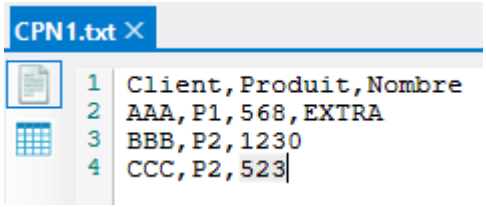
Utiliser Integrator pour nettoyer manuellement les donnÃ©es crÃ©Ã©es

Description

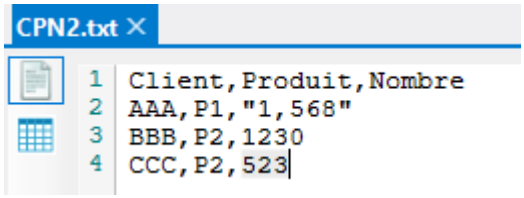
Supposons que nous ayons plusieurs fichiers texte en entrÃ©e dans un script Integrator et n'ayant pas strictement la mÃªme structure. Les donnÃ©es peuvent Ãªtre nettoyÃ©es en modifiant les propriÃ©tÃ©s d'entrÃ©e et en ajoutant des calculs au sein d'Integrator.

Le premier fichier texte a le contenu suivant :



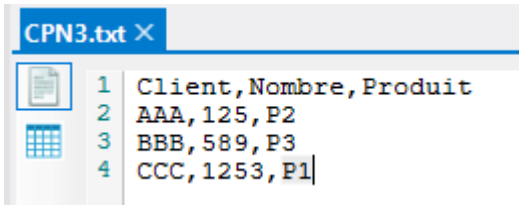
On constate que sur la premiÃ¨re ligne de donnÃ©e, un champ supplÃ©mentaire est prÃ©sent.

Le deuxiÃ¨me fichier texte possÃ¨de le contenu suivant :



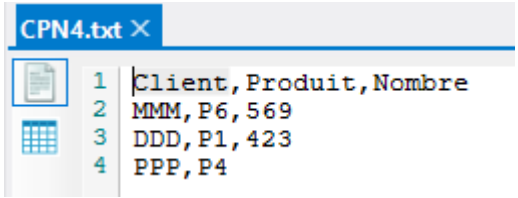
Sur la premiÃ¨re ligne de donnÃ©e, le sÃ©parateur de milliers pour les nombres est la virgule.

Le troisiÃ¨me fichier texte contient les donnÃ©es suivantes :



L'ordre d'apparition des colonnes est diffÃ©rent des deux autres fichiers.

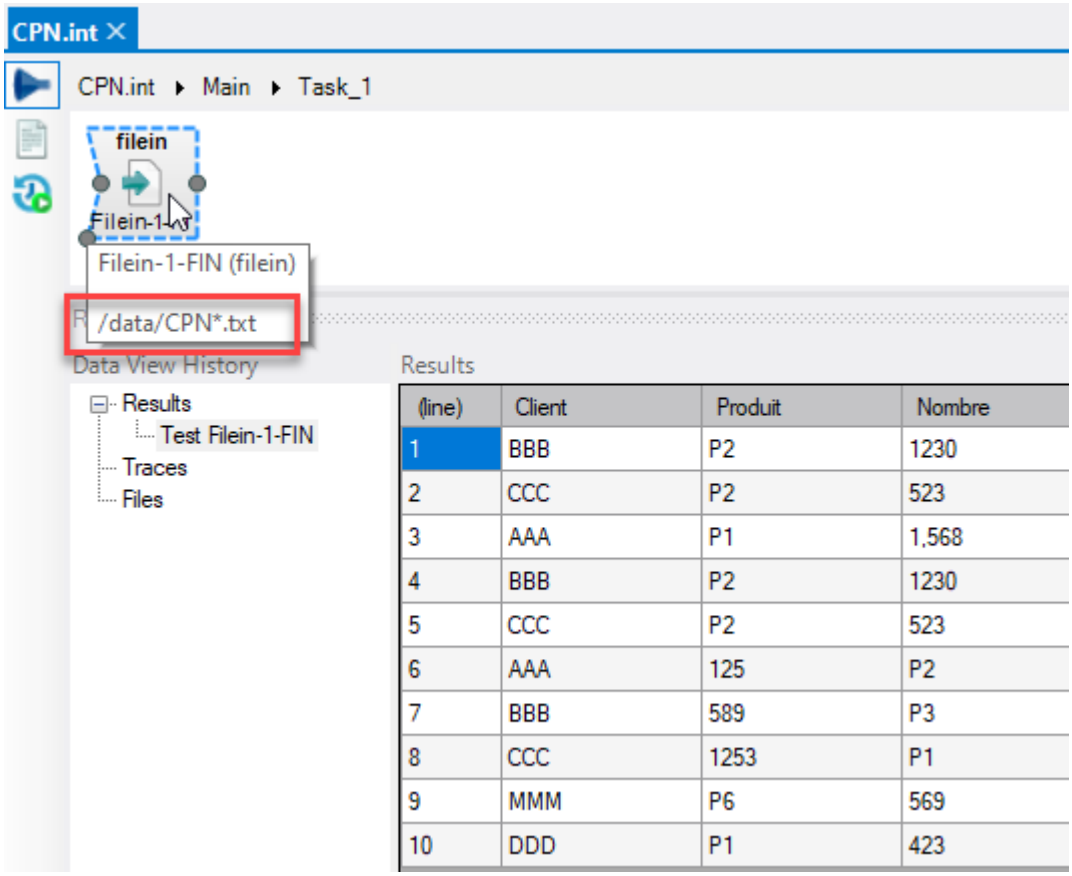
Enfin un quatriÃ¨me fichier est peuplÃ© de la maniÃ¨re suivante :



On constate qu'il manque une valeur de champ sur la derniÃ¨re ligne.

Â© Bimaxis - Tous droits rÃ©servÃ©s

L'utilisation d'un script Integrator permettant la consolidation des donnÃ©es des 4 fichiers prÃ©cÃ©dents donne ceci :



(line)	Client	Produit	Nombre
1	BBB	P2	1230
2	CCC	P2	523
3	AAA	P1	1,568
4	BBB	P2	1230
5	CCC	P2	523
6	AAA	125	P2
7	BBB	589	P3
8	CCC	1253	P1
9	MMM	P6	569
10	DDD	P1	423

On constate que seules ***10*** lignes sont affichÃ©es alors que la somme des lignes des 4 fichiers donne ***12***.

Certaines lignes sont donc ignorÃ©es.

Pour rÃ©tablir les lignes manquantes, la premiÃ¨re modification Ã effectuer est de paramÃ©trer l'option ***Ignore_Extra_Column*** Ã ***true*** au niveau de l'objet d'entrÃ©e Filein.

La deuxiÃ¨me modification Ã effectuer est de paramÃ©trer l'option ***Ignore_Line_End*** Ã ***true*** au niveau de l'objet d'entrÃ©e Filein.

Une fois les options prÃ©citÃ©es modifiÃ©es, un test sur l'objet Filein donne ceci :

Â© Bimaxis - Tous droits rÃ©servÃ©s

Results

(line)	Client	Produit	Nombre
1	AAA	P1	568
2	BBB	P2	1230
3	CCC	P2	523
4	AAA	P1	1,568
5	BBB	P2	1230
6	CCC	P2	523
7	AAA	125	P2
8	BBB	589	P3
9	CCC	1253	P1
10	MMM	P6	569
11	DDD	P1	423
12	PPP	P4	

On a bien maintenant 12 lignes en sortie.

On a cependant un problÃ¨me au niveau des valeurs des champs. Concernant les lignes 7 Ã 9 on a une permutation des valeurs des champs ***Produit*** et ***Nombre***.

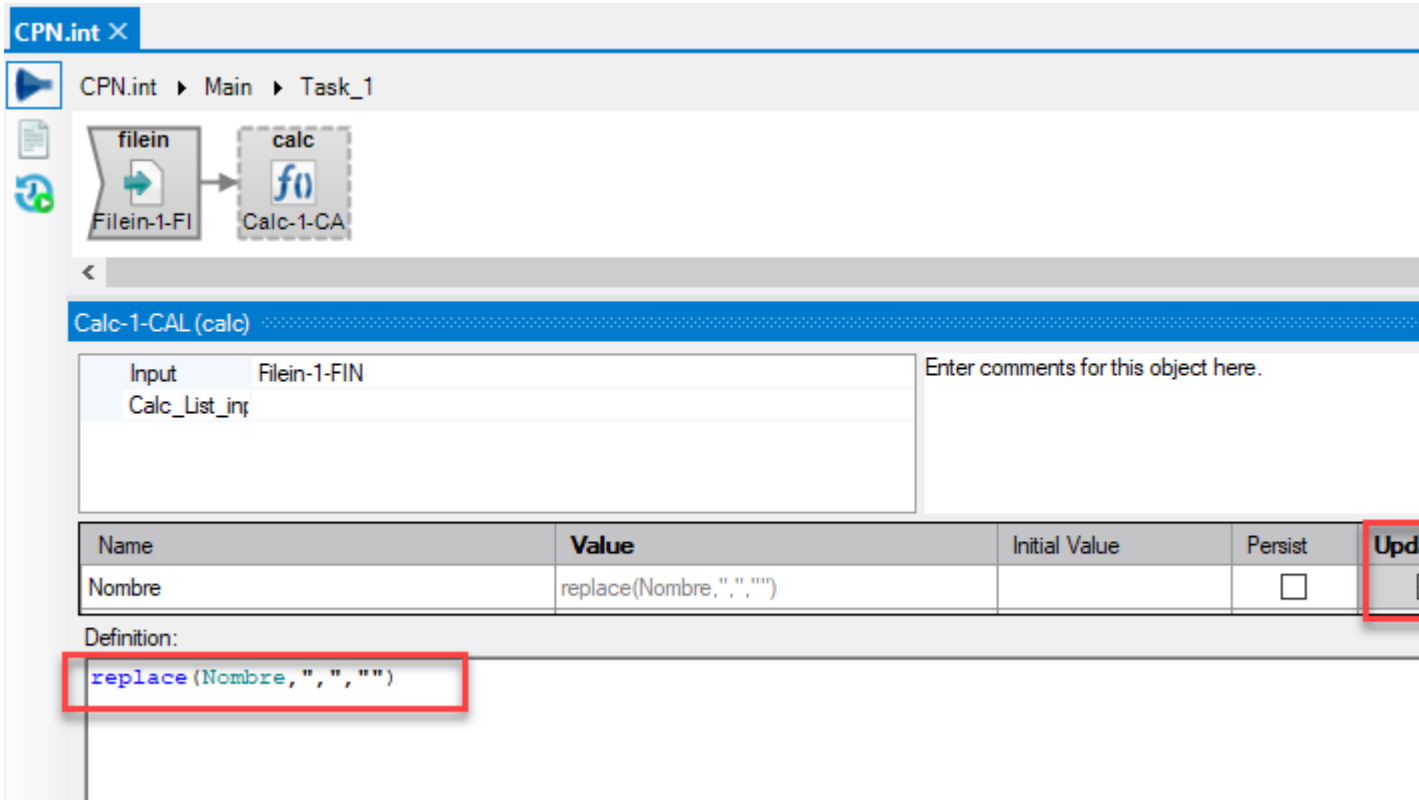
La modification de l'option ***Union*** Ã ***true*** au niveau de l'objet d'entrÃ©e Filein, permet d'obtenir une alimentation correcte des deux champs :

(line)	Client	Produit	Nombre
1	AAA	P1	568
2	BBB	P2	1230
3	CCC	P2	523
4	AAA	P1	1,568
5	BBB	P2	1230
6	CCC	P2	523
7	AAA	P2	125
8	BBB	P3	589
9	CCC	P1	1253
10	MMM	P6	569
11	DDD	P1	423
12	PPP	P4	

Il reste Ã modifier le formatage du nombre sur la ligne 4 en supprimant le sÃ©parateur de milliers pour passer de la valeur ***1.568*** Ã ***1568***.

L'ajout d'un objet Calc au flux de donnÃ©es avec la formule de calcul suivante :

Â© Bimaxis - Tous droits rÃ©servÃ©s



et l'option ***Update*** cochÃ©e donne ceci :

Results

(line)	Client	Produit	Nombre
1	AAA	P1	568
2	BBB	P2	1230
3	CCC	P2	523
4	AAA	P1	1568
5	BBB	P2	1230
6	CCC	P2	523
7	AAA	P2	125
8	BBB	P3	589
9	CCC	P1	1253
10	MMM	P6	569
11	DDD	P1	423
12	PPP	P4	

Tags

1. Data Integrator
2. script
3. Visual Integrator

Â© Bimaxis - Tous droits rÃ©servÃ©s