# Alimentation

## Normalisation des traitements d’erreurs

En cas d’erreur lors d’un traitement de fichier Excel® par Talend pour l’alimentation, il est important de préciser la marche à suivre :

### Log d’erreur dans un fichier texte (format .txt)

L’inscription de l’erreur se fera dans un fichier créé au début du traitement (pour le traitement d’un fichier .xls(x) de données, on créera un fichier de log d’erreur. Ce fichier se nommera de la sorte :

errorLog\_FILETYPE\_DATE\_ID

FILETYPE : Le FILETYPE (ou type de fichier) est une chaîne de trois caractère définissant le type de traitement auquel correspond le fichier de log d’erreur. Les 3 chaîne possibles (et aucun autre type de saura être accepté) sont :

* *INI pour un fichier d’initialisation*
* *MEN pour un fichier mensuel*
* *ANU pour un fichier annuel*

DATE : la DATE sera à écrire sous la forme définie par la norme ISO 8601 avec les 3 premiers niveaux de granularité à savoir Année, Mois, et Quantième. Pour rappel, la notation de la date avec ces trois données en ISO 8601 se fait de cette façon :

*Pour le 26 Janvier 2011 :* ***2011-01-26***

ID : L’ID (ou identifiant) est un indice incrémental sur deux digits en cas de doublon de traitement dans la journée. Le début d’incrémentation sera **00**, puis sera incrémenté de **1** en cas de nouveau fichier dans la même journée. À noter que l’incrémentation sera à faire lors de la création du fichier : il faudra vérifier l’existence d’un fichier avec le même FILETYPE et la même DATE, et le cas échéant, prendre le dernier ID pour l’incrémenter de **1**

Exemple : Voici un exemple de nom de fichier de log d’erreurs pour le traitement d’un fichier de données *mensuelles*, effectué [le traitement] le *3 février 2011*, alors qu’il y a déjà eu *2 traitements* identiques juste auparavant :

errorLog\_MEN\_2011-02-03\_02.txt

##### Le contenu du fichier sera organisé de la manière suivante :

* Une ligne d’entête
* Une ligne de 5 tirets (caractère Unicode/ASCII **002D**)
* Une liste d’erreurs (une erreur par ligne, les lignes seront terminées par un CRLF (hexadécimal **0D0A**)

Entête - L’entête sera organisé de la manière suivante :

TIME\_FILENAME

TIME : Le TIME représente sous la norme ISO 8601 la date et l’heure de la création du fichier. On utilisera 6 niveaux de granularité (Année, Mois, Quantième, Heure, Minute, Seconde), sachant que les secondes pourront être fixées à 00 sans problème. On utilisera de préférence l’heure UTC, mais pour plus de simplicité, l’heure système peut servir de base, à condition que l’on précise le décalage de fuseau par rapport à UTC. Pour rappel, la notation de la date avec ces trois données en ISO 8601 se fait de cette façon :

*Pour le 26 Janvier 2011 à 12h37 et 45 secondes à Paris :* ***2011-01-26T12:37:45+01:00*** *L’heure suivante correspond au même moment :* ***2011-01-26T11:37:45Z***

FILENAME : Le FILENAME est simplement le nom du fichier traité, ainsi que sons extension. L’écriture se fera de la façon suivante : **NomDuFichier.Extension**

Contenu - Le contenu sera organisé de la façon suivante :

LINE1\_LINE2\_ERRORCODE\_DESCRIPTION

LINE1 : Le numéro de ligne LINE1 correspond au numéro de la ligne à laquelle s’est posé un problème dans le fichier source.

LINE2 : Le numéro de ligne LINE2 correspond au numéro de la ligne à laquelle a été recopié la ligne litigieuse dans le fichier Excel® de contrôle d’erreur.

ERRORCODE : Le code d’erreur ERRORCODE correspond au code fourni par le référentiel d’erreurs du traitement en cours. Il permet de repérer le type d’erreur.   
*À noter que si le ERRORCODE contient un tiret bas (ou underscore : caractère ASCII/Unicode* ***005F****) ce dernier devra être remplacé par un tiret (caractère Unicode/ASCII* ***002D****) pour éviter des problèmes ultérieurs*.

DESCRIPTION : Cette chaîne est libre en caractère hormis CRLF (0D et 0A) et tiret bas (5F). Elle représente la description textuelle de l’erreur précisée par le ERRORCODE.

### Inscription dans un fichier Excel® (format .xls ou si possible .xlsx POINT À CLARIFIER)

Ce fichier Excel® servira de base pour traiter les données qui n’ont pas été intégrées dans la base.

*ATTENTION : dans le cas d’une intégration d’initialisation, les warnings ne seront pas intégrés dans ce fichier .xls ou .xlsx, car les données sont déjà insérées dans la base. On retrouve cependant les warnings, avec leurs codes correspondant dans le fichier .txt de log d’erreurs, pour correction ultérieure.*

Lors des traitements des fichiers Excel® source, pour chaque nouvelle ligne de donnée (que l’on appellera par la suite *enregistrement*) la boucle de traitement analyse la validité des données. En cas de non-conformité des données de l’enregistrement, ce dernier sera ignoré lors de l’insertion en base de données, et sera recopié dans le fichier Excel® cible, ou fichier Excel® de récupération.

Le nommage de ce fichier se fera exactement de la même manière que le nommage du fichier txt précédemment explicité. Seule l’extension changera (extension .xls ou .xlsx **POINT À CLARIFIER**) entre ces deux fichiers.

L’organisation du contenu de ce fichier dépendra de l’organisation du fichier Excel® source, et lui sera semblable en tous points (nombre de feuilles, nombre de colonnes par feuilles, etc).

La première ligne de chaque feuille sera une copie de la première ligne de chaque feuille du fichier Excel® source. On copiera ainsi les titres de colonnes.

Pour la suite, chaque enregistrement (dans le fichier Excel® source qui contiendra des données litigieuses sera recopié sur une nouvelle ligne de la feuille correspondante, dans le fichier Excel® de récupération.

Nous obtiendrons alors une ligne par erreur, et ce dans chaque feuille.