de données, le nom de la base de données ainsi que les informations d'authentification.

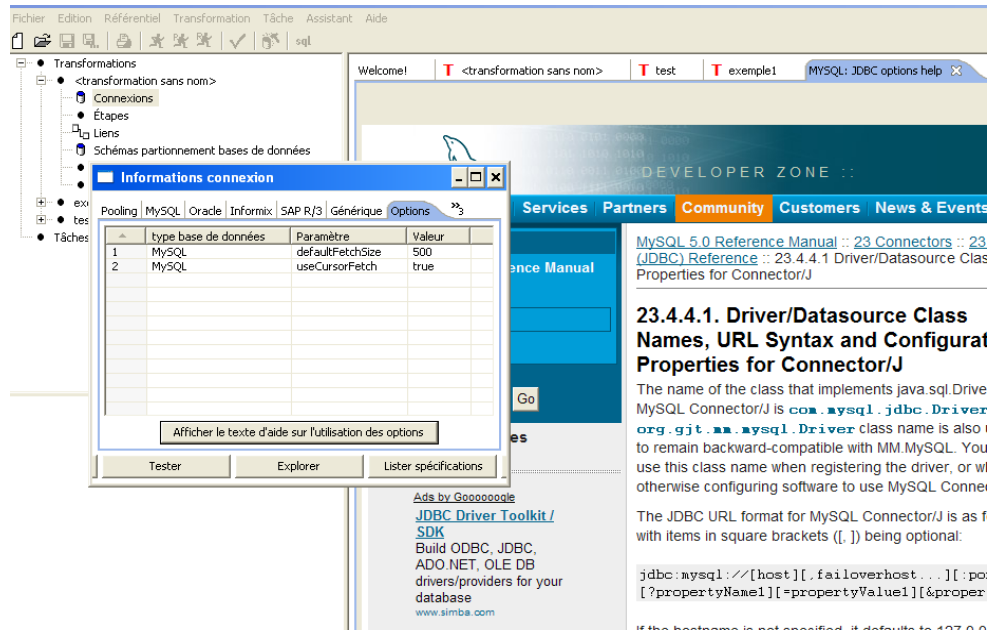
A screenshot of the 'Informations connexion' dialog box, specifically the 'Général' tab. The dialog has several input fields and dropdown menus. The 'Nom connexion' field is empty. The 'Fournisseur base de données' dropdown is set to 'MySQL'. The 'Méthode d'accès' dropdown is set to 'Native (JDBC)'. The 'Nom du serveur hôte' field is empty. The 'Nom de la base de données' field is empty. The 'Numéro du port' field contains '1521'. The 'Nom d'utilisateur' field is empty. The 'Mot de passe' field is empty. At the bottom, there are buttons for 'OK', 'Tester', 'Explorer', 'Lister spécifications', and 'Annuler'.

4.5.1.2. Onglet « Pooling »

L'onglet pooling vous permet de configurer votre connexion afin qu'elle utilise les fonctionnalités de « pooling » et définir les paramètres qu'elle utilisera (taille initiale, taille maximum).

4.5.1.3. Onglet Options

Cet onglet vous permet d'ajouter des paramètres spécifiques pour votre connexion. Un bouton d'aide affiche



l'aide en ligne depuis le site de l'éditeur de la base de données.

4.5.1.4. Onglet SQL

Cet onglet vous permet de saisir des commandes SQL qui seront exécutées juste après la connexion à la base de données. Vous aurez peut-être besoin de cela pour des raisons de licences, configuration, traçage, journalisation etc.

4.5.1.5. Onglet Grappe (ndt en anglais Cluster)

Cet onglet vous permet d'activer l'exécution en grappe (ndt cluster en anglais) et de créer des connexions à des partitions de base de données. Pour créer une nouvelle partition de base de données, veuillez entrer un identifiant de partition et un nom d'hôte, un numéro de port, une base de données, un nom d'utilisateur et un mot de passe.

4.5.2. Modifier une connexion

Double-cliquez sur le nom de la connexion dans la vue de gauche pour la modifier. Vous pouvez aussi effectuer un clic droit et sélectionner « Éditer ».

4.5.3. Dupliquer une connexion

Effectuez un clic droit sur le nom de la connexion et sélectionnez « Dupliquer ».

4.5.4. Copier vers le presse papier

Effectuez un clic droit sur le nom de la connexion et sélectionnez « Copier vers le presse papier ». Cela copie le code XML décrivant la connexion vers le presse-papier.

4.5.5. Supprimer une connexion

Effectuez un clic droit sur le nom de la connexion et sélectionnez « Supprimer » afin de supprimer la connexion (ndt il n'y aura pas de confirmation mais vous pouvez annuler l'action édition|annulation ou CTRL+Z).

4.5.6. Tester une connexion

Dans la fenêtre d'édition de la connexion (voir ci-dessus), sélectionnez le bouton « Tester ». Si spoon arrive à se connecter, un message de succès d'affiche après un délai assez court.

4.5.7. Exécuter une instruction SQL

Effectuez un clic droit sur le nom de la connexion et sélectionnez « éditeur SQL ». Voir [#Éditeur SQL](#) pour plus de détails.

4.5.8. Vider le cache de la connexion

Pour accélérer les connexions, Spoon utilise un cache de base de données. Utilisez cette option si les informations dans le cache ne sont plus d'actualité. C'est le cas, par exemple, quand les tables de la bases de données ont été changées dans leur structure, créées ou supprimées.

4.5.9. Explorer

Cette option lance un explorateur de données pour la connexion sélectionnée. Voir [Explorateur de base de données](#) pour plus de détails.

4.6. Bases de données non supportées

Si vous voulez accéder à une base de données non supportées, faitesle nous savoir et nous essaierons de trouver une solution. Quelques bases de données ne sont pas supportées dans cette version en raison du manque d'utilisateurs.

Veuillez noter qu'il est peut-être possible de lire des données depuis ces bases de données en utilisant le pilote générique ou grâce à une connexion ODBC.

5. Éditeur SQL

5.1. Description

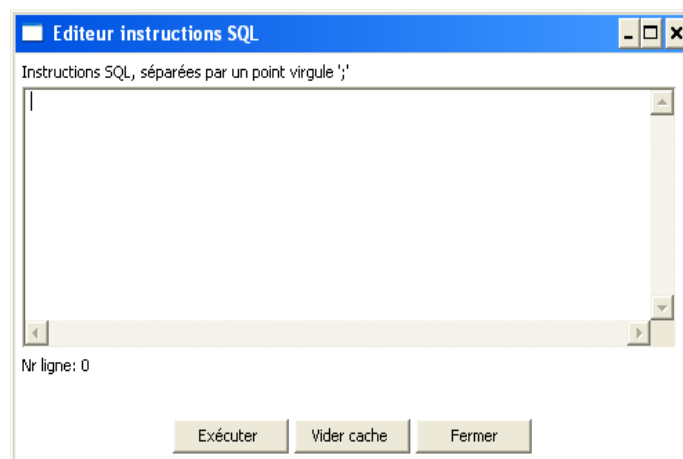
Il est parfois sympathique de disposer d'un éditeur SQL, spécialement quand vous voulez créer des tables. L'éditeur SQL fourni dans Spoon vous permet de faire ceci. En fait, vous pouvez même utiliser la plupart des instructions DDL (Data Definition Language) comme "create/alter table", "create index" et "create sequence". Les instructions SQL sont générées automatiquement pour vous dans la fenêtre de l'éditeur SQL.

NOTE: Les instructions SQL multiples doivent être séparées par un point-virgule (;).

NOTE: Avant d'envoyer à la base de données les instructions SQL, Spoon enlève les retour chariot, les retour à la ligne et les points-virgule de séparation.

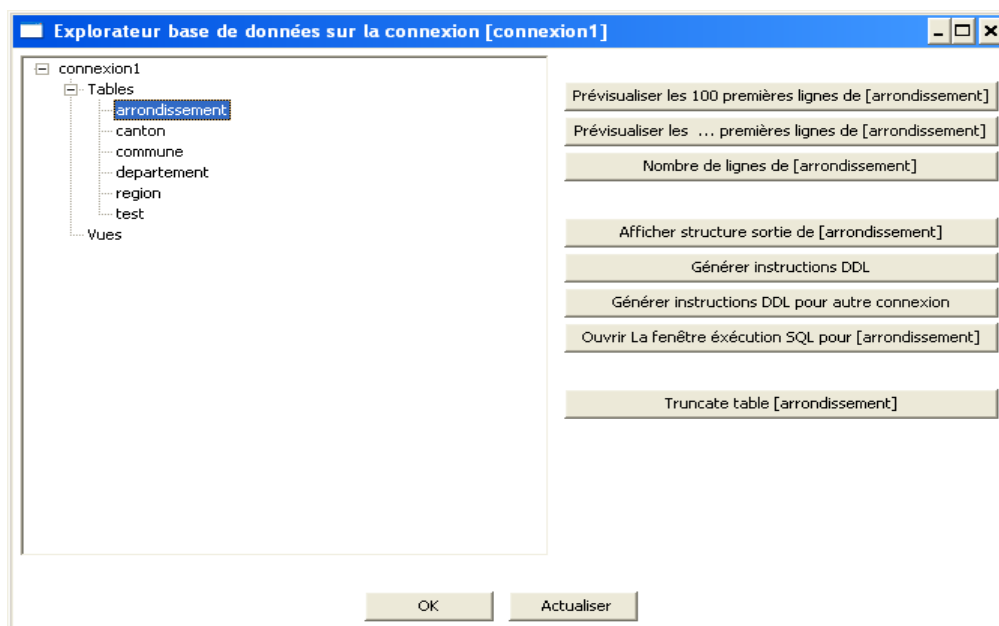
NOTE: Kettle nettoie le cache de connexion à chaque fois que vous envoyez des instructions DDL.

5.2. Capture d'écrans



6. Explorateur de base de données

6.1. Capture d'écran



6.2. Description

L'explorateur de base de données vous permet d'explorer la base de données assignée par votre connexion. Pour le moment, cet explorateur vous montre seulement les tables et les schéma/catalogues accessibles.

Il est possible avec un clic droit sur une table ou une vue d'effectuer les actions suivantes :

- Afficher les 100 première lignes de la table (action possible en double-cliquant sur la table))
- Afficher les n premières lignes de la table, en renseignant n
- Afficher le nom de lignes de la table
- Afficher la structure de la table
- Générer des instructions DDL (create table) pour cette table
- Générer des insctructions DDL pour cette table avec une autre connexion
- Ouvrir l'éditeur SQL pour cette table avec une instruction SQL (select * from table)

7. Liens

7.1. Description

Un lien relie une étape à une autre. La direction du flux de données est indiqué par une flèche sur la représentation graphique. Un lien peut être activé ou désactivé (pour des besoins de test, par exemple).

7.1.1. Liens dans le contexte « transformation »

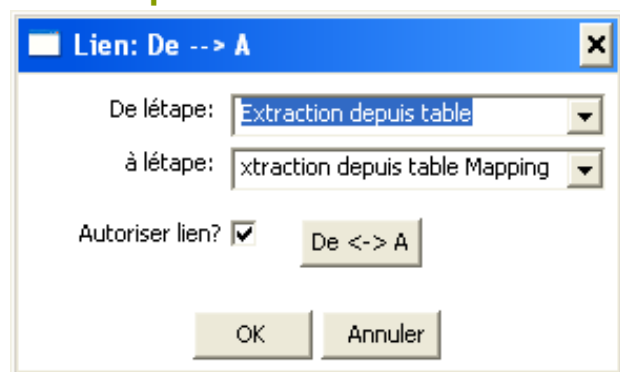
Lorsqu'un lien est désactivé au sein d'une transformation, l'étape cible du lien désactivé ne reçoit plus de flux de données en entrée. Cela peut donc conduire à des résultats erronés en aval du lien désactivé. Par exemple, vous n'aurez plus la possibilité de récupérer les champs en entrée au sein de l'étape en aval du lien tant que celui-ci sera désactivé.

7.1.2. Liens dans le contexte « tâche »

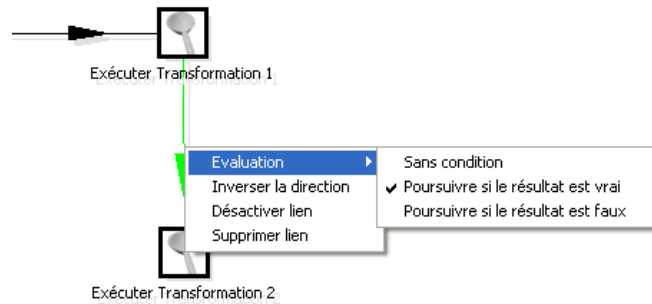
Le lien au sein d'une tâche précise la condition dans laquelle la tâche suivante doit être exécutée, ceci, quel que soit l'ordre de ces tâches. Vous pouvez définir les conditions pour enchaîner l'exécution de la prochaine tâche en cliquant droit sur le lien et en sélectionnant « Évaluation » :

- « Sans condition » signifie sur la tâche suivante sera exécutée sans se soucier du résultat de la tâche précédente.
- "Poursuivre si le résultat est vrai" signifie que la tâche suivante sera exécutée seulement si le résultat de la tâche précédente est vrai, ce qui veut dire que l'exécution a été un succès, le fichier a été trouvé, la table a été trouvée, sans erreur, sans évaluation à faux...
- "Poursuivre si le résultat est faux" signifie que la tâche suivante sera exécutée seulement si le résultat de la tâche précédente est faux, ce qui veut dire que l'exécution a été un échec, le fichier n'a pas été trouvé, la table n'a pas été trouvée, il y a eu des erreurs, l'évaluation est fausse...

7.2. Capture d'écrans



édition d'un lien au sein d'une transformation



édition d'un lien au sein d'une tâche

7.3. Créer un lien

Vous pouvez créer un nouveau lien entre deux étapes en utilisant l'une des méthodes suivantes

- Glisser la souris entre deux étapes en utilisant le bouton du milieu
- Glisser la souris entre deux étapes en maintenant la touche SHIFT (⇧) enfoncée et le clic gauche de la souris.
- Sélectionner deux étapes dans le menu, cliquer droit et sélectionner « nouveau lien »
- Sélectionner deux étapes dans la représentation graphique (CTRL + clic gauche), effectuer un clic droit sur la sélection et sélectionnez « Nouveau lien »

7.4. Séparer un lien

Vous pouvez insérer une nouvelle étape dans un lien en glissant cette étape sur le lien dans la représentation graphique. Le lien sera alors dessiné en gras. Lorsque vous lâchez le bouton gauche de la souris, un dialogue s'affiche vous demandant la confirmation de séparation du lien. Cette manipulation ne fonctionne qu'avec des étapes connectées ensemble.

7.5. Boucles

Les boucles ne sont pas autorisées dans les transformations. Spoon détermine les champs et les lignes de données en fonction des étapes précédentes. Des boucles pourraient donc être infinies ou donner des résultats indéfinis.

Les boucles **sont autorisées** dans les tâches parce que Spoon (anciennement Chef) exécute séquentiellement toutes les étapes. Faites attention que vos boucles ne soient pas sans fin. Pour cela, consultez le paragraphe . Une entrée d'étape « Créer évaluation » vous permet de sortir de ce type de cas en définissant un compteur (Voir [#Évaluation](#) Pour plus de détails).

8. Variables

8.1. Utilisation des variables

Des variables peuvent être utilisées tout au long de l'utilisation de Pentaho Data Integration, et notamment au sein de toutes les étapes des transformations et toutes les entrées de tâches. Les variables peuvent être définies en les paramétrant au sein d'une étape « Création variables » ou en les définissant dans le fichier `kettle.properties` dans le répertoire :

- `$HOME/.kettle` (Unix/Linux/OSX)
- `C:\Documents and Settings\<username>\.kettle\` (Windows)

Vous pouvez également les récupérer en utilisant [#Récupérer les variables](#) ou en définissant des chaînes méta-données comme ceci :

- `${VARIABLE}` or
- `%%VARIABLE%%`

Les deux formats peuvent être utilisés ensemble; le premier provient de l'environnement UNIX, le second de l'environnement Windows. Les boîtes de dialogue qui peuvent utiliser de telles variables sont repérées avec un signe dollar de couleur rouge :

Mot de passe \$

8.2. Portée des variables

La portée d'une variable dépend de l'endroit où elle est définie.

8.2.1. Variables d'environnement

La première utilisation de variables (et la seule dans les versions < 0) concernait les variables d'environnement. Ceci est rendu possible en passant à la machine virtuelle Java les variables avec l'option `-D`. C'est comme cela que l'on peut spécifier le chemin d'accès aux répertoires de façon indépendante de la plate-forme, par exemple en utilisant la variable `${java.io.tmpdir}`. Cette variable pointe vers le répertoire `/tmp` en environnement Unix/Linux/OSX et vers `C:\Documents and Settings\<username>\Local Settings\Temp` en environnement Windows.

Un des problèmes avec l'utilisation des variables d'environnement est que cette utilisation n'est pas dynamique par nature et de nombreux soucis arrivent si vous essayez de les utiliser de façon dynamique. Dès que vous utilisez deux ou plus transformations ou tâches en même temps sur un même serveur (par exemple dans l'environnement de travail Pentaho), vous risquez d'avoir des conflits. De plus, les

changements de variables d'environnement sont visibles de tous les programmes qui tournent sur la machine virtuelle Java (JVM) (ndt leur portée est celle de la JVM).

8.2.2. Les variables de Kettle

La portée des variables d'environnement ([Variables d'environnement](#)) étant trop importante, nous avons introduit des variables Kettle, locales à la tâche pour laquelle la variable est définie. L'étape « Création variable » dans une transformation vous permet de définir quels sont les tâches pour les quelles la variable est définie (tâche parent, tâche grand parent, tâche racine).

8.2.3. Les variables internes

Ces variables sont toujours définies :

| Nom de la variable | exemple de valeur |
|-------------------------------|---------------------|
| Internal.Kettle.Build.Date | 2006/09/06 18:01:39 |
| Internal.Kettle.Build.Version | 9 |
| Internal.Kettle.Version | 0 |

Ces variables sont définies pour une transformation :

| Nom de la variable | exemple de valeur |
|--|--|
| Internal.Transformation.Filename.Directory | D:\Kettle\samples |
| Internal.Transformation.Filename.Name | Dénormalisation - 2 séries de paires clé-valeur.ktr |
| Internal.Transformation.Name | Exemple de dénormalisation - 2 séries de paires clé-valeur |
| Internal.Transformation.Repository.Directory | / |

Ces variables sont définies pour une tâche :

| Nom de la variable | exemple de valeur |
|---------------------------------|-------------------|
| Internal.Job.Filename.Directory | /home/matt/jobs |
| Internal.Job.Filename.Name | Nested jobs.kjb |
| Internal.Job.Name | test de tâche |

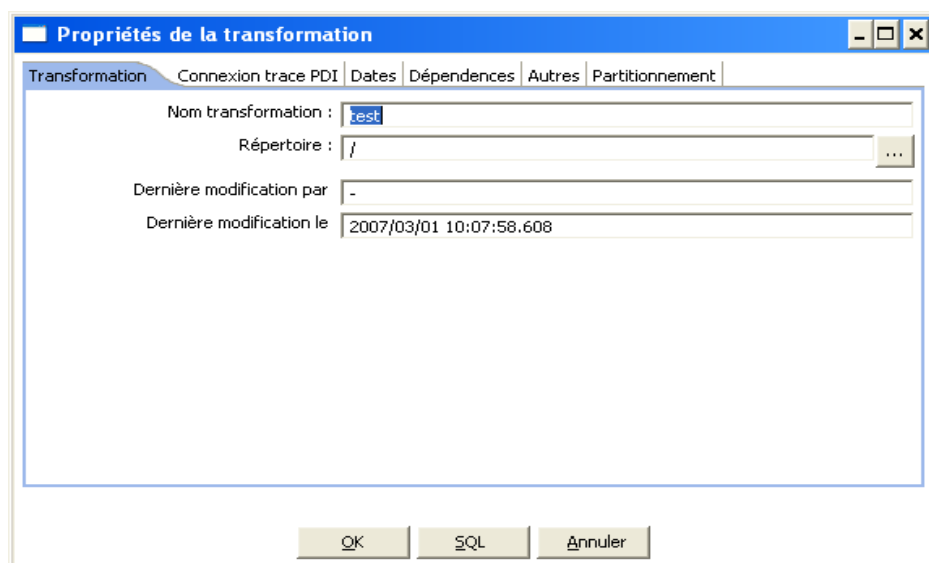
| | |
|-----------------------------------|---|
| Internal.Job.Repository.Directory | / |
|-----------------------------------|---|

9. Paramétrage des transformations

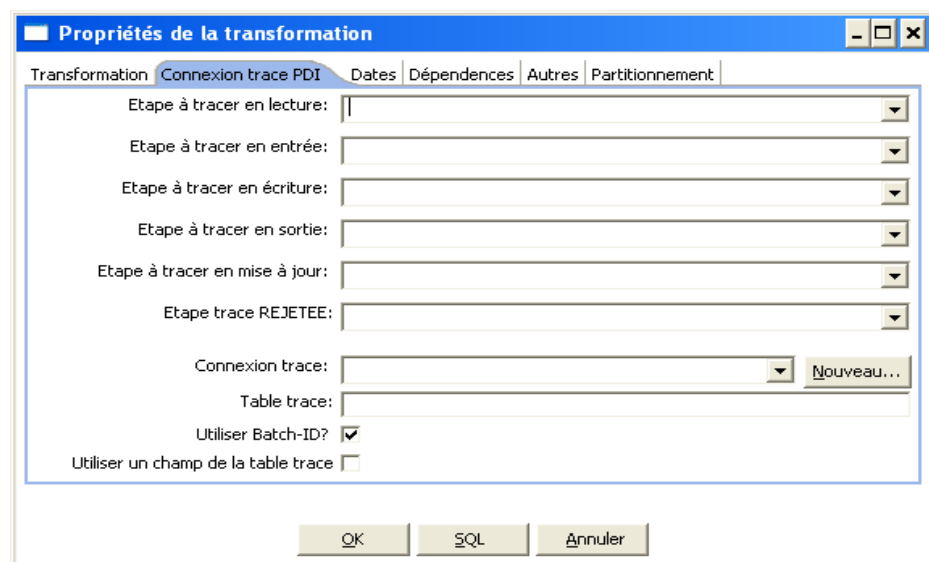
9.1. Description

Quelques options permettent de contrôler le comportement des transformations et de leur journalisation. Pour paramétrer ces options, sélectionner l'item de menu Transformation|Paramètres (ou CTRL+T)

9.2. Captures d'écran



l'onglet Transformation



l'onglet connexion trace PDI

Propriétés de la transformation

Transformation | Connexion trace PDI | **Dates** | Dépendences | Autres | Partitionnement

Connexion Maxdate: [dropdown]

Table Maxdate: [text field]

Champ Maxdate: [text field]

Maxdate décalage (seconds): 0.0

Max. date difference (seconds): 0.0

OK SQL Annuler

l'onglet "Dates"

9.3. Paramètres

| Paramètres | Description |
|---------------------------|---|
| Nom de la transformation | le nom de la transformation, information requise si vous voulez sauvegarder dans un référentiel (ndt ou dans un fichier) |
| Répertoire | le répertoire du référentiel dans lequel vous désirez sauvegarder la transformation. |
| Dernière modification par | affiche le nom du dernier utilisateur qui a modifié la transformation (ndt dans le cas d'un référentiel, c'est l'utilisateur connecté au référentiel; dans les autres cas, l'utilisateur peut être indéterminé) |
| Dernière modification le | affiche la date et l'heure de la modification enregistrée (ndt et non seulement modification, vous ne pourrez pas ainsi déterminer une modification en cours). |
| Étape à tracer en lecture | use the number of read lines from this step to write to the log table. Read means: read from source steps. |
| Étape à tracer en entrée | use the number of input lines from this step to write to the log table. Input means: input from file or database. |
| Étape à tracer en | use the number of written lines from this step to write to the log table. Written |

| Paramètres | Description |
|---|--|
| sortie | means: written to target steps. |
| OUTPUT log step | use the number of output lines from this step to write to the log table. Output means: output to file or database. |
| UPDATE log step | use the number of updated lines from this step to write to the log table. Update means: updated in a database. |
| Connexion trace | utilise cette connexion pour écrire une table de journalisation. |
| Table trace | définit le nom de la table de journalisation (par exemple L_ETL) |
| Utiliser Batch-ID? | activé si la table de journalisation possède un identifiant de batch. Désactivez-le pour être compatible avec les version de Spoon/Pan < |
| Utiliser un champ de table trace | Cette option permet d'enregistrer le texte de journalisation dans un champ de type CLOB au sein de la table de journalisation. Ainsi, vous serez en capacité d'avoir au sein de la même table la journalisation avec le résultat. Désactivez-le pour être compatible avec les version de Spoon/Pan < |
| connexionMaxdate connection | get the upper limit for a date range on this connection. |
| Maxdate table | get the upper limit for a date range in this table. |
| Maxdate field | get the upper limit for a date range in this field. |
| Maxdate offset | increases the upper date limit with this amount. Use this for example, If you find that the field DATE_LAST_UPD has a maximum value of 2004-05-29 23:00:00, but you know that the values for the last minute are not complete. In this case, simply set the offset to - |
| Maximum date difference | sets the maximum date difference in the obtained date range. This will allow you to limit job sizes. |
| Dépendances | this table allows you to enter all the dependencies. For example if a dimension is depending on 3 lookup tables, we have to make sure that these lookup tables have not changed. If the values in these lookup tables have changed, we need to extend the date range to force a full refresh of the dimension. The dependencies allow you to look up whether a table has changed in case you have a "data last changed" column in the table. |
| Nb de lignes dans une plage | Ce paramètre vous permet de changer la taille du buffer (zone mémoire tampon) entre deux étapes connectées. Vous ne devriez jamais (rarement) changer ce paramètre. Si vous ne disposez que de peu de mémoire ou de temps d'exécution sur une machine, vous pourriez baisser ce paramètre. |
| Rétro-contrôle journalisé ? | Ce paramètre définit si oui ou non le programme journalise une information de contrôle pendant l'exécution d'une étape. Par défaut, ce paramètre est activé et une information est affichée toutes les 5000 lignes exécutées. |
| Fréquence du rétro-contrôle (nb lignes du | Ce paramètre définit le nombre de lignes entre deux journalisations de contrôle si le paramètre ci-dessus est activé. Augmentez ce paramètre si vos flux |

| Paramètres | Description |
|-----------------------------|---|
| flux) | d'informations sont importants et/ou que vous voulez alléger votre fichier journal. |
| Use unique connections | This allows use to open one unique connection per defined and used database connection in the transformation. Checking this option is required in order to allow a failing transformation to be rolled back completely. |
| Fichier des objets partagés | Ce paramètre définit le chemin du fichier XML utilisé pour sauvegarder les objets partagés, comme les connexions aux bases de données, les schémas de grappe (ndt en anglais clustering) et autres. |
| Manage thread priorities? | Ce paramètre vous permet de désactiver la gestion interne des priorités entre les processus (ndt en anglais thread), fondé essentiellement sur le nombre de lignes dans les flux d'entrée et de sortie. Cela peut vous servir dans des situations simples pour lesquelles le coût de cette gestion est trop important au regard du bénéfice à l'exécution.. |
| Partitionnement | Cet onglet vous fournit une liste de partitions de base s de données possibles. Pour créer une nouvelle partition, utilisez le bouton « Nouveau ». Le bouton « Récupérer partitions » essaiera de récupérer les partitions qui ont été définies pour cette connexion. Pour plus de détails, voir Connexion aux bases de données . |

9.4. Extra

- Récupérer les dépendances (ndt onglet dépendances): le programme essaie de récupérer automatiquement les dépendances.
- Bouton SQL (ndt en bas du dialogue): génère l'instruction SQL nécessaire pour créer la table « trace » dans la base de données pointée par la connexion choisie pour la trace.

10. Les étapes au sein d'une transformation

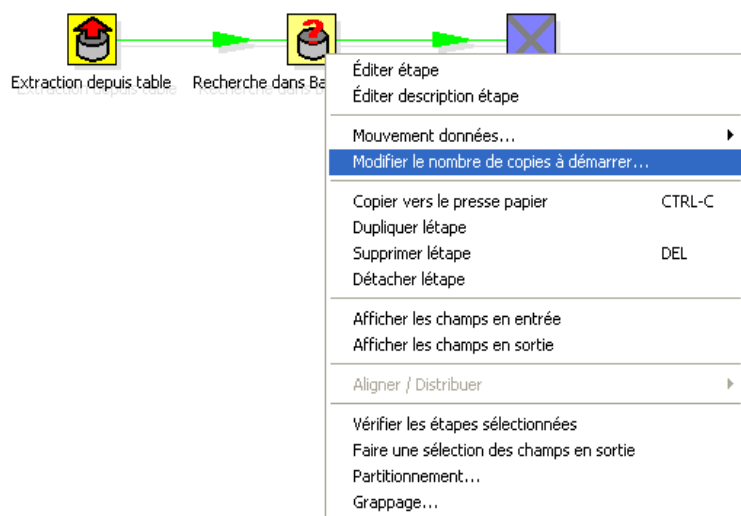
10.1. Description

Une transformation est composée d'étapes. Une étape peut fournir de nombreux services, comme lire un fichier texte, une table dans une base de données, réaliser un filtre sur des valeurs.... Pour connaître tous les types possibles d'étapes, voir ci dessous [Types d'étapes au sein d'une transformation](#).

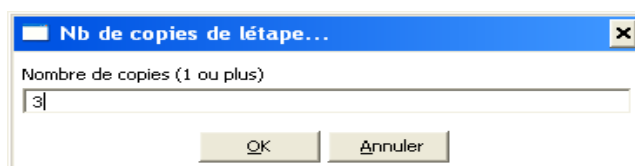
10.2. Lancer plusieurs copies de la même étape

Parfois, il peut être utile de lancer plusieurs fois la même étape. Par exemple, pour des raisons de performance, il peut vous arriver de devoir lancer une lecture de base trois fois ou plus, ceci parce que les connexions à cette base ont une certaine latence; lancer plusieurs fois la même étape occupe les connexions et diminue effectivement la latence (ndt: on est dans le cambouis et c'est normalement le serveur de la base de données qui doit gérer ceci; par contre, il peut y avoir des limites de taille dans le flux qui exige de relancer plusieurs fois la même étape pour épuiser la quantité d'informations à lire ou écrire).

Vous pouvez simplement lancer plusieurs copies d'une étape au sein d'une transformation en changeant le paramètre « Modifier le nombre de copies à démarrer ». Pour cela, cliquez droit sur l'étape désirée sur la représentation graphique de la transformation et sélectionnez ce paramètre



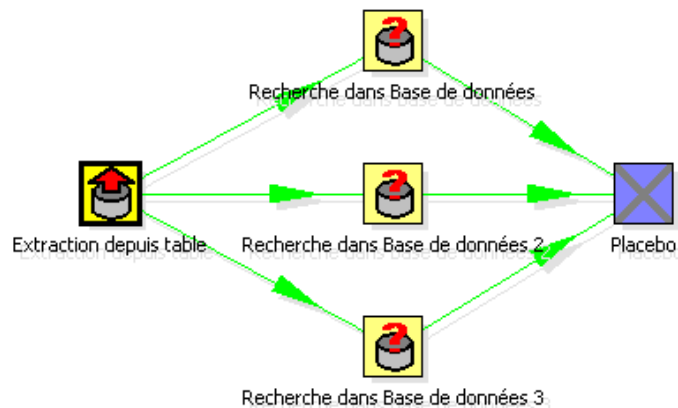
Vous aurez alors ce dialogue :



Si vous saisissez le nombre 3, vous verrez alors :



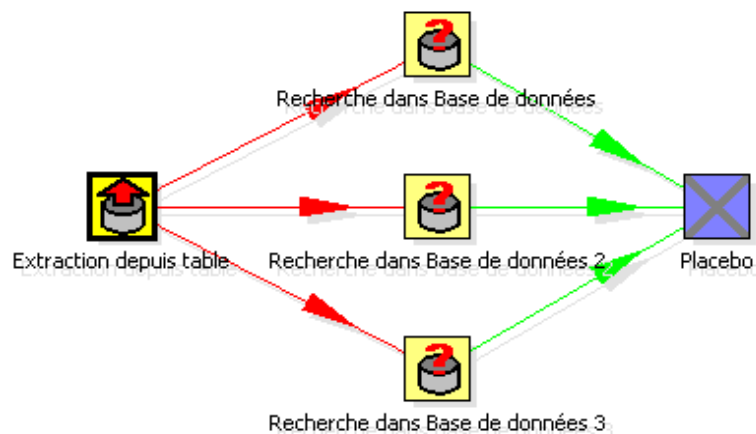
Ce qui est équivalent techniquement à :



10.3. Distribuer ou copier ?

Dans l'exemple ci-dessus, les liens qui relient les différentes étapes sont en vert. Ceci indique que les lignes sont distribuées entre les étapes cibles. Dans ce cas, cela signifie que la première ligne provenant de l'étape « Extraction depuis table » est dirigée sur l'étape « Recherche dans Base de données », la deuxième ligne vers « Recherche dans Base de données 2 », la troisième vers « Recherche dans Base de données 3 », la quatrième vers « Recherche dans Base de données » etc.

Si nous cliquons droit sur la première étape et que nous sélectionnons « Copier les données vers les étapes suivantes » dans le sous-menu « Mouvement », alors les liens deviendront rouges :



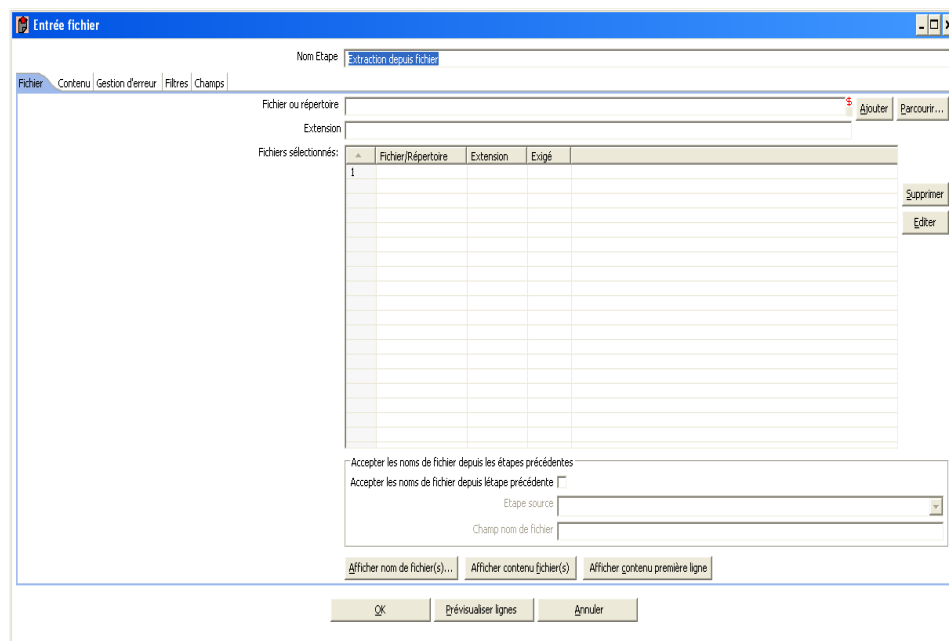
Toutes les lignes vont alors être distribuées à chacune des étapes suivantes, ce qui signifie que les lignes trouvées par les étapes « Recherche dans Base de données » vont être triplées à l'arrivée dans la dernière étape.

NOTE: Les processus utilisés par toutes ces étapes sont différents. L'ordre dans lequel les lignes arrivent sera donc différent et non garanti pour ces différentes situations.

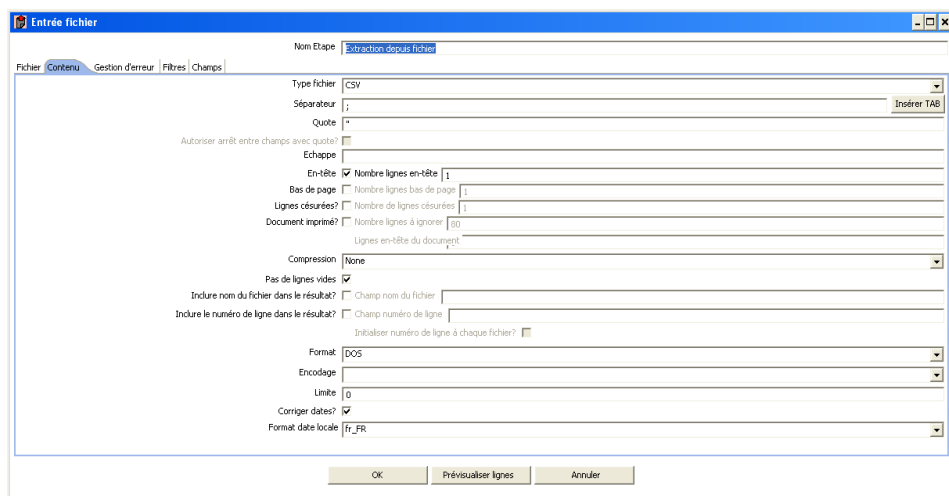
10.4. Types d'étapes au sein d'une transformation

10.4.1. Extraction depuis fichier (ndt en anglais text file input)

10.4.1.1. Captures d'écran



"onglet "Fichier"



l'onglet "contenu"

Vous trouverez ci-dessous des exemples d'expressions régulières utilisables :

| Nom de fichier | Expression régulière | Signification |
|----------------|----------------------|--|
| /dirA/ | .*userdata.*\.txt | Tous les fichiers du répertoire /dirA/ comportant la chaîne de caractères « userdata » dans le nom et finissant par « .txt » |
| /dirB/ | AAA.* | Tous les fichiers du répertoire /dirB/ dont le nom commence par la chaîne de caractères « AAA » |
| /dirC/ | [A-Z][0-9].* | Tous les fichiers du répertoire /dirC/ dont le nom commence par une lettre majuscule suivie immédiatement par un chiffre (A0-Z9) |

Vous trouverez ci-dessous tous les paramètres possibles

| Paramètre | Description |
|-----------------------|---|
| Fichier ou Répertoire | Vous pouvez saisir dans ce champ le chemin et/ou le nom du fichier ou du répertoire à inclure. NOTE: utilisez le bouton « Ajouter » pour ajouter le fichier/répertoire ou l'expression régulière à la liste des fichiers. |
| Expression régulière | Vous pouvez saisir l'expression régulière désirée |
| File type | This can be either CSV or Fixed length. Based on this selection, Spoon will launch a different helper GUI when you press the "get fields" button in the last "fields" tab. |
| Requis (O/N) | Si un fichier est requis et qu'il n'est pas trouvé, alors une erreur est générée. Dans le cas contraire ou le fichier n'est pas requis et que celui-ci n'est pas trouvé, cette source n'est pas utilisée sans erreur générée. |

Accepter les noms de fichier depuis les étapes précédentes
 Accepter les noms de fichier depuis l'étape précédente ☒

Etape source

Champ nom de fichier

10.4.1.4.2. Accepter des noms de fichiers depuis l'étape précédente

Ce paramètre vous permet une plus grande flexibilité dans les combinaisons avec les autres étapes comme "Récupération nom fichiers". Vous pouvez construire de toute pièce un nom de fichier et le transmettre à cette étape. Ainsi, le nom de fichier peut provenir de n'importe quelle source : d'un fichier texte, une table de base de données, etc.

| Paramètre | Description |
|--|--|
| Accepter les noms de fichier depuis les étapes précédentes | Ce paramètre active l'option de récupération d'un nom de fichier depuis une étape précédente |
| Étape source | L'étape dont provient le nom de fichier |
| Champ du nom de fichier | Le champ du flux de sortie de l'étape source spécifiée dans lequel devrait se trouver le ou les noms de fichier. |

10.4.1.4.3. Contenu

L'onglet « Contenu » vous permet de définir le format des informations contenues dans les fichiers « texte » qui

The screenshot shows the 'Entrée fichier' dialog box with the 'Contenu' tab selected. The 'Nom étape' is 'Extraction depuis fichier'. The settings include:

- Type fichier: CSV
- Séparateur: ;
- Quote: "
- Autoriser arrêt entre champs avec quote? ☐
- Echappe: \
- En-tête: ☒ Nombre lignes en-tête: 1
- Bas de page: ☐ Nombre lignes bas de page: 1
- Lignes césurées? ☐ Nombre de lignes césurées: 1
- Document imprimé? ☐ Nombre lignes à ignorer: 80
- Lignes en-tête du document: 1
- Compression: None
- Pas de lignes vides: ☒
- Inclure nom du fichier dans le résultat? ☐ Champ nom du fichier:
- Inclure le numéro de ligne dans le résultat? ☐ Champ numéro de ligne:
- Initialiser numéro de ligne à chaque fichier? ☐
- Format: DOS
- Encodage:
- Limite: 0
- Corriger dates? ☒
- Format date locale: fr_fr

Buttons at the bottom: OK, Prévisualiser lignes, Annuler.

onglet "contenu"

seront lues. Voici la liste de ces paramètres :

| Paramètre | Description |
|-------------------|--|
| Type des fichiers | Le type de fichier peut être soit de longueur fixe, soit délimité (CSV). Se fondant sur cette sélection, Spoon lancera des assistants différents lorsque vous pressez le bouton « Récupérer champs » dans l'onglet « Champs ». |
| Délimiteur | Cela peut être un ou plusieurs caractères délimitant chaque champ. Typiquement, c'est un point-virgule (;) ou une tabulation (ndt il y a un bouton à droite pour générer une tabulation). |
| Entouré par | Certains champs peuvent être entourés par une paire de chaînes de caractères |

| Paramètre | Description |
|---|---|
| | afin de permettre une séparation des champs. Ce paramètre est optionnel. |
| Échappement | Ce paramètre définit le caractère d'échappement (ou une séquence de caractères) si vous avez échappé des caractères dans vos données. |
| En-tête et nombre de lignes d'en-tête | Activez ce paramètre si votre fichier texte comprend un en-tête (premières lignes du fichier). Vous pouvez préciser le nombre de lignes d'en-tête. |
| Fin de fichier & nombre de lignes de fin de fichier | Activez ce paramètre si votre fichier texte comprend une fin de fichier (dernières lignes du fichier). Vous pouvez préciser le nombre de lignes de fin de fichier. |
| Césure & nombre de lignes césurées | Activez ce paramètre si vous utilisez des lignes de données qui ont été césurées (ndt wrap en anglais), souvent à cause d'une limite en nombre de caractères par ligne. Veuillez noter que les en-têtes et fin de fichier ne sont jamais considérées césurées. |
| Document paginé, nombre de lignes d'en-tête et nombre de lignes par page. | Vous pouvez utiliser ce paramètre en dernier ressort si vos données sont structurées pour une impression (lignes d'en-tête toutes les n lignes). Il vous faut alors préciser le nombre de lignes d'en-tête à ignorer et le nombre de lignes à prendre en compte après cet en-tête. |
| Compression | Activez ce paramètre si votre fichier est placé dans une archive compressée en sélectionnant le format de compression.. NOTE: <i>Pour le moment, seul le premier fichier de l'archive est lu.</i> |
| Pas de lignes vides | Les lignes vides ne sont pas envoyées dans le flux de sortie (étape(s) suivante(s)). |
| Inclure le nom de fichier dans la sortie | Activez ce paramètre si vous désirez que le nom du fichier fasse partie du flux de sortie (ndt: concrètement, ce sera une colonne de plus dans chacune des lignes, attention aux performances). |
| Nom du champ « Nom de Fichier » | Saisissez ici le nom du champ qui contiendra le nom du fichier. |
| Inclure le numéro de ligne dans la sortie ? | Activez ce paramètre si vous désirez que le numéro de ligne fasse partie du flux de sortie (ndt: concrètement, ce sera une colonne de plus dans chacune des lignes, attention aux performances). |
| Nom de champ « Numéro de ligne » | Saisissez ici le nom du champ qui contiendra le numéro de ligne. |
| Initialiser le numéro de ligne à chaque fichier ? | Activez ce paramètre si vous désirez que le numéro de ligne soit initialisé (ndt à zéro) à chaque nouveau fichier dans le flux d'entrée. (ndt combiné avec le nom de fichier, vous pouvez ainsi tracer au sein même du flux les données de la transformation, quelque soit l'ordre dans lequel celles-ci sont traitées) |
| Format | Ce paramètre peut-être soit « DOS », soit « UNIX » soit « mixed ». Les fichiers « UNIX » présentent des lignes terminées par le caractère unicode U+000A (Line feed ou LF). Les fichiers DOS utilisent deux caractères unicode en fin de ligne: U+000D (Carriage Return CR) suivi de U+000A (LF). Si vous définissez le paramètre à « mixed », aucune vérification n'est faite. |

| Paramètre | Description |
|------------------------|---|
| Encodage | Ce paramètre définit l'encodage des caractères. Si vous le laissez à blanc, alors Spoon utilisera l'encodage par défaut de votre environnement. Pour utiliser l'unicode, précisez UTF-8 ou UTF-. Lors de la première utilisation, Spoon cherchera au sein de votre environnement les encodages possibles. |
| Limite | Ce paramètre définit le nombre maximum de lignes lues dans un fichier. 0 signifie qu'il n'y a pas de limites. (ndt ce paramètre vous permet de limiter les exécution de fichiers trop grands... avec Spoon, vous pouvez rapidement générer des fichiers énormes sans vous en rendre compte). |
| Corriger les dates ? | Désactiver ce paramètre si vous désirez une stricte interprétation des dates. Si ce paramètre est actif, une date comme le 32 janvier sera interprétée comme le 1er février. |
| Localisation des dates | La localisation des dates est utilisée pour interpréter les dates. Interpréter une date comme « February 2nd, 2006 » avec une localisation en français (fr_FR) ne fonctionnera pas car « Fébruary » ne spoon attendra « Février ». |

10.4.1.4.4. Gestion d'erreur

L'onglet « Gestion d'erreur » permet de définir comment cette étape interagit avec le gestionnaire d'erreur.

| Option | Description |
|----------------------------------|---|
| Ignorer les erreurs ? | Activez ce paramètre si vous désirez ignorer les erreurs durant l'analyse des données. |
| Ignorer les lignes d'erreur | Activez ce paramètre si vous désirez ignorer les lignes qui contiennent des erreurs. Notez que vous pouvez aussi générer un fichier qui contient les lignes en erreur ou les lignes pour lesquelles il y a eu erreur (ndt pour un traitement ultérieur manuel par exemple). Si les lignes contenant des erreurs ne sont pas ignorées, le champ qui n'a pas réussi à être correctement analysé sera de valeur nulle. |
| Nom du champ « nombre d'erreur » | Saisissez ici le nom du champ qui contiendra le nombre d'erreurs de la ligne. |
| Nom du champ « champ(s) en | Saisissez ici le nom du champ qui contiendra la liste des champs en erreur de la ligne |

| Option | Description |
|--|---|
| erreur » | |
| Nom du champ « description de l'erreur » | Saisissez ici le nom du champ qui contiendra la description de l'erreur d'analyse du ou des champs de la ligne. |
| Répertoire des alertes | Lorsque des alertes sont créées, elles seront placées dans ce répertoire. Le nom du fichier d'alerte sera de la forme <répertoire alerte>/nom du fichier.<date_heure>.<extension alerte> |
| Répertoire des erreurs | Lorsque des erreurs sont créées, elles seront placées dans ce répertoire. Le nom du fichier d'erreur sera de la forme <répertoire erreur>/nom du fichier.<date_heure>.<extension erreur> |
| Répertoire des lignes en erreur | Lorsque des lignes sont en erreur et mises de côté (voir le paramètre « ignorer les lignes d'erreur »), ces lignes sont placées dans ce répertoire. Le nom du fichier contenant les lignes en erreur sera de la forme <répertoire lignes en erreur>/nom du fichier.<date_heure>.<extension lignes en erreur>. |

10.4.1.4.5. Filtres

| # | Filter string | Filter position | Stop on filter |
|-----|---------------|-----------------|----------------|
| 001 | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Cet onglet « Filtres » vous permet de définir les lignes que vous désirez ignorer dans le fichier texte en entrée (ndt la documentation ici n'est pas suffisamment détaillée et explicite)

| Paramètre | Description |
|-------------------------|--|
| Chaîne filtre | La chaîne de caractères à chercher |
| Position filtre | La position à laquelle la chaîne de recherche doit être dans la ligne. La première position dans la ligne est Si vous précisez un onmbre négatif, alors le filtre est cherché dans toute la ligne. |
| S'arrêter sur le filtre | Sélectionnez O (Oui) si vous désirez arrêter l'étape d'extraction depuis ce fichier si la chaîne « filtre » est trouvée. |

10.4.1.4.6. Champs

| Paramètre | Description |
|-----------------|---|
| Nom | Nom du champ |
| Type | Type du champ (ndt en version Number, string, date, boolean, integer, big number, serializable, binary) |
| Format | Voir le paragraphe Formats pour une description complète des formats possibles |
| Position | (ndt oubli dans la documentation anglaise) |
| Longueur | Pour le type <i>Number</i> : le nombre total de chiffres dans le nombre; Pour le type <i>String</i> : le nombre total de caractères; Pour le type <i>Date</i> : la longueur de la chaîne de caractères à restituer (i.e. 4 restituera seulement l'année). |
| Précision | Pour le type <i>Number</i> : le nombre de chiffres après la virgule; Pour les types <i>String</i> , <i>Date</i> , <i>Boolean</i> : non utilisé; |
| Devise | utilisé pour interpréter les nombres comme \$10,00 ou € |
| Décimal | Le signe décimal peut être un point "." (10;00) ou une virgule "," () |
| Groupe | Un groupe peut être une virgule "," (10;00) ou un point "." () |
| Nul si | Traiter cette valeur comme nulle |
| Défaut | La valeur par défaut de ce champ si aucune valeur n'est spécifiée (valeur manquante, champ vide) |
| Enlever espaces | Détermine si on enlève ou non les espaces autour de la chaîne de caractères et si on ne le fait qu'à gauche, qu'à droite ou pour les deux (ndt en anglais trim) |
| Répéter | O/N: Si la valeur correspondante de cette ligne est vide, répéter la dernière valeur rencontrée non vide pour ce champ. |

10.4.1.5. Formats

10.4.1.5.1. Format des nombres

L'information sur le format des nombres est issue de la documentation « Sun Java API documentation », que vous pouvez trouver ici: <http://java.sun.com/j2se/2/docs/api/java/text/DecimalFormat.html>

| Symbole | Ou? | Localisé | Signification |
|---------------|----------------------------|----------|--|
| 0 | Nombre | Oui | Chiffre |
| # | Nombre | Oui | Chiffre, le zéro ne s'affichant pas |
| . | Nombre | Oui | Séparateur décimal ou de devise |
| - | Nombre | Oui | signe moins |
| , | Nombre | Oui | Séparateur pour les groupes de 3 chiffres |
| E | Nombre | Oui | Séparateur entre la mantisse et l'exposant en notation scientifique. Le suffixe et le préfixe n'ont pas besoin d'être entourés de quote. |
| ; | Frontière d'un sous masque | Oui | Sépare le sous masque positif ou négatif |
| % | Préfixe ou suffixe | Oui | Multiplie par cent et affiche comme un pourcentage |
| \u2030 | Préfixe ou suffixe | Oui | Multiplie par cent et affiche comme un pour mille |
| ¤ (\u00A4) | Préfixe ou suffixe | Non | Signe monétaire, remplacé par le signe de la devise. Si doublé, il est remplacé par le signe international de la devise. Lorsqu'il est présent dans un masque, le séparateur de la décimale monétaire est utilisé sauf s'il existe déjà un séparateur décimal. |
| ' | Préfixe ou suffixe | Non | Utiliser une apostrophe (' quote) dans un préfixe ou un suffixe, par exemple, "'#'" formate 123 en "'#123'". Pour créer une quote elle-même, utilisez-en deux, par exemple "'# o'clock'". |

Notation scientifique

Dans un masque, le caractère significatif de l'exposant (E) doit être suivi immédiatement de un ou plusieurs chiffres. Exemple: "###E0" formate le nombre 1234 en "".

10.4.1.5.2. Format des nombres

L'information sur le format des dates est issue de la documentation « Sun Java API documentation », que vous pouvez trouver ici: <http://java.sun.com/j2se/2/docs/api/java/text/SimpleDateFormat.html>

| Lettre | Composant Date ou Heure | Présentation | Exemples |
|--------|---------------------------------|-------------------|---------------------------------------|
| G | Era designator | Texte | AD |
| y | Année | Année | 1996; 96 |
| M | Mois dans l'année | Mois | July; Jul; 07 |
| w | Semaine dans l'année | Nombre | 27 |
| W | Semaine dans le mois | Nombre | 2 |
| D | Jour dans l'année | Nombre | 189 |
| d | Jour dans le mois | Nombre | 10 |
| F | Jour de la semaine dans le mois | Nombre | 2 |
| E | Jour de semaine | Texte | Tuesday; Tue |
| a | Marqueur matin/après-midi am/pm | Texte | PM |
| H | Heure (0-23) | Nombre | 0 |
| k | Heure (1-24) | Nombre | 24 |
| K | Heure (0-11) am/pm | Nombre | 0 |
| h | Heure (1-12) am/pm | Nombre | 12 |
| m | Minute | Nombre | 30 |
| s | Seconde | Nombre | 55 |
| S | Milliseconde | Number | 978 |
| z | Fuseau horaire | GMT | Pacific Standard Time; PST; GMT-08:00 |
| Z | Fuseau horaire | RFC 822 time zone | -0800 |

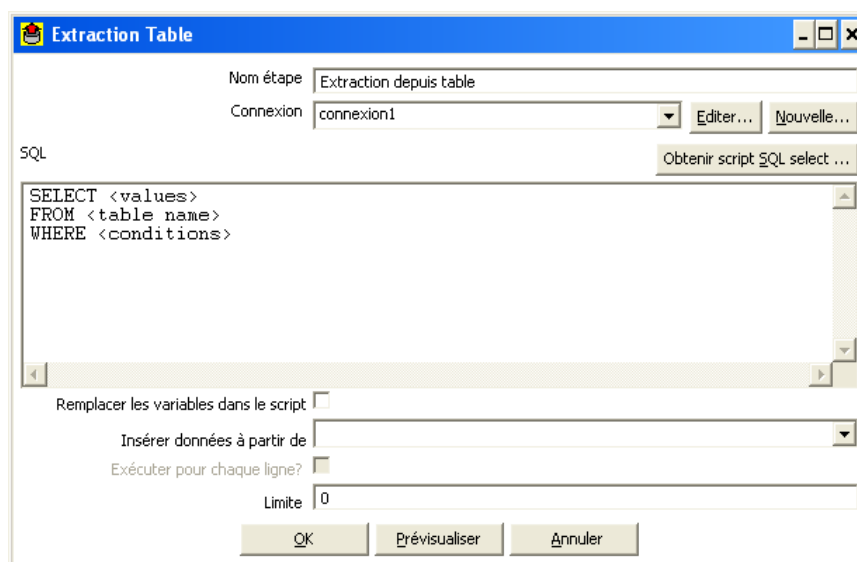
10.4.1.6. Extras

| Fonction/Bouton | Description |
|--|--|
| Afficher nom de fichier(s) onglet « Fichier » | Ce bouton vous permet d'afficher la liste de tous les fichiers sélectionnés (ndt dans l'onglet « Fichier »). Veuillez noter que si la transformation est exécutée sur un serveur distant, le résultat pourra ne pas être correct (ndt le système de fichier de ce serveur n'est peut-être pas accessible). |

| | |
|---|---|
| Afficher contenu fichier(s) onglet « Fichier » | Ce bouton montre les premières lignes du fichier texte sélectionné. Vérifiez que le format du fichier est correct. EN cas de doute, essayez les deux formats possibles DOS et UNIX. |
| Afficher contenu première ligne onglet « Fichier » | Ce bouton vous aide à trouver la première ligne dans les fichiers complexes avec plusieurs lignes d'en-tête. |
| Récupérer Champs onglet « Champs » | Ce bouton vous permet de deviner la structure de la ligne de votre fichier. Dans le cas d'un fichier délimité de type CSV, cette récupération devrait être automatiquement assez proche de la réalité. Si vous spécifiez un format fixe, vous devrez fixer les limites de chacun des champs en utilisant un assistant |
| Prévisualiser lignes tous onglets | Ce bouton vous permet de prévisualiser les lignes générées par cette étape. |

10.4.2. Extraction depuis table (ndt en anglais Table input)

10.4.2.1. Captures d'écran



10.4.2.2. Icône



10.4.2.3. Description générale

Cette étape est utilisée pour extraire des données depuis une base de données, en utilisant une connexion et le langage SQL. Des requêtes basiques SQL sont générées automatiquement, pour vous faciliter la tâche.

10.4.2.4. Paramètres

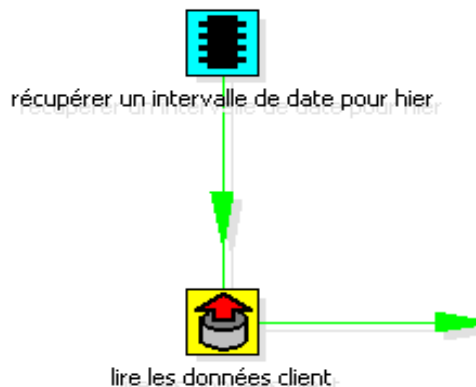
| Paramètre | Description |
|--|---|
| Nom de l'étape | Ce nom doit être unique au sein de la transformation. |
| Connexion | La connexion utilisée pour lire les données. |
| SQL | La requête SQL utilisée pour récupérer les données. |
| Remplacer les variables dans le script | |
| Insérer les données depuis l'étape | Ce paramètre permet de définir le nom d'une étape utilisée pour récupérer une information que vous pouvez utiliser au sein de la requête SQL avec le marqueur ? (point d'interrogation). (ndt attention à l'ordre des informations récupérées. Une seule étape est possible. Aussi, si vous avez besoin de récupérer des informations provenant de plusieurs étapes, créez cette information dans une première étape avant de l'utiliser ici) |
| Limite | Ce paramètre définit le nombre de lignes qui seront lues depuis la base. la valeur 0 signifie toutes les lignes. |

10.4.2.5. Exemple

Si, par exemple, vous saisissez la requête SQL suivante :

```
SELECT * FROM customers WHERE changed_date BETWEEN ? AND ?
```

Cette requête a besoin de deux dates lues dans une étape précédente.



NOTE: Les dates peuvent être fournies par exemple par une étape du type « Récupération info système ».

L'étape « Lire les données client » ressemble à cela :

| | Nom | Type | |
|---|-------|---------------|--|
| 1 | début | Hier 00:00:00 | |
| 2 | fin | Hier 23:59:59 | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Et l'étape « récupérer un intervalle de date pour hier » :

Voir aussi [#Récupération des données système](#)

10.4.2.6. Extras

| Fonction/Bouton | Description |
|-----------------|--|
| Prévisualiser | Ce bouton vous permet de prévisualiser cette étape. Pour cela, Spoon effectue deux étapes, l'étape à prévisualiser et une étape factice (ndt en anglais dummy). Vous pouvez voir le journal d'exécution en cliquant sur le bouton « Journal » au sein de la fenêtre de prévisualisation. |

10.4.3. Récupération des données système

10.4.3.1. Capture d'écrans

Informations Système

Nom étape : récupérer un intervalle de date pour hier

Champs:

| | Nom | Type |
|---|-------|---------------|
| 1 | début | Hier 00:00:00 |
| 2 | fin | Hier 23:59:59 |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

OK Annuler

Sélectionner un type d'info

Veuillez sélectionner le type d'information

- Date système (variable)
- Date système (fixé)
- Plage Date départ
- Plage Date fin
- start date range (Job)
- end date range (Job)
- Hier 00:00:00
- Hier 23:59:59
- Aujourd'hui 00:00:00
- Aujourd'hui 23:59:59
- Demain 00:00:00
- Demain 23:59:59
- Premier jour du mois dernier 00:00:00
- Dernier jour du mois dernier 23:59:59
- Premier jour du mois en cours 00:00:00
- Dernier jour du mois en cours 23:59:59
- Premier jour du prochain mois 00:00:00
- Dernier jour du prochain mois 23:59:59
- Copie détape
- Nom de la transformation
- Nom du fichier de transformation
- Le dernier utilisateur à avoir modifié la transformation
- La date de la dernière modification de la transformation
- Batch ID transformation
- Batch ID de la tâche parente
- Nom de l'hôte
- L'adresse IP
- ligne commande argument 1
- ligne commande argument 2
- ligne commande argument 3
- ligne commande argument 4
- ligne commande argument 5
- ligne commande argument 6
- ligne commande argument 7
- ligne commande argument 8
- ligne commande argument 9
- ligne commande argument 10
- Version Pentaho Data Integration
- Sous version Pentaho Data Integration
- Date de livraison

OK Annuler



10.4.3.2. Icône

10.4.3.3. Description générale

Cette étape récupère des informations depuis l'environnement de l'outil Pentaho Data Integration (Kettle).

Il y a plusieurs types d'information que vous pouvez récupérer

| Type d'information | Description |
|--|---|
| Date système (variable) | Date système, changera à chaque fois que vous demandez une date. |
| Date système (fixe) | Date système, déterminée au début de la transformation. |
| Début plage de date (transformation) | Début d'une plage de dates, fondée sur l'information du journal ETL log. Voir aussi Paramétrage des transformations |
| Fin plage de date (transformation) | End of date range, based upon information in ETL log table. See, also Paramétrage des transformations |
| Début plage de date (tâche) | Start of date range based upon information in the ETL log table. See also Paramétrage des tâches |
| Fin plage de date (tâche) | End of date range based upon information in the ETL log table. See also Paramétrage des tâches |
| Hier 00:00:00 | Hier, début de la journée |
| Hier 23:59:59 | Hier, fin de la journée |
| Aujourd'hui 00:00:00 | Aujourd'hui, début de la journée |
| Aujourd'hui 23:59:59 | Aujourd'hui, fin de la journée |
| Demain 00:00:00 | Demain, début de la journée |
| Demain 23:59:59 | Demain, fin de la journée |
| Premier jour du mois dernier 00:00:00 | Début du mois dernier |
| Dernier jour du mois dernier 23:59:59 | Fin du mois dernier |
| Premier jour du mois en cours 00:00:00 | Début du mois en cours |

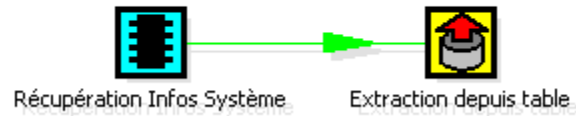
| Type d'information | Description |
|--|---|
| Dernier jour du mois en cours 23:59:59 | Fin du mois en cours |
| Premier jour du mois prochain 00:00:00 | Début du mois prochain |
| Dernier jour du mois prochain 23:59:59 | Fin du mois prochain |
| Copie d'étape | Nombre de copies (ou numéro de la copie ?) de l'étape (voir aussi Lancer plusieurs copies de la même étape). |
| Nom de la transformation | Nom de la transformation. |
| Nom du fichier de la transformation | Nom du fichier de la transformation (si XML et non référentiel) |
| Le dernier utilisateur à avoir modifié la transformation | Nom du dernier'utilisateur ayant modifié la transformation (ndt dans le cas Référentiel, l'utilisateur est connu) |
| La date de la dernière modification de la transformation | Date et heure de la dernière sauvegarde de la transformation |
| Identifiant du processus batch de la transformation | L'identifiant trouvé dans le journal d'exécution pour l'exécution « batch » de cette transformation |
| Identifiant du processus batch de la tâche | L'identifiant trouvé dans le journal d'exécution pour l'exécution « batch » de la tâche contenante. |
| Nom de l'Hôte | Nom d'hôte du serveur |
| Adresse IP | Adresse IP du serveur |
| ligne commande argument 1 à 10 | Argument 1 à 10 de la ligne de commande |
| Version Pentaho Data Integration | La version de Kettle (par exemple. 0) |
| Numéro de Build de la librairie Kettle | La version dite « Build » de la librairie Kettle (par exemple 960). (ndt Une librairie Keetle est construite (build) à chaque fois que le chef de projet le décide. (fréquence quotidienne début 2007)) |
| Date de Build de la librairie Kettle | Retourne la date de construction de la dernière version de la librairie Kettle |

10.4.3.4. Paramètres

- **Nom de l'étape:** Nom de l'étape. Ce nom doit être unique au sein d'une même transformation.
- **Champs:**les champs de sortie.

10.4.3.5. Utilisation

Le premier type d'utilisation est de récupérer une information du système:

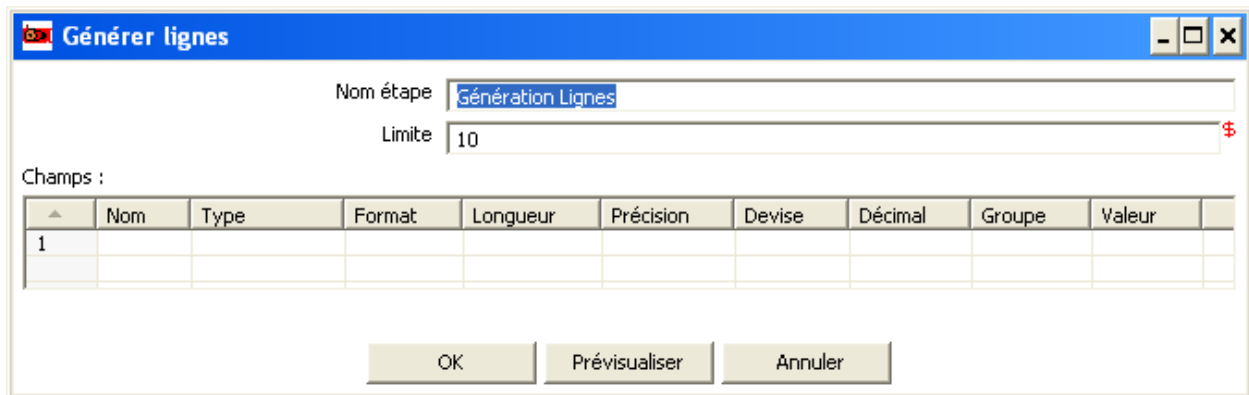


Depuis la version 0, cette étape accepte également des lignes en entrée. Les valeurs sélectionnées s'ajoutent aux lignes du flux d'entrée.



10.4.4. Génération de lignes (ndt en anglais Generate Rows)

10.4.4.1. Captures d'écran



Nom étape

Limite

Champs :

| | Nom | Type | Format | Longueur | Précision | Devise | Décimal | Groupe | Valeur | |
|---|-----|------|--------|----------|-----------|--------|---------|--------|--------|--|
| 1 | | | | | | | | | | |

OK Prévisualiser Annuler

10.4.4.2. Icône



10.4.4.3. Description générale

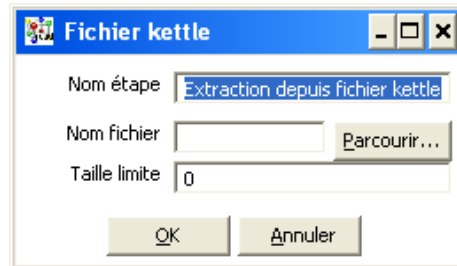
Cette étape génère en flux de sortie un nombre de lignes, par défaut vides, mais pouvant facultativement contenir des champs statiques.

10.4.4.4. Paramètres

- **Nom de l'étape:** Nom de l'étape. Ce nom doit être unique au sein d'une même transformation.
- **Limite:** définit le nombre maximum de lignes que vous voulez générer.
- **Champs:** les champs statiques que vous désirez inclure dans le flux de sortie.

10.4.5. Extraction depuis fichier kettle (ndt en anglais de-serialize from file formerly Cube Input)

10.4.5.1. Captures d'écran



10.4.5.2. Icône



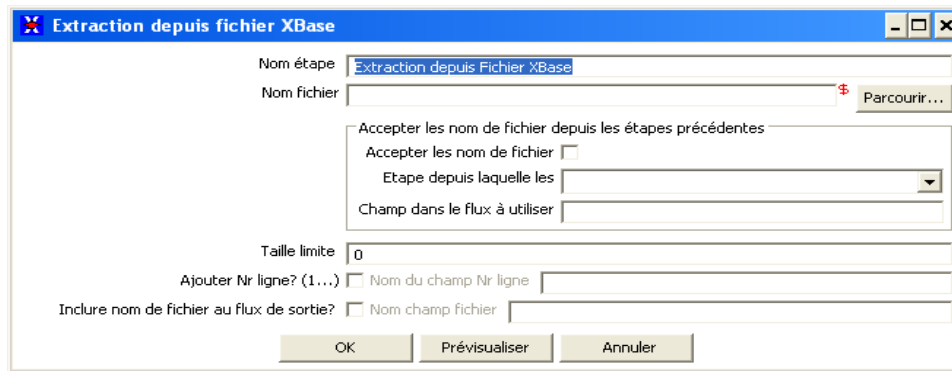
10.4.5.3. Description générale

Lit des lignes de données depuis un fichier contenant un cube (binaire) Kettle. (voir aussi le paragraphe sur l'alimentation d'un tel fichier)

NOTE: Cette étape ne devrait pas être utilisée comme un outil de création de cube. En effet, le format utilisé n'est pas un standard.

10.4.6. Extraction depuis fichier Xbase (ndt en anglais XBase input)

10.4.6.1. Captures d'écran



10.4.6.2. Icône



10.4.6.3. Description générale

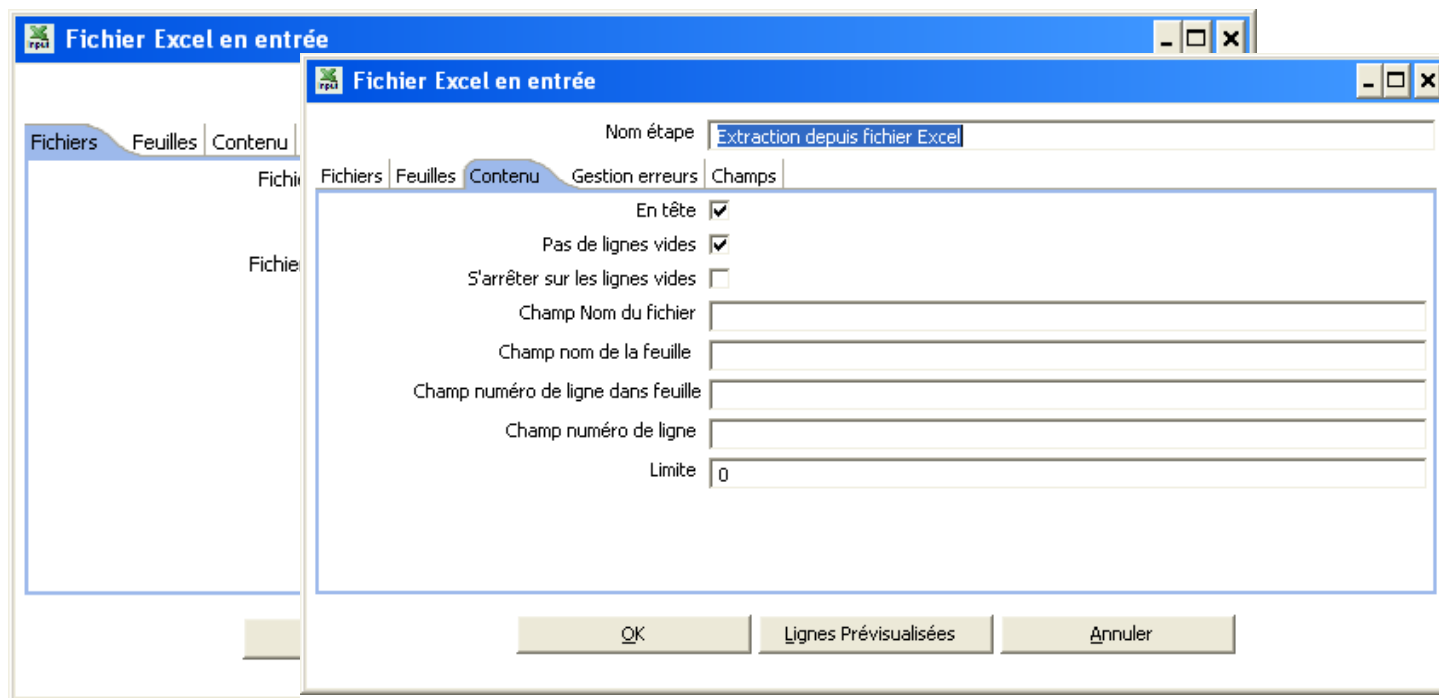
Cette étape rend possible la lecture de données provenant de la plupart des fichiers de type DBF ou dérivés. (dBase III/IV, Foxpro, Clipper, ...)

10.4.6.4. Paramètres

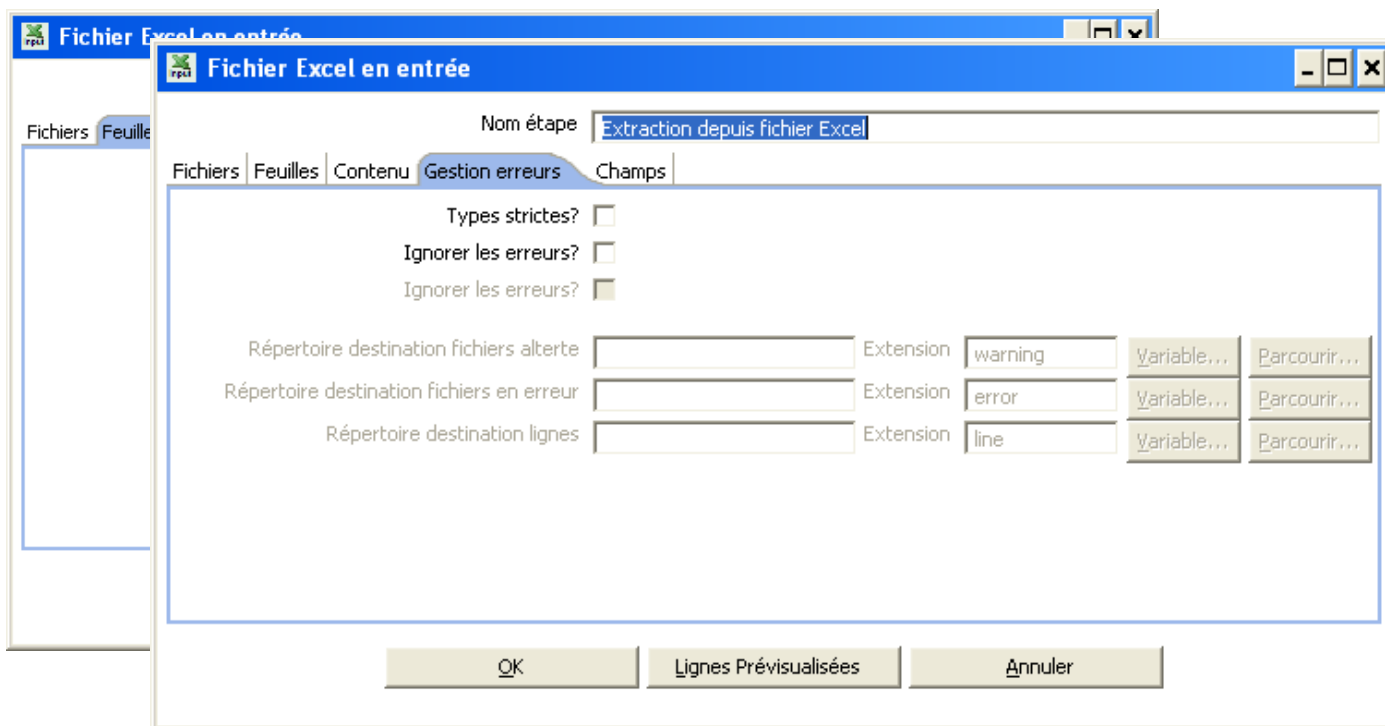
- **Nom de l'étape** : Nom de l'étape. Ce nom doit être unique au sein d'une même transformation.
- **Nom de fichier** : nom du fichier, paramètre acceptant les variables d'environnement
- **Taille Limite** : définit le nombre de lignes à lire, 0 signifie toutes
- **Ajouter numéro de lignes ?**: Ce paramètre permet d'ajouter un champ contenant le numéro de la ligne.
- **Nom du champ 'Numéro de ligne'** : Nom du champ
- **Inclure nom de fichier au flux de sortie ?**: Ce paramètre permet d'ajouter un champ contenant le nom du fichier Xbase
- **Nom du champ 'Nom de fichier'** : Nom du champ

10.4.7. Extraction depuis un fichier excel (ndt en anglais Excel input)

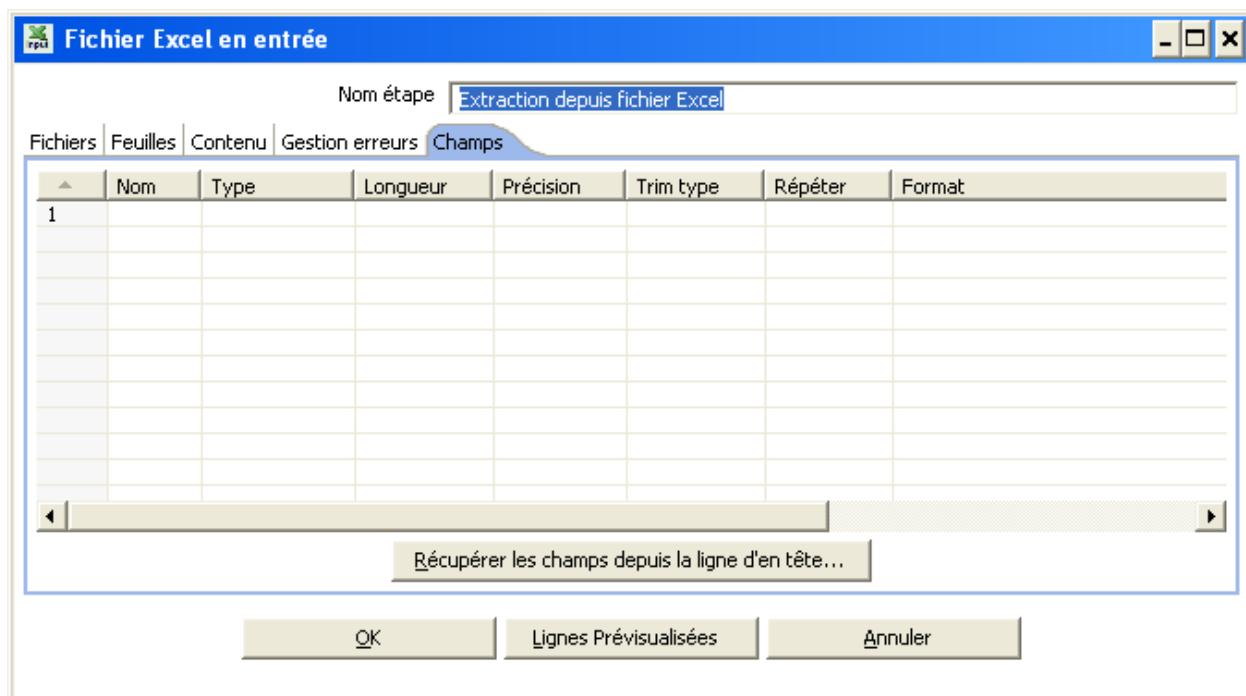
10.4.7.1. Captures d'écrans



l'onglet "Contenu"



l'onglet "Gestion d'erreur"



l'onglet "Champs"

10.4.7.2. Icône



10.4.7.3. Description générale

Cette étape vous permet de lire des données depuis un ou plusieurs fichiers excel. Cel fonctionne depuis tous les environnements surportant Kettle (ndt même si vous n'avez pas d'application excel sur cet environnement).

10.4.7.4. Paramètres

- **Nom de l'étape** : Nom de l'étape. Ce paramètre doit être unique au sein de la transformation.

Onglet « Fichier »:

- les nom de fichiers, avec le support des variables d'environnement.
- Voir aussi : [.Nom de fichier](#) et [Accepter des noms de fichiers depuis l'étape précédente](#)

Onglet « Feuilles » :

- Les feuilles à importer. Utiliser le bouton « Récupérer les noms de feuilles » pour générer une liste des feuilles possible de façon automatique. Veuillez noter que vous devez préciser la ligne et la colonne de départ (endroit où l'étape commence à lire) par leur coordonnées.

Onglet Contenu :

- **En-tête:** Activez ce paramètre si les feuilles présentent un en-tête qui devra être ignoré.
- **Pas de lignes vides :** Activez ce paramètre si vous ne désirez pas de lignes vides dans le flux de sortie de cette étape.
- **S'arrêter sur les lignes vides :** Activez ce paramètre si vous désirez arrêter de lire la feuille courante si l'étape lit une ligne vide.
- **Champ 'Nom du fichier' :** Nom du champ dans lequel inclure le nom du fichier excel en entrée de la ligne.
- **Champ 'Nom de la feuille' :** Nom du champ dans lequel inclure le nom de la feuille excel en entrée de la ligne.
- **Champ 'Numéro de ligne dans la feuille' :** Nom du champ dans lequel inclure le numéro de la ligne de la feuille excel.
- **Champ 'Numéro de ligne écrite' :** Nom du champ dans lequel inclure le numéro de la ligne écrite dans le flux de sortie, à commencer par 1 jusqu'à avoir épuisé toutes les lignes des feuilles des fichiers excel source, dans la limite du paramètre ci-après.
- **Limite :** précisez la limite en nombre de lignes à lire. 0 signifie toutes les lignes.

Onglet Génération d'erreur :

- Veuillez vous reporter à [Gestion d'erreur](#) pour une description détaillée de la gestion d'erreur.

Onglet Champs :

- Vous pouvez préciser ici les champs qui doivent être lus depuis les fichiers excel. Un bouton « Récupérer les champs » vous permet de générer automatiquement une liste de champs.
- Afin de permettre une conversion logique, cet onglet supporte dorénavant le formatage des données. Par exemple, si vous désirez lire une date et que vous récupérez une chaîne de caractères dans Excel, vous pourrez préciser un masque de formatage. Note: dans le cas d'une conversion de nombre à date (exemple: 20051028 --> 28 Octobre 2005), vous devriez simplement préciser un masque de conversion du type yyyyMMdd car la conversion « nombre à chaîne de caractères » est implicite et est réalisée avant la conversion « chaîne de caractères à date ».

10.4.8. Extraction depuis fichier XML (ndt en anglais XML Input)

10.4.8.1. Captures d'écran

The screenshot shows the 'Fichier XML en entrée' dialog box with the 'Fichier' tab selected. The 'Nom étape' field contains 'Extraction depuis fichier XML'. Below the tabs, there are input fields for 'Fichier ou répertoire' and 'Extension', each with an 'Ajouter' button. A table titled 'Fichiers sélectionnées:' has columns 'Fichier/Répertoire' and 'Extension'. The first row is numbered '1'. To the right of the table are 'Supprimer' and 'Editer' buttons. Below the table is an 'Afficher nom de fichier(s)...' button. At the bottom are 'OK', 'Prévisualiser lignes', and 'Annuler' buttons.

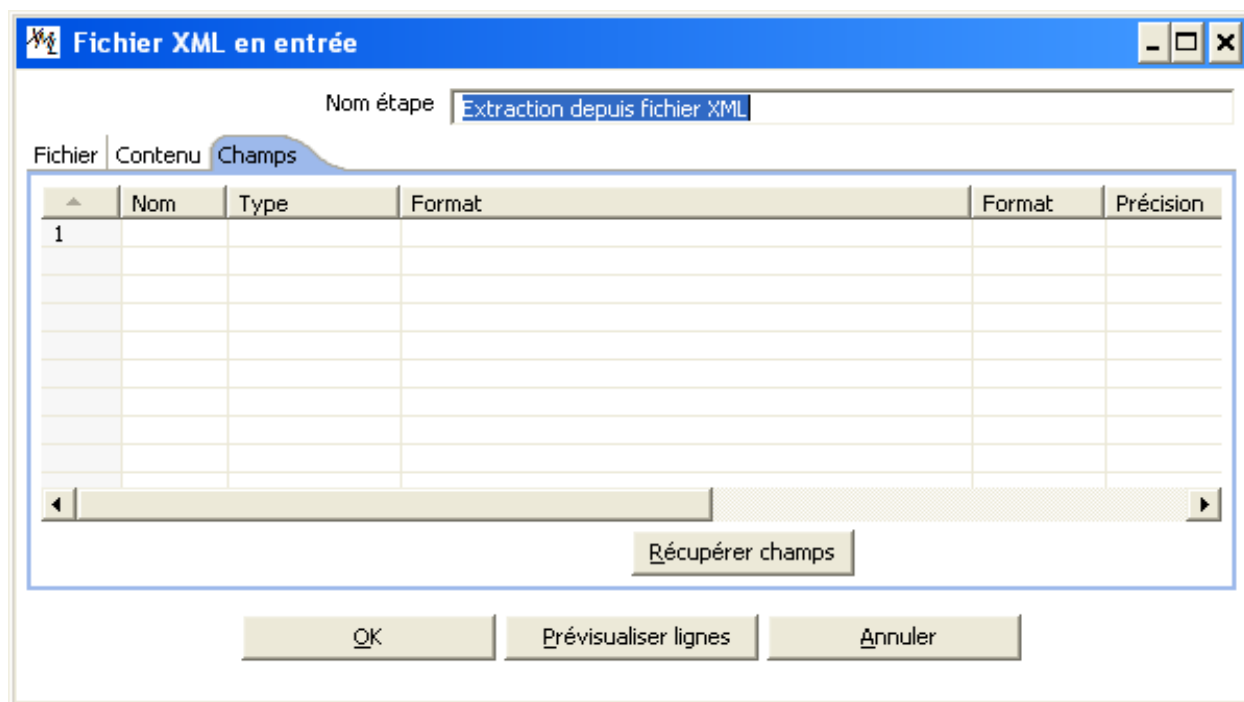
| | Fichier/Répertoire | Extension |
|---|--------------------|-----------|
| 1 | | |
| | | |
| | | |
| | | |
| | | |
| | | |

l'onglet "Fichier"

The screenshot shows the 'Fichier XML en entrée' dialog box with the 'Contenu' tab selected. The 'Nom étape' field contains 'Extraction depuis fichier XML'. Below the tabs, there are checkboxes for 'Inclure le nom de fichier en sortie?' and 'Inclure numéro de ligne en sortie?', each with a corresponding input field. Below these are input fields for 'Limite' and 'Nbr de lignes de début à ignorer', both set to '0'. A table titled 'Location' has columns 'Eléments'. The first row is numbered '1'. At the bottom are 'OK', 'Prévisualiser lignes', and 'Annuler' buttons.

| | Eléments |
|---|----------|
| 1 | |
| | |
| | |

l'onglet "Contenu"



l'onglet "Champs"

10.4.8.2. Icône



10.4.8.3. Description générale

Cette étape vous permet de lire les données provenant d'un fichier dans un format XML.

Pour ce faire, l'application affiche une interface qui vous permet de définir le nom des fichiers, les parties répétitives des fichiers ainsi que les champs que vous désirez lire. Vous pouvez définir ces champs par le chemin de l'élément de l'attribut. Vous avez la possibilité d'entrer un masque de conversion des données ou méta-données.

10.4.8.4. Paramètres

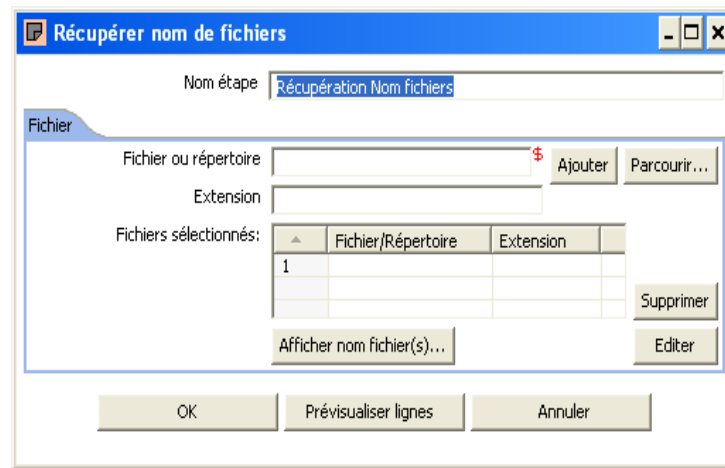
| Option | Description |
|-----------------------|--|
| Nom de l'étape | Nom de l'étape. Ce nom doit être unique au sein d'une même transformation. |
| Fichier ou répertoire | Ce paramètre permet de définir le chemin des fichiers et/ou des répertoires à parcourir. NOTE: Utilisez le bouton « Ajouter » pour ajouter le répertoire ou le fichier à la liste dessous. |

| Option | Description |
|--|--|
| Expression régulière | Vous pouvez utiliser une expression régulière pour sélectionner les fichiers en entrée. |
| Inclure nom de fichier en sortie & nom du champ | Activez ce paramètre si vous désirez que le nom du fichier XML soit inclus dans un champ au sein du flux de sortie. Précisez le nom du champ souhaité. |
| Inclure numéro de lignes en sortie et nom du champ | Activez ce paramètre si vous désirez que le numéro de ligne écrite en sortie (démarrage à 1) soit inclus dans un champ au sein du flux de sortie. Précisez le nom du champ souhaité. |
| Limite | Définissez ici le nombre de lignes maximum que vous désirez lire. 0 signifie toutes les lignes. |
| Nombre de lignes d'en-tête à ignorer | Définissez avec ce paramètre le nombre de lignes à ignorer au début du fichier XML. |
| Chemin | <p>Ce paramètre vous permet de définir le chemin des éléments répétitifs à prendre en compte dans le fichier XML.</p> <p>Par exemple, si vous lisez depuis le fichier XML :</p> <pre> <Rows> <Row> <Field1>...</Field1> ... </Row> ... </Rows> </pre> <p>Vous devrez préciser le chemin suivant : Rows, Row</p> <p>Note: Si vous utilisez la racine XML (ici Rows) comme élément répétitif. La sortie contiendra alors une seule ligne.</p> |
| Champs | <ul style="list-style-type: none"> Nom : Nom du champ Type : Type du champ (ndt en version Number, string, date, boolean, integer, big number, serializable, binary) Format : Voir le paragraphe Formats pour une description complète des formats possibles Position : Longueur : <ul style="list-style-type: none"> Pour le type Number: le nombre total de chiffres dans le nombre; Pour le type String: le nombre total de caractères; Pour le type Date: la longueur de la chaîne de caractères à restituer (i.e. 4 restituera seulement l'année). |

| Option | Description |
|--------|---|
| | <ul style="list-style-type: none"> • Précision : <ul style="list-style-type: none"> ○ Pour le type Number: le nombre de chiffres après la virgule; ○ Pour les types String, Date, Boolean: non utilisé; • Devise : utilisé pour interpréter les nombres comme \$10,00 ou € • Décimal : Le signe décimal peut être un point "." (10;00) ou une virgule "," (10,00) • Groupe : Un groupe peut être une virgule "," (10;00) ou un point "." (10,00) • Enlever espaces : Détermine si on enlève ou non les espaces autour de la chaîne de caractères et si on ne le fait qu'à gauche, qu'à droite ou pour les deux (ndt en anglais trim) • Répéter :O/N: Si la valeur correspondante de cette ligne est vide, répéter la dernière valeur rencontrée non vide pour ce champ. • Position: la position de l'élément XML ou de son attribut. Vous utiliserez la syntaxe suivante pour définir cette position : <u>Le premier élément appelé « foo »</u> : E=foo/1 <u>Le premier attribut appelé « bar »</u> : A=bar/1 <u>Le premier attribut appelé « bar » dans le deuxième élément appelé « foo »</u> : E=foo/2, A=bar/1 NOTE: Vous pouvez générer de façon automatique toutes les positions possibles dans le fichier XML en utilisant le bouton « Récupérer Champs ». NOTE: Les fichiers pour lesquels toute l'information est enregistrée dans l'élément racine du document XML sont également supportés. Le caractère R permet de récupérer cette information. |

10.4.9. Récupération nom fichier (ndt en anglais Get File Names)

10.4.9.1. Captures d'écran



10.4.9.2. Icône



10.4.9.3. Description générale

Cette étape vous permet de récupérer de l'information sur les noms des fichiers de votre système de fichier local.

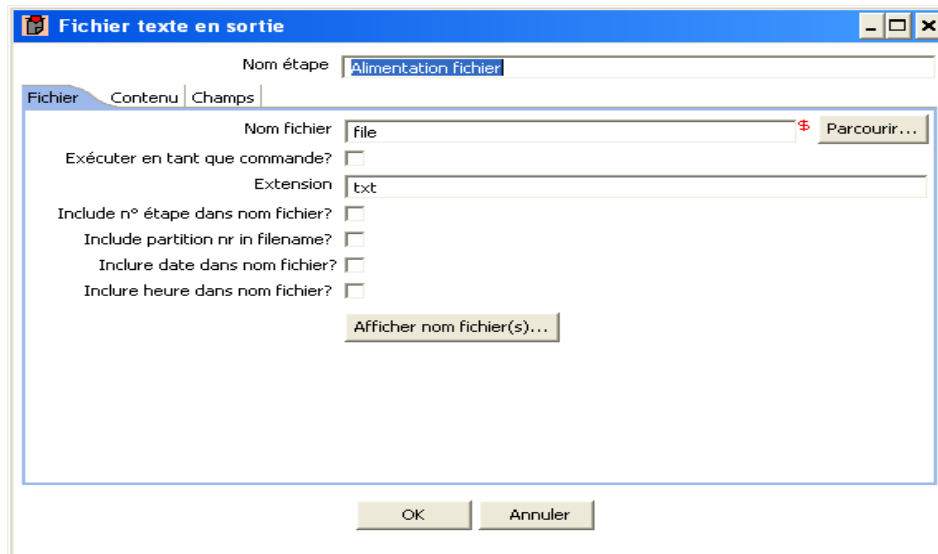
Veuillez consulter également le paragraphe [Nom de fichier](#) pour plus de détails sur la façon d'utiliser ces paramètres.

Le flux de sortie de cette étape est composé de :

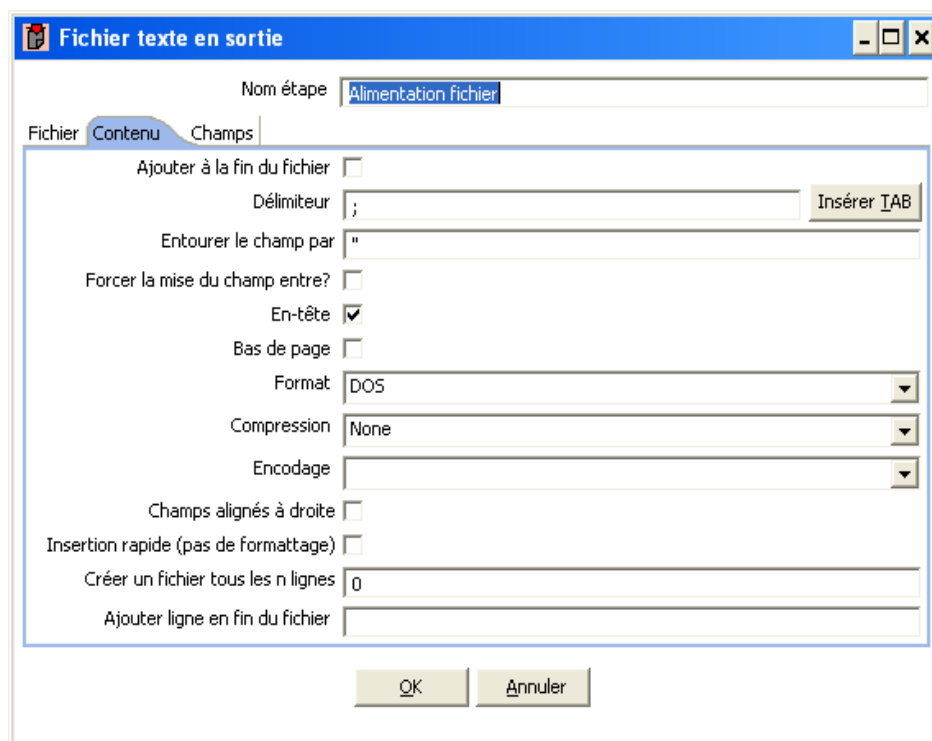
- Nom de fichier : le chemin complet du fichier (/tmp/kettle/somefile.txt)
- Nom court de fichier : seulement le nom du fichier, sans le chemin (somefile.txt)
- Chemin : seulement le chemin du fichier (/tmp/kettle/)

10.4.10. Alimentation de fichier (ndt en anglais Text File Output)

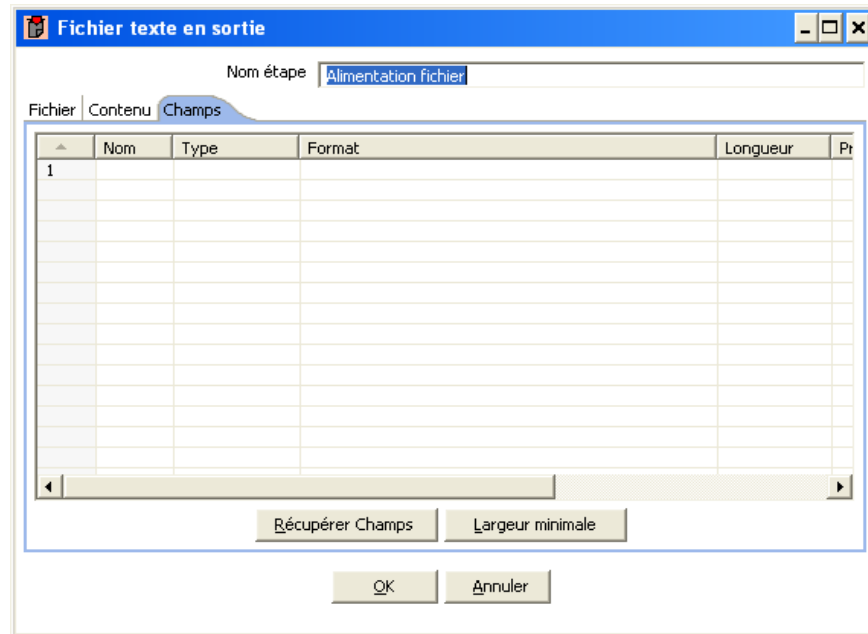
10.4.10.1. Captures d'écrans



l'onglet "Fichier"



l'onglet "Contenu"



l'onglet "Champs"

10.4.10.2. Icône



10.4.10.3. Description générale

Vous pouvez utiliser l'étape « Alimentation fichier » pour exporter le un flux de données dans un fichier au format texte. Les plus courants des formats textes sont les formats « dits » délimités de type CSV (Comma Separated Values) qui peuvent être lus par les tableurs.

10.4.10.4. Paramètres

| Paramètre | Description |
|---------------------------------|---|
| Nom de l'étape | Ce nom doit être unique au sein de la même transformation. |
| Nom du fichier | Ce paramètre vous permet de saisir le chemin complet du fichier en sortie. |
| Exécuter en tant que commande ? | Si vous activez ce paramètre, alors le flux de sortie sera envoyée comme flux d'entrée à la commande ou au script que vous spécifierez (ndt on s'éloigne du concept de fichier pour revenir à un flux de sortie pris comme stdout de cette étape et stdin d'une commande à spécifier) |

| Paramètre | Description |
|--|--|
| Extension | Ajoute un point et l'extension saisie à la fin du nom de fichier (.txt) |
| Inclure numéro d'étape dans le nom du fichier ? | Si vous exécutez l'étape en plusieurs copie (voir Lancer plusieurs copies de la même étape), le numéro de copie utilisé est inclus dans le nom du fichier, avant l'extension. (_0). |
| Inclure numéro de partition dans le nom du fichier ? | Ce paramètre permet d'inclure le numéro de la partition des données dans le nom du fichier en sortie (ndt dans le cas d'utilisation de la notion de grappage ou cluster). |
| Inclure la date dans le nom du fichier | Ce paramètre permet d'inclure la date système dans le nom du fichier de sortie. (_20041231). |
| Inclure l'heure dans le nom de fichier | Ce paramètre permet d'inclure l'heure système dans le nom du fichier de sortie. (_235959). |
| Ajouter à la fin du fichier | Activez ce paramètre si vous désirez que le flux de données soit ajouter à la fin du fichier, sans l'effacer au préalable. |
| Délimiteur | Vous pouvez préciser ici le(s) caractère(s) qui délimite(nt) les champs au sein d'une ligne de données. C'est typiquement un point-virgule « ; » ou une tabulation. |
| Entouré par | Une chaîne de caractères peut entourer les champs. Cela permet d'inclure des caractères délimiteurs au sein des champs sans risque de confusion. Ce paramètre est facultatif. |
| Forcer la mise du champ entre ? | Ce paramètre permet de forcer tous les champs de sortie à être entourés par la chaîne de caractères saisie ci-dessus. |
| En-tête | Activez cette option si vous souhaitez que le fichier texte se dote d'une ligne |

| Paramètre | Description |
|--------------------------------------|---|
| | d'en-tête. (première ligne du fichier). |
| Fin de fichier | Activez cette option si vous désirez que le fichier de sortie soit doté d'une ligne de fin de fichier. (Dernière ligne dans le fichier). |
| Format | Ce paramètre peut-être soit « DOS », soit « UNIX ». Les fichiers « UNIX » présentent des lignes terminées par le caractère unicode U+000A (Line feed ou LF). Les fichiers DOS utilisent deux caractères unicode en fin de ligne: U+000D (Carriage Return CR) suivi de U+000A (LF). |
| Encodage | Précisez ici l'encodage de caractère à utiliser dans le fichier de sortie. Laissez à vide ce paramètre pour utiliser l'encodage par défaut de l'environnement. Pour utiliser l'Unicode, saisissez UTF-8 or UTF- A la première utilisation, Spoon cherchera dans l'environnement quels sont les encodages possibles. |
| Compression | Ce paramètre vous permet de compresser le fichier (zip our gzip) NOTE: <i>Pour le moment, seulement un fichier peut être placé dans l'archive compressée.</i> |
| champs normalisés à droite ? | Activez ce paramètre si vous souhaitez que tous les champs soient « normalisés » à droite, c'est-à dire que spoon ajoute des espaces ou tronque le champ pour qu'il ait la longueur spécifié dans l'onglet « Champs ». (ndt, c'est une façon de faire du texte fixe en délimité, ce qui n'est pas performant en terme de taille de fichier) |
| Insertion rapide (pas de formatage) | Activez ce paramètre afin d'améliorer les performance du processus en cas de gros volumes de données. Les formatages demandés dans l'onglet « Champs » ne seront pas réalisés. |
| Créer un fichier toutes les n lignes | Si ce nombre est différent de 0, alors un nouveau fichier sera créé toutes les n lignes. (ndt, attention aux conventions de nommages des fichiers si vous utilisez cette option conjointement aux inclusion des numéros d'étapes ou de partitions dans les noms de fichiers) |

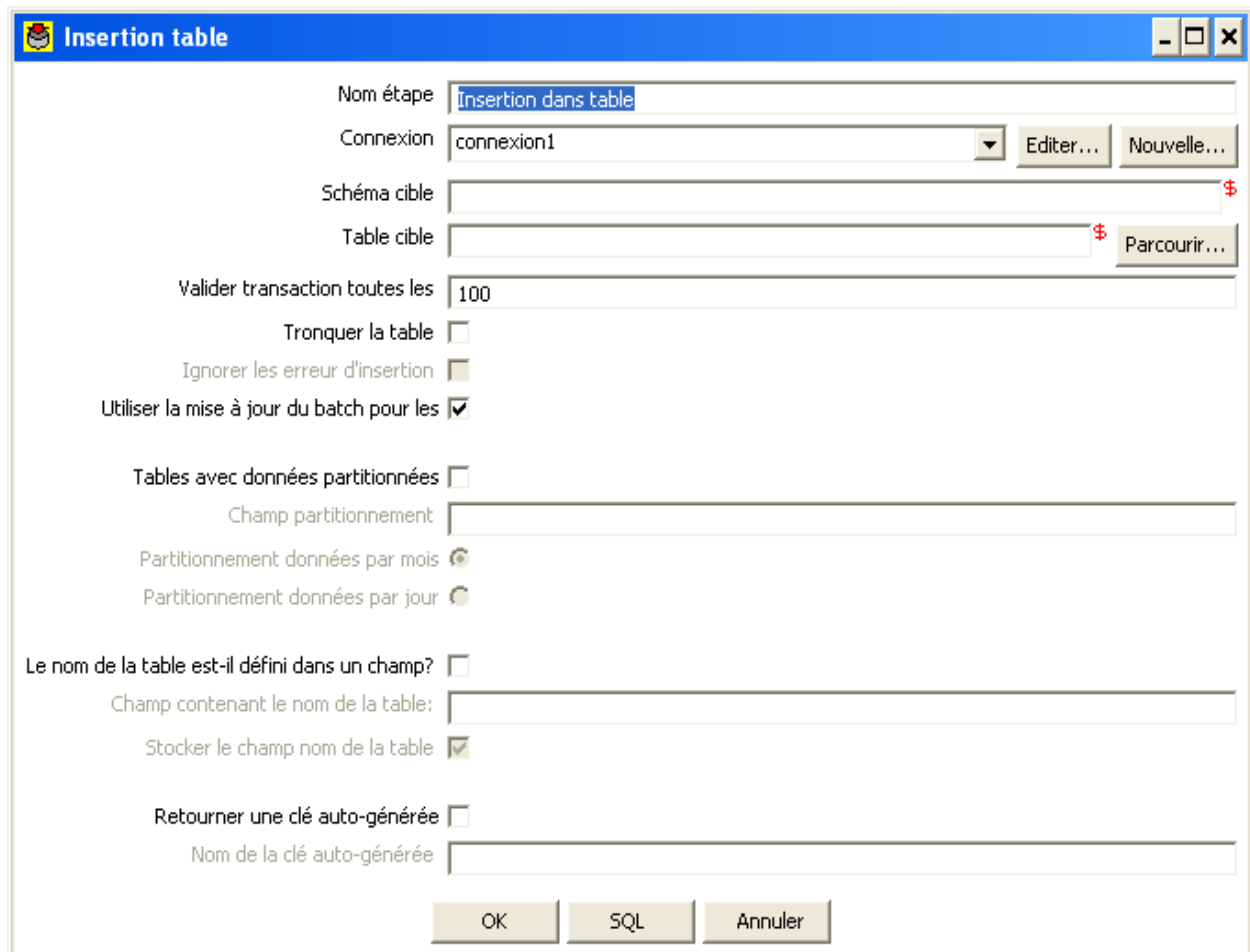
| Paramètre | Description |
|---------------------------------|---|
| Ajouter ligne en fin de fichier | Ce paramètre vous permet de saisir une fin alternative pour le fichier. |
| Champs | <ul style="list-style-type: none"> Nom : nom du champ Type : Type du champ Format : voir Formats pour une description complète des formats Longueur : <ul style="list-style-type: none"> Nombre : le nombre des chiffres significatifs dans le nombre; Chaîne de caractères : le nombre de caractères; Date : longueur de la date imprimée (si 4, il y aura seulement l'année). Précision : <ul style="list-style-type: none"> Nombre r: le nombre de décimales après la virgule; Chaîne de caractères : non utilisé; Date : non utilisé. Devise : symbole utilisé pour représenter des devises comme \$10,00 ou € Décimal : le symbole utilisé pour délimiter la partie entière et décimale "." (10,00) ou "," () Groupe : le symbole utilisé pour grouper les chiffres d'un grand nombre par trois "," (10,00) ou "." () Null: Si la valeur d'un champ est nulle, ce paramètre permet de préciser la chaîne de caractères à insérer dans le flux de sortie |

10.4.10.5. Extras

| Fonction/Bouton | Description |
|---------------------------|--|
| Afficher noms de fichiers | <p>Ce bouton permet de récupérer la liste de fichiers qui seront générés par cette étape.</p> <p>NOTE: Ceci est une simulation et peut dépendre de données que spoon ne connais pas à ce moment là (nombre de lignes, etc.)</p> |
| Largeur minimale | <p>Ce bouton permet de minimiser la largeur des champs pour que la longueur de ceux-ci soit minimale. Par exemple, si vous disposez d'un nombre comme 0000001, on a envie d'écrire 1, etc.</p> <p>Les champs en sortie ne seront alors plus aussi longs que l'information en entrée.</p> |

10.4.11. Insertion dans table (ndt en anglais Table output)

10.4.11.1. Captures d'écran



Insertion table

Nom étape: Insertion dans table

Connexion: connexion1 [Editer...] [Nouvelle...]

Schéma cible: [] \$

Table cible: [] \$ [Parcourir...]

Valider transaction toutes les: 100

Tronquer la table: ☐

Ignorer les erreur d'insertion: ☐

Utiliser la mise à jour du batch pour les: ☒

Tables avec données partitionnées: ☐

Champ partitionnement: []

Partitionnement données par mois: ☐

Partitionnement données par jour: ☐

Le nom de la table est-il défini dans un champ?: ☐

Champ contenant le nom de la table: []

Stocker le champ nom de la table: ☒

Retourner une clé auto-générée: ☐

Nom de la clé auto-générée: []

[OK] [SQL] [Annuler]

10.4.11.2. Icône



10.4.11.3. Description générale

Cette étape vous permet d'enregistrer un flux de données dans une table de base de données.

10.4.11.4. Paramètres

| Paramètres | Description |
|---|---|
| Nom de l'étape | Ce nom doit être unique au sein de la même transformation. |
| Connexion | La connexion de base de données utilisée pour écrire les données. |
| Schéma cible | Le nom du schéma de la table cible des données. Ce paramètre est important si la source de données permet des noms avec des points `.`. |
| Table cible | Le nom de la table cible des données. (ndt Un bouton « parcourir » vous permet de récupérer les tables visibles depuis la connexion spécifiée) |
| Valider transaction toutes les | <p>Spoon utilise une transaction pour insérer les lignes dans la table. Celle-ci est validée toutes les N lignes si N n'est pas égal à 0. Si ce paramètre n'est pas renseigné ou négatif ou nul, Spoon n'utilise pas les transactions (plus lent)</p> <p>NOTE: <i>Les transactions ne sont pas supportées par toutes les plate-formes.</i></p> |
| Tronquer la table | Activez ce paramètre si vous désirez que la table soit tronquée avant que la première ligne soit insérée. |
| Ignorer les erreurs d'insertion | Ce paramètre activé fait que le système ignore toutes les erreurs d'insertion que peut lui renvoyer le serveur de bases de données, comme « clé primaire déjà existante ». Cependant, un maximum de 20 avertissements seront journalisés. Ce paramètre n'est pas utilisé dans le cas de traitement « batch ». |
| Utiliser une mise à jour différée pour les insertions | Ce paramètre est activé par défaut. Cela permet d'effectuer les requêtes d'insertion par lot en différé, ce qui est plus rapide. |
| Tables avec données partitionnées | Activer cette option afin de séparer les données dans plusieurs tables. Par exemple, lorsque vous désirez insérer des données de ventes annuelles dans des tables de ventes mensuelles SALES_200510, SALES_200511, SALES_200512, ... |

| Paramètres | Description |
|--|--|
| | <p>SI vous utilisez ce paramètre sur un système non partitionné et/ou qui ne permet pas les insertions multiples réalisera une requête « UNION ALL » de toutes les vues dans la table maître des tables partitionnées.</p> <p>La vue SALES vous permet de voir l'ensemble des ventes :</p> <pre> CREATE OR REPLACE VIEW SALES AS SELECT * FROM SALES_200501 UNION ALL SELECT * FROM SALES_200502 UNION ALL SELECT * FROM SALES_200503 UNION ALL SELECT * FROM SALES_200504 </pre> |
| Le nom de la table est-il défini dans un champ ? | <p>Utilisez ce paramètre afin de séparer les données dans une ou plusieurs tables. Le nom de la table cible est défini dans le champ saisi.</p> <p>Par exemple, si vous enregistrez les données des clients selon le genre (féminin ou masculin, F ou M), les données seront enregistrées dans deux tables finissant par F ou M. Il y a un paramètre pour exclure le champ contenant le nom de la table dans les tables partitionnées.</p> |
| Retourner une clé auto-générée | Activez ce paramètre si vous désirez que la clé soit auto-générée à chaque insertion de ligne dans la table. |
| Nom de la clé auto-générée | Précisez ici le nom du nouveau champ dans le flux de sortie qui contiendra la clé auto-générée. |

10.4.11.5. Extras

| Fonction/Bouton | Description |
|-----------------|---|
| SQL | Ce bouton permet de générer l'instruction SQL pour créer automatiquement la table liée au flux de sortie. |
| Vérification | Ce bouton permet de vérifier si tous les champs sont valides dans la table cible et vice-versa. |

10.4.12. Insertion/Mise à jour Table (ndt en anglais Insert / Update)

10.4.12.1. Captures d'écran

Insertion / Mise à jour

Nom étape: Insertion / Mise à jour table

Connexion: [dropdown] [Editer...] [Nouvelle...]

Schéma cible: [text box]

Table cible: Table de recherche [Parcourir...]

Valider transaction toutes les: 100

Ne pas effectuer de mise à jour: ☐

Les clés de recherche:

| ▲ | Champ table | Comparateur | Champ flux1 | Champ flux2 | Récupérer champs |
|---|-------------|-------------|-------------|-------------|------------------|
| 1 | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

Champs mis à jour:

| ▲ | Champ table | Champ flux | Mise à jour | Récupérer champs mis à jour |
|---|-------------|------------|-------------|-----------------------------|
| 1 | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

[OK] [SQL] [Annuler]

10.4.12.2. Icône



10.4.12.3. Description générale

Cette étape réalise une insertion ou mise à jour selon le contexte. Si la ligne du flux d'entrée n'existe pas dans la table cible, alors elle est insérée. Si la ligne est trouvée et que les champs à mettre à jour sont les mêmes, alors rien n'est fait. Enfin, si tous les champs à mettre à jour ne sont pas les mêmes, alors la ligne est mise à jour sur les champs qui ont été précisés (ndt comprenez bien ce fonctionnement car il peut paraître non naturel selon les cas)

10.4.12.4. Options

- **Nom de l'étape** : Ce nom doit être unique au sein d'une transformation.
- **Connexion** : nom de la connexion à la base de données contenant la table cible.
- **Schéma cible** : nom du schéma de la table cible.
- **Table cible** : nom de la table cible.

- **Valider la transaction toutes les** : nombre de lignes à changer (insertion/mise à jour) avant de réaliser une instruction de confirmation (ndt en anglais commit)
- **Ne pas effectuer de mise à jour** : Si ce paramètre est activé, alors seules les insertions seront réalisées.
- **Clés de recherche** : Vous pouvez ici préciser une liste de valeurs de champs et des opérateurs de comparaison. Vous pouvez utiliser les opérateurs suivants : =, <>, <, <=, >, >=, LIKE, BETWEEN, IS NULL, IS NOT NULL
- **Champs mis à jour** : Précisez tous les champs de la tables que vous désirez insérer/mettre à jour en n'oubliant pas d'inclure les clés. Veuillez noter que vous pouvez éviter la mise à jour de certains champs en saisissant « N » dans la colonne « mise à jour ».

10.4.12.5. Extra

- **bouton "Récupérer champs"** : ce bouton permet de récupérer les nom des champ depuis le flux de données en entrée.
- **bouton "Récupérer champs mis à jour"** : ce bouton permet de récupérer les champs mis à jour depuis le flux d'entrée.
- **bouton "Valider"** : ce bouton permet de valider si les champs en entrée devant être mis à jour sont présents dans la table cible.
- **bouton "SQL"** : ce bouton permet de générer l'instruction SQL de cr »ation de la table cible et de ses index.

10.4.13. Mise à jour Table (ndt en anglais Update)

10.4.13.1. Captures d'écran

Mise à jour

Nom étape:

Connexion:

Schéma cible:

Table cible:

Valider transaction toutes les:

Ignorer la recherche infructueuse? ☐ Champ flag (clé trouvée)

Clé(s) de recherche de valeurs:

| | Champ table | Comparateur | Champ flux2 | Champ flux2 |
|---|-------------|-------------|-------------|-------------|
| 1 | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

Champs mis à jour:

| | Champ table | Champ flux |
|---|-------------|------------|
| 1 | | |
| | | |

10.4.13.2. Icône



10.4.13.3. Description générale

Cette étape est identique à l'étape **Insertion/Mise à jour Table** si ce n'est que l'insertion n'est jamais faite dans cette étape. SEULE la mise à jour est réalisée.

10.4.14. Suppression dans Table (ndt en anglais Delete)

10.4.14.1. Captures d'écran

Nom étape: Suppression dans table

Connexion: [dropdown] [Editer...] [Nouvelle...]

Schéma cible: [empty]

Table cible: table de recherche [Parcourir...]

Valider transaction toutes les: 100

Clé(s) de recherche:

| | Champ table | Comparateur | Champ flux1 | Champ flux2 | |
|---|-------------|-------------|-------------|-------------|--|
| 1 | | | | | |

[Récupérer champs]

[OK] [Annuler]

10.4.14.2. Icône

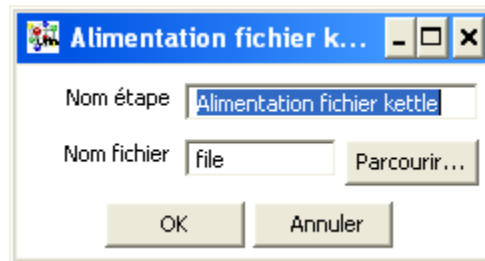


10.4.14.3. Description générale

Cette étape est identique à l'étape [Mise à jour Table](#) si ce n'est qu'au lieu d'être mises à jour, les lignes sont supprimées.

10.4.15. S rialiser dans un fichier (ndt en anglais Serialize to file formerly Cube File Output)

10.4.15.1. Captures d cran



10.4.15.2. Ic ne



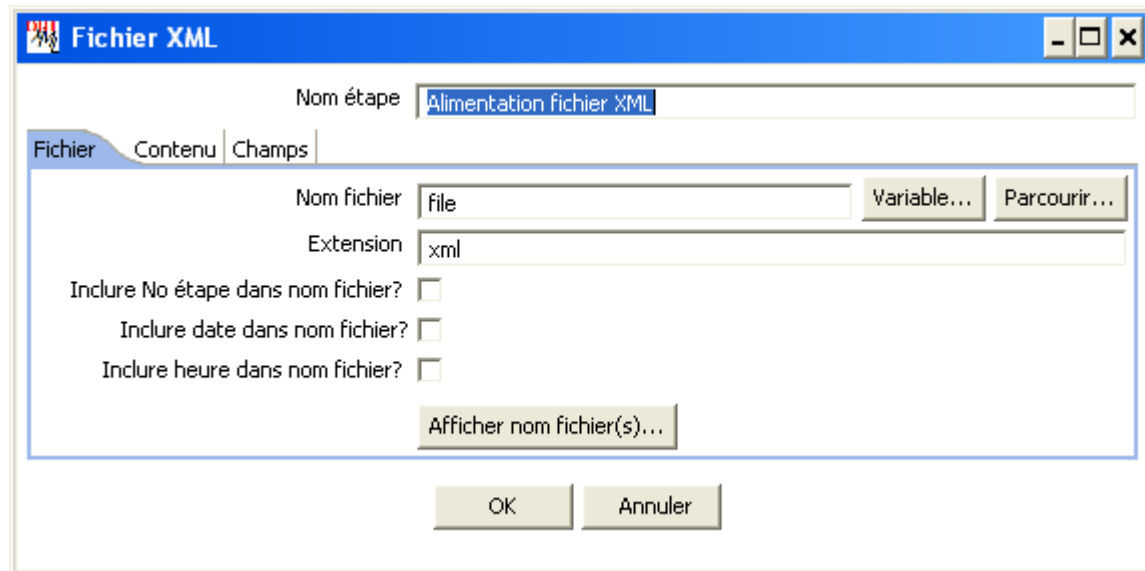
10.4.15.3. Description g n rale

Cette  tape permet d'enregistrer des lignes sous une forme binaire dans un fichier. L'avantage par rapport   un fichier texte   plat (ndt en anglais flat), c'est que le contenu n'a pas besoin d' tre analys    la relecture. Les donn es sont compr hensibles directement par Kettle. C'est pourquoi les meta-data sont stock es sous forme binaire en interne dans Kettle.

NOTE: Cette  tape ne devrait probablement pas  tre utilis e comme outil de construction de cube, car ce format n'est pas du tout standard, mais seulement interne   Kettle.

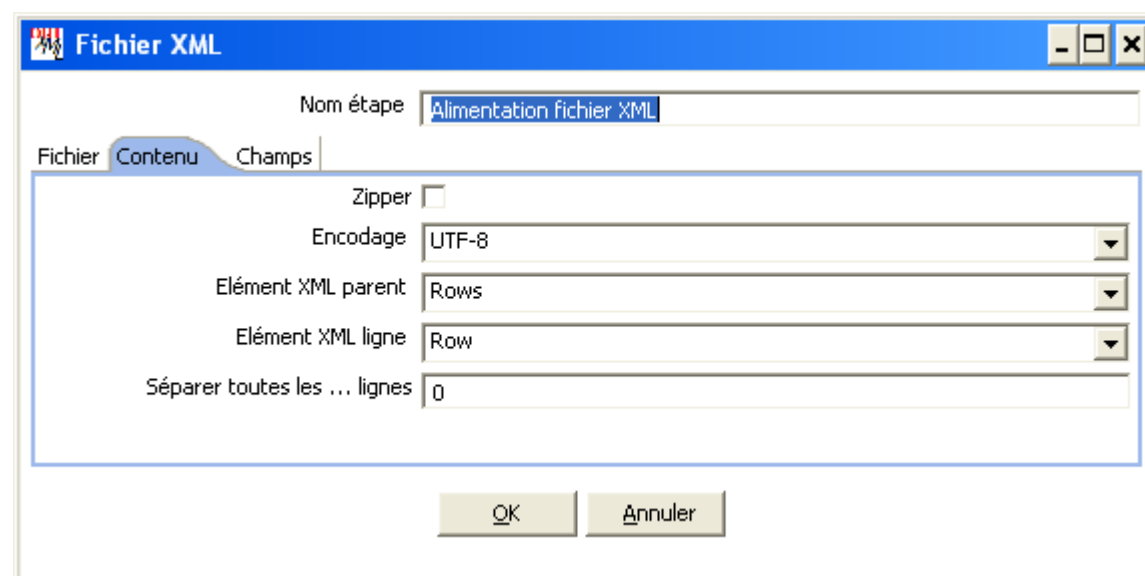
10.4.16. XML Output

10.4.16.1. Captures d'écrans



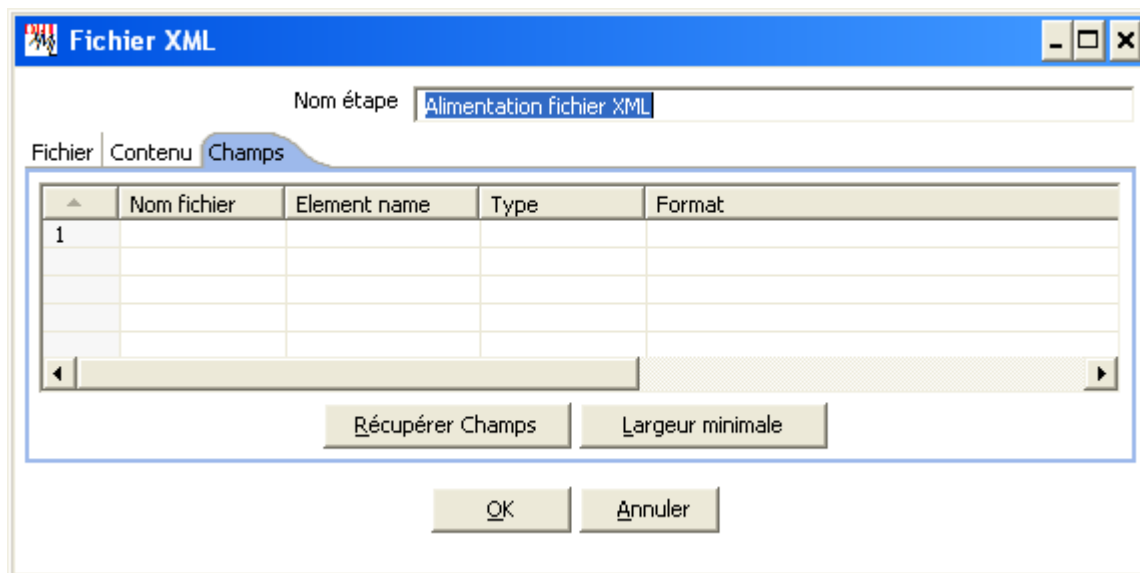
The screenshot shows the 'Fichier XML' dialog box with the 'Fichier' tab selected. The 'Nom étape' field contains 'Alimentation fichier XML'. The 'Nom fichier' field contains 'file', and the 'Extension' field contains 'xml'. There are three checkboxes: 'Inclure No étape dans nom fichier?' (unchecked), 'Inclure date dans nom fichier?' (unchecked), and 'Inclure heure dans nom fichier?' (unchecked). A button 'Afficher nom fichier(s)...' is located below the checkboxes. At the bottom are 'OK' and 'Annuler' buttons.

l'onglet "Fichier"



The screenshot shows the 'Fichier XML' dialog box with the 'Contenu' tab selected. The 'Nom étape' field contains 'Alimentation fichier XML'. The 'Zipper' checkbox is unchecked. The 'Encodage' dropdown menu is set to 'UTF-8'. The 'Élément XML parent' dropdown menu is set to 'Rows'. The 'Élément XML ligne' dropdown menu is set to 'Row'. The 'Séparer toutes les ... lignes' field contains '0'. At the bottom are 'OK' and 'Annuler' buttons.

l'onglet "Contenu"



l'onglet "Champs"

10.4.16.2. Icône



10.4.16.3. Description générale

Cette étape vous permet d'écrire les lignes de n'importe quelle source dans un ou plusieurs fichiers au format XML.

10.4.16.4. Paramètres

| Paramètres | Description |
|--|--|
| Nom de l'étape | Nom de l'étape. Il doit être unique au sein d'une transformation. |
| Nom de fichier | Ce paramètre permet de spécifier le nom et le chemin de fichier de sortie. |
| Extension | Ce paramètre permet d'ajouter automatiquement l'extension saisie à la fin du nom du fichier (.xml). |
| Inclure le numéro d'étape dans le nom du fichier | SI vous exécutez l'étape en plusieurs copies () le numéro de la copie peut être inclus dans le nom du fichier de sortie en activant ce paramètre, avant |

| | |
|--|---|
| | l'extension (_0). |
| Inclure la date dans le nom du fichier | Ce paramètre permet d'inclure la date système dans le nom du fichier de sortie. (_20041231). |
| Inclure l'heure dans le nom de fichier | Ce paramètre permet d'inclure l'heure système dans le nom du fichier de sortie. (_235959). |
| Séparer toutes les n lignes | Le nombre maximum de lignes à inclure dans un fichier avant d'en créer un nouveau. |
| Compression | <p>Ce paramètre vous permet de compresser le fichier (zip ou gzip)</p> <p>NOTE: <i>Pour le moment, seulement un fichier peut être placé dans l'archive compressée.</i></p> |
| Encodage | <p>Précisez ici l'encodage de caractère à utiliser dans le fichier de sortie. Laissez à vide ce paramètre pour utiliser l'encodage par défaut de l'environnement. Pour utiliser l'Unicode, saisissez UTF-8 or UTF- A la première utilisation, Spoon cherchera dans l'environnement quels sont les encodages possibles.</p> <p>Ce code sera inscrit dans l'en-tête xml du fichier.</p> |
| Élément XML parent | Le nom de l'élément racine du fichier XML. |
| Élément XML ligne | Le nom de l'élément associé à une ligne dans le fichier XML. |
| Champs | <ul style="list-style-type: none"> • Nom : nom du champ • Type : Type du champ • Format : voir Formats pour une description complète des formats • Longueur : |

- Nombre : le nombre des chiffres significatifs dans le nombre;
 - Chaîne de caractères : le nombre de caractères;
 - Date : longueur de la date imprimée (si 4, il y aura seulement l'année).
- Précision :
 - Nombre r: le nombre de décimales après la virgule;
 - Chaîne de caractères : non utilisé;
 - Date : non utilisé.
- Devise : symbole utilisé pour représenter des devises comme \$10,00 ou €
- Décimal : le symbole utilisé pour délimiter la partie entière et décimale "." (10,00) ou "," ()
- Groupe : le symbole utilisé pour grouper les chiffres d'un grand nombre par trois "," (10,00) ou "." ()
- Null: Si la valeur d'un champ est nulle, ce paramètre permet de préciser la chaîne de caractères à insérer dans le flux de sortie

10.4.17. Alimentation fichier Excel (ndt en anglais Excel Output)

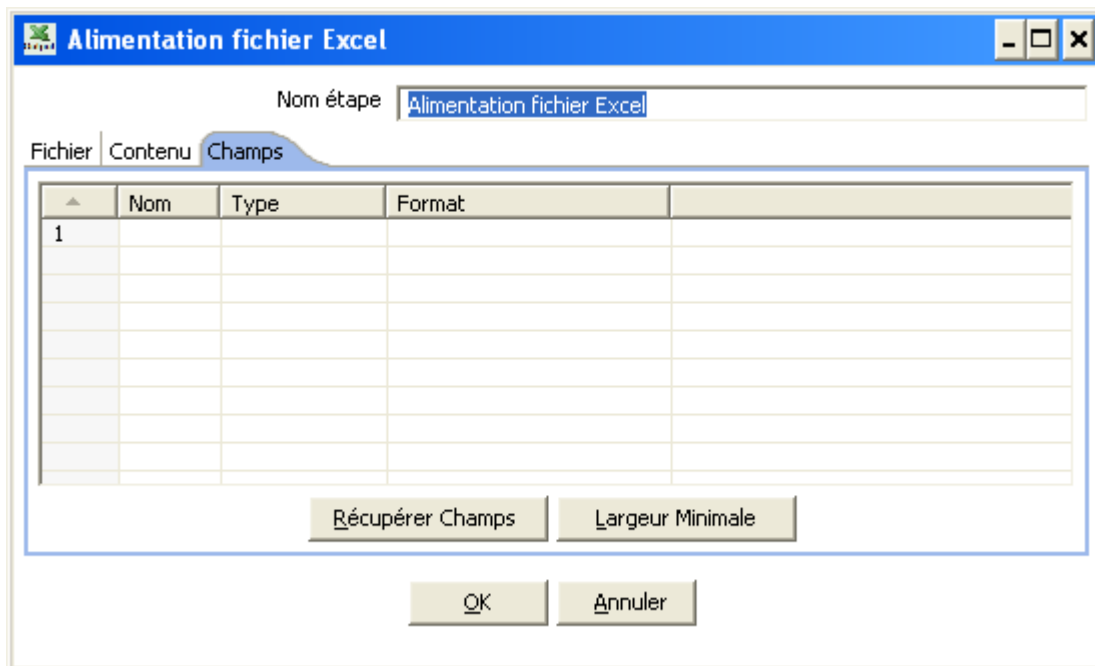
10.4.17.1. Captures d'écrans

The screenshot shows the 'Alimentation fichier Excel' dialog box with the 'Fichier' tab selected. The 'Nom étape' field contains 'Alimentation fichier Excel'. The 'Fichier' tab is active, showing fields for 'Nom fichier' (set to 'file') and 'Extension' (set to 'xls'). There are checkboxes for 'Inclure n°étape dans nom de fichier?', 'Inclure date dans Nom de fichier?', 'Inclure heure dans Nom de fichier?', and 'Protéger la feuille?'. A 'Mot de passe' field is also present. A 'Parcourir...' button is next to the 'Nom fichier' field. A 'Visualiser Nom de fichier(s)...' button is at the bottom. 'OK' and 'Annuler' buttons are at the bottom right.

onglet "Fichier"

The screenshot shows the 'Alimentation fichier Excel' dialog box with the 'Contenu' tab selected. The 'Nom étape' field contains 'Alimentation fichier Excel'. The 'Contenu' tab is active, showing checkboxes for 'En-tête' (checked), 'Fin de fichier', and 'Mettre à jour modèle Excel'. There is a dropdown for 'Encodage' and a text field for 'Créer un fichier tous les n lignes' (set to '0'). There is a checkbox for 'Utiliser un fichier modèle' and a 'Fichier modèle Excel' field (set to 'template.xls') with a 'Parcourir...' button. 'OK' and 'Annuler' buttons are at the bottom right.

onglet "Contenu"



onglet "Champs"

10.4.17.2. Icône



10.4.17.3. Description générale

Cette étape vous permet d'écrire les lignes de n'importe quelle source dans un ou plusieurs fichiers au format Excel. Cela fonctionne sur tous les environnements supportant Kettle, qu'il y ait ou non le logiciel Excel.

10.4.17.4. Paramètres

- **Nom de l'étape** : Nom de l'étape. Il doit être unique au sein d'une transformation.

onglet « Fichier » :

- **Nom de fichier** : Ce paramètre permet de spécifier le nom et le chemin de fichier de sortie, avec le support des variables d'environnement
- **Extension** : Ce paramètre permet d'ajouter automatiquement l'extension saisie à la fin du nom du fichier (.xls).
- **Inclure n° d'étape dans nom de fichier** : Ce paramètre permet d'inclure dans le nom de fichier le numéro d'étape (ceci permet la création en parallèle de fichier)

- **Inclure date dans le nom de fichier** : Ce paramètre permet d'inclure la date système dans le nom du fichier.
- **Inclure heure dans le nom de fichier** : Ce paramètre permet d'inclure l'heure système dans le nom du fichier
- **Protéger la feuille** : Ceci permet de protéger avec un mot de passe la feuille cible du flux de données
- **Mot de passe** : mot de passe qui protège la feuille cible des données.

onglet « Contenu » :

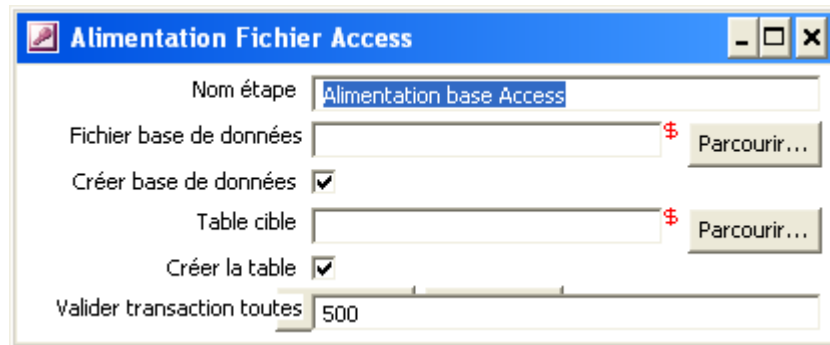
- **En-tête** : Activez ce paramètre si vous désirez ajouter un en-tête au flux de données.
- **Pied de fichier** : Activez ce paramètre si vous désirez ajouter un pied de fichier au flux de données.
- **Encodage** : Précisez ici l'encodage utilisé pour les données dans le tableur. Laissez vide pour utiliser la configuration par défaut de l'environnement.
- **Créer un fichier toutes les n lignes** : Ce paramètre permet de créer un nouveau de fichier à chaque fois que le fichier en sortie dépasse le nombre de lignes saisi (hors en-tête et pied de fichier)
- **Utiliser un fichier modèle** : Activez ce paramètre si vous comptez utiliser un gabarit de fichier en sortie.
- **Fichier modèle excel** : Précisez ici le chemin de votre fichier gabarit.
- **Mettre à jour le fichier modèle** : Activez ce paramètre si le flux de sortie doit ajouter les données au fichier modèle.

onglet « Champs »

- Vous pouvez préciser ici les champs qui devront être exportés dans le fichier au format Excel. Un bouton vous permet de récupérer l'ensemble des champs présents dans le flux en entrée de l'étape.
- Vous pouvez préciser le format des données comme utilisé dans Excel (et non Kettle)

10.4.18. Alimentation base Access (ndt en anglais Microsoft Access Output)

10.4.18.1. Captures d'écran



10.4.18.2. Icône



10.4.18.3. Description générale

Cette étape vous permet d'écrire les lignes de n'importe quelle source dans une base de données au format Access ou de la créer si besoin.

10.4.18.4. Paramètres

- **Nom de l'étape** : Nom de l'étape. Il doit être unique au sein d'une transformation.
- **Nom de fichier** : Ce paramètre permet de spécifier le nom et le chemin de fichier de sortie, avec le support des variables d'environnement
- **Créer bas de données** : Activez ce paramètre pour créer une nouvelle base de données.
- **Table cible** : Précisez ici la table dans laquelle le flux de données sera écrit.
- **Créer la table** : Activez ce paramètre pour créer la table si besoin.
- **Valider les transactions toutes les n lignes** : Définissez ici le nombre de lignes validant l'écriture dans la base de données (ndt en anglais commit size)

10.4.19. Database lookup

Step name: CustomerRegionLookup

Connection: ol_olap

Lookup table: statedim

Enable cache? ☒

Cache size in rows (0=cache everything): 0

The key(s) to look up the value(s):

| | Table field | Comparator | Field1 | Field2 |
|---|-----------------|------------|-----------|--------|
| 1 | state | = | state | |
| 2 | dim_effect_date | <= | update_dt | |
| 3 | dim_expire_date | >= | update_dt | |

Values to return from the lookup table :

| | Field | New name | Default | Type |
|---|-------------|----------|---------|--------|
| 1 | region_name | | | String |

Do not pass the row if the lookup fails ☐

Fail on multiple results? ☐

Order by:

OK Get Fields Get lookup fields Cancel

10.4.19.1. Screen dump

10.4.19.2. Icône

10.4.19.3. Description générale

This step type allows you to look up values in a database table.

10.4.19.4. Options

- **Step name:** Name of the step. This name has to be unique in a single transformation.
- **Connection:** The database connection used to write data to.
- **Lookup table:** The name of the table where we do the lookup.
- **Enable cache:** This option caches database lookups. This means that we expect the database to return the same value all the time for a certain lookup value.

IMPORTANT: if other processes are changing values in the table where you do the lookup, it might be unwise to cache values. However, in all other cases, enabling this option can seriously increase the