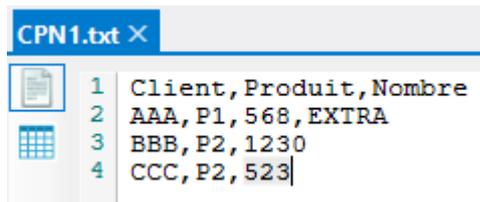


Utiliser Spectre pour nettoyer manuellement les données créées

Description

Supposons que nous ayons plusieurs fichiers texte en entrée dans un script .build de Spectre et n'ayant pas strictement la même structure. Les données peuvent être nettoyées en ajoutant des calculs au sein de Spectre.

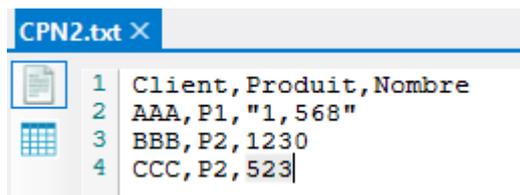
Le premier fichier texte a le contenu suivant :



```
1 Client,Produit,Nombre
2 AAA,P1,568,EXTRA
3 BBB,P2,1230
4 CCC,P2,523|
```

On constate que sur la première ligne de donnée, un champ supplémentaire est présent.

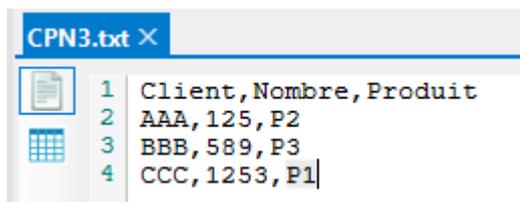
Le deuxième fichier texte possède le contenu suivant :



```
1 Client,Produit,Nombre
2 AAA,P1,"1,568"
3 BBB,P2,1230
4 CCC,P2,523|
```

Sur la première ligne de donnée, le séparateur de milliers pour les nombres est la virgule.

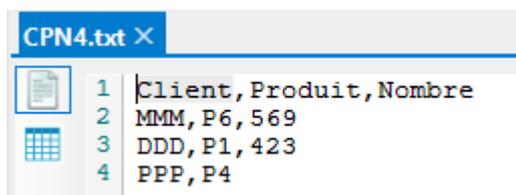
Le troisième fichier texte contient les données suivantes :



```
1 Client,Nombre,Produit
2 AAA,125,P2
3 BBB,589,P3
4 CCC,1253,P1|
```

L'ordre d'apparition des colonnes est différent des deux autres fichiers.

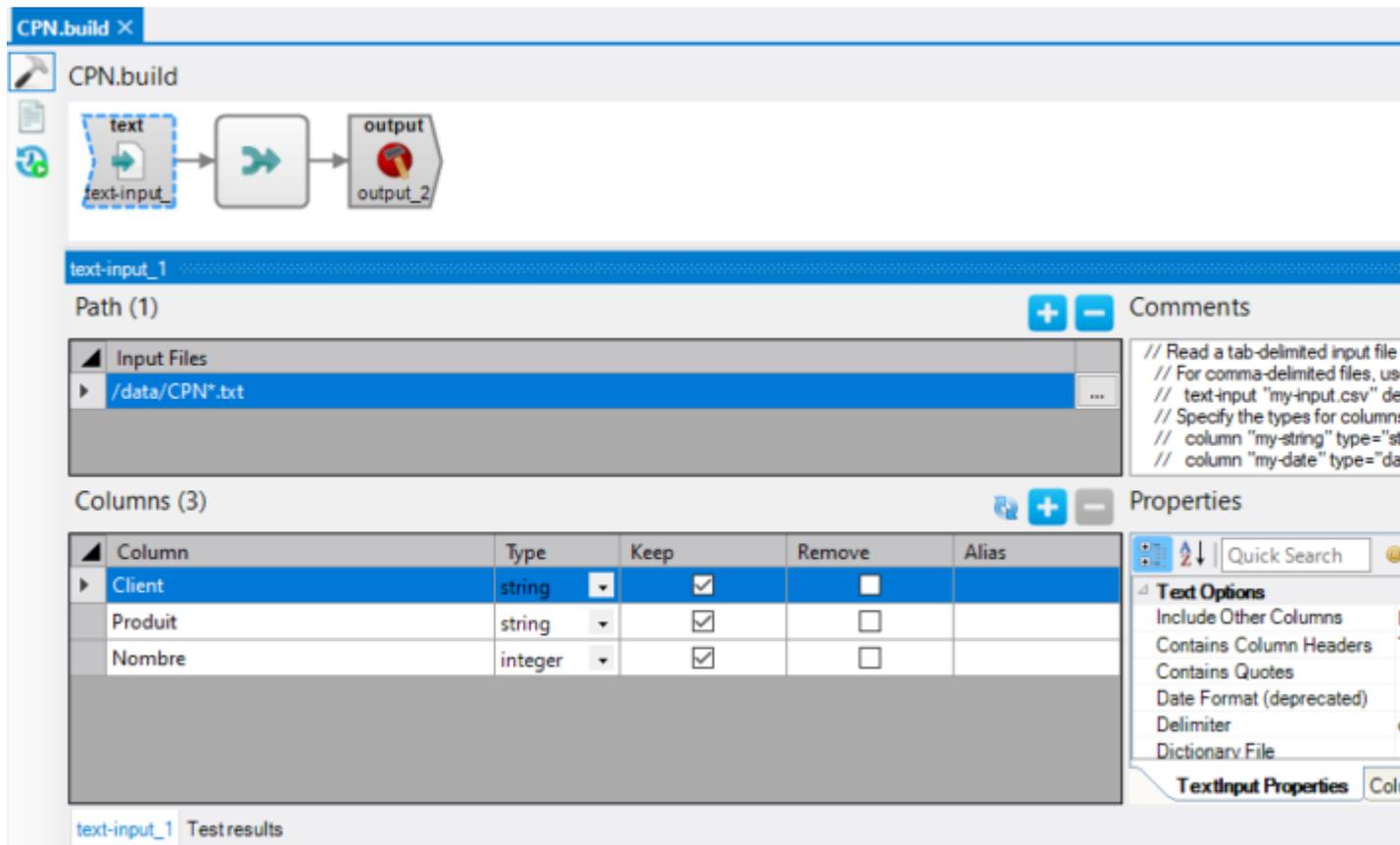
Enfin un quatrième fichier est peuplé de la manière suivante :



```
1 Client,Produit,Nombre
2 MMM,P6,569
3 DDD,P1,423
4 PPP,P4|
```

On constate qu'il manque une valeur de champ sur la dernière ligne.

L'utilisation d'un script build Spectre permettant la consolidation des données des 4 fichiers précédents :



donne ceci pour **Nombre** défini en **Integer** :

Testresults			
(line)	Client	Produit	Nombre
1	AAA	P1	568
2	BBB	P2	1230
3	CCC	P2	523
4	AAA	P1	<unknown>
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	<unknown>

donne ceci pour **Nombre** défini en **Double** ou en **String** :

Test 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	P2	125
8	BBB	P3	589
9	CCC	P1	1253
10	MMM	P6	569
11	DDD	P1	423
12	PPP	P4	<unknown>

On retrouve bien l'intégralité des lignes, cependant il y a un problème d'affichage sur la colonne **Nombre** pour les lignes 4 et 12.

Afin d'avoir un bon formatage du nombre sur la ligne 4, un objet **Calc** dans le flux données avec les propriétés suivantes :

The screenshot shows the CPN.build interface. At the top, a workflow diagram consists of four nodes: 'text' (text-input_1), 'calc' (Calc-1), a connector, and 'output' (output_2). Below this, the 'Calc-1' configuration window is open. It has two panes: 'Available columns (3)' and 'Operations (1)'. The 'Available columns' pane contains a table with columns 'Column Name', 'Type', and 'Source'. The 'Operations' pane contains a table with columns 'Operation' and 'Expression'. The 'Expression' pane on the right shows the formula: `replace(value("Nombre"),",","")`.

Column Name	Type	Source
Client	string	text-inp...
Produit	string	text-inp...
Nombre	string	text-inp...

Operation	Expression
Nombre	replace(value("Nombre"),",","")

Pour obtenir le résultat suivant :

Test 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	<unknown>

Tags

1. script



2. Spectre