

# Compte rendu TP1 ETL

Ahmed Gharbi

## 1. Exercice 1

- Générer 10 lignes avec un champs de test de valeur = 1

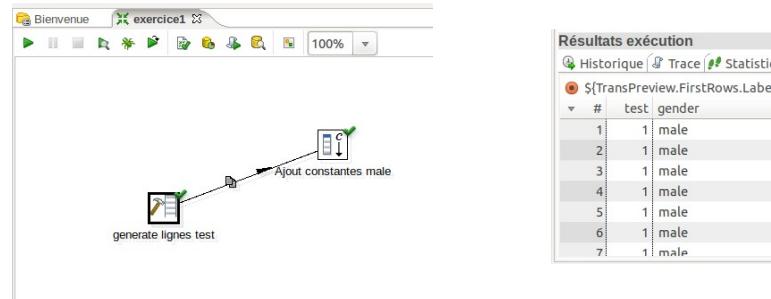
configuration du champs test avec valeur 1.



- Ajouter une colonne ‘gende’ de valeur male.



- Prévisualiser le résultatat



⇒ On remarque que la colonne ‘gender’ est ajoutée à l’ensemble des attributs. L’ajout constante permettant d’ajouter une ou plusieurs constante au flux d’entré.

- Rajouter une colonne ‘cola’ dans la génération de lignes de valeur test.

The screenshot shows the 'Générer lignes' (Generate Rows) configuration dialog and its results.

**Configuration Dialog:**

- Nom étape:** generate lignes test
- Limite:** 10
- Never stop generating rows:**
- Interval in ms (delay):** 5000
- Current row time field name:** now
- Previous row time field name:** FiveSecondsAgo

**Champs:**

#	Nom	Type	Format	Longueur	Précision	Devise	Décimal	Groupe	Valeur	Mettre chaîne vide
1	test	Integer							1	N
2	cola	String							test	N

**Résultats exécution:**

#	test	cola	gender
1	1	test	male
2	1	test	male
3	1	test	male
4	1	test	male
5	1	test	male
6	1	test	male
7	1	test	male

⇒ lorsqu'on à rajouter une colonne 'cola' dans la génération de lignes on aperçoit les trois colonnes 'test' et 'cola' et 'gender' ,ajouter par l'ajout constante, dans la prévisualisation.

- Rajouter un checksum après la génération de ligne :

- De type CRC32
- Récupérer les champs en entrée
- Le champ de sortie sera checksum

The screenshot shows a Talend job flow and its preview results.

**Job Flow:**

```

graph LR
    G1[generate lignes test] --> A[Ajout constantes male]
    A --> G2[generate lignes test]
    G2 --> A2[Ajout checksum]
    A2 --> G3[generate lignes test]
    G3 --> A3[Ajout checksum]
    A3 --> D[Données prévisualisées]
  
```

**Données prévisualisées:**

#	test	cola	gender	checksum
1	1	test	male	1435361449
2	1	test	male	1435361449
3	1	test	male	1435361449
4	1	test	male	1435361449
5	1	test	male	1435361449
6	1	test	male	1435361449
7	1	test	male	1435361449
8	1	test	male	1435361449
9	1	test	male	1435361449
10	1	test	male	1435361449

⇒ cette étape calcule les sommes de contrôle pour un ou plusieurs champs dans le flux d'entrée et l'ajoute à la sortie en tant que nouveau champ.

- Numérotez les lignes

The screenshot shows a Talend job flow and its preview results.

**Job Flow:**

```

graph LR
    G1[generate lignes test] --> A[Ajout constantes male]
    A --> G2[generate lignes test]
    G2 --> A2[Ajout checksum]
    A2 --> A3[Ajout séquence]
    A3 --> D[Données prévisualisées]
  
```

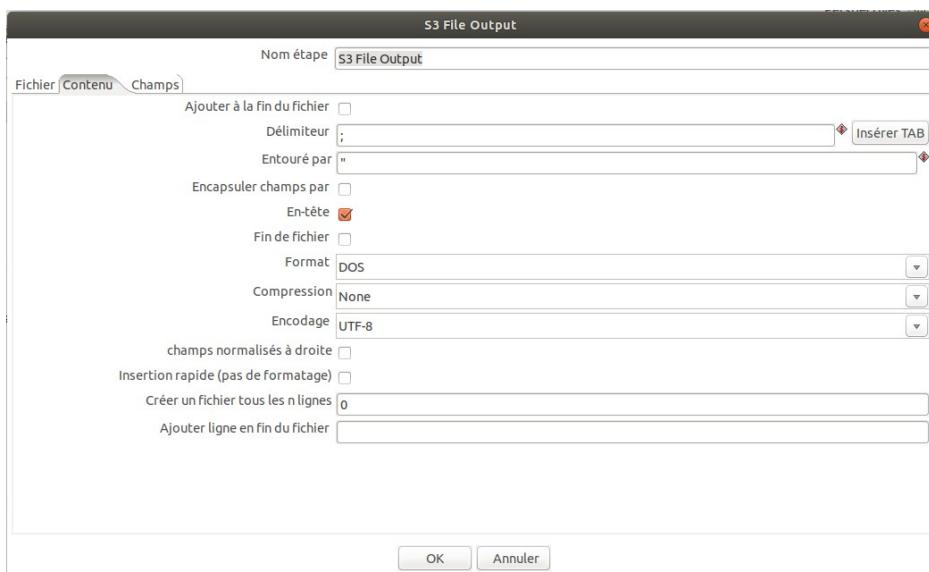
**Données prévisualisées:**

#	test	cola	gender	checksum	valuename
1	1	test	male	1435361449	1
2	1	test	male	1435361449	2
3	1	test	male	1435361449	3
4	1	test	male	1435361449	4
5	1	test	male	1435361449	5
6	1	test	male	1435361449	6
7	1	test	male	1435361449	7
8	1	test	male	1435361449	8
9	1	test	male	1435361449	9
10	1	test	male	1435361449	10

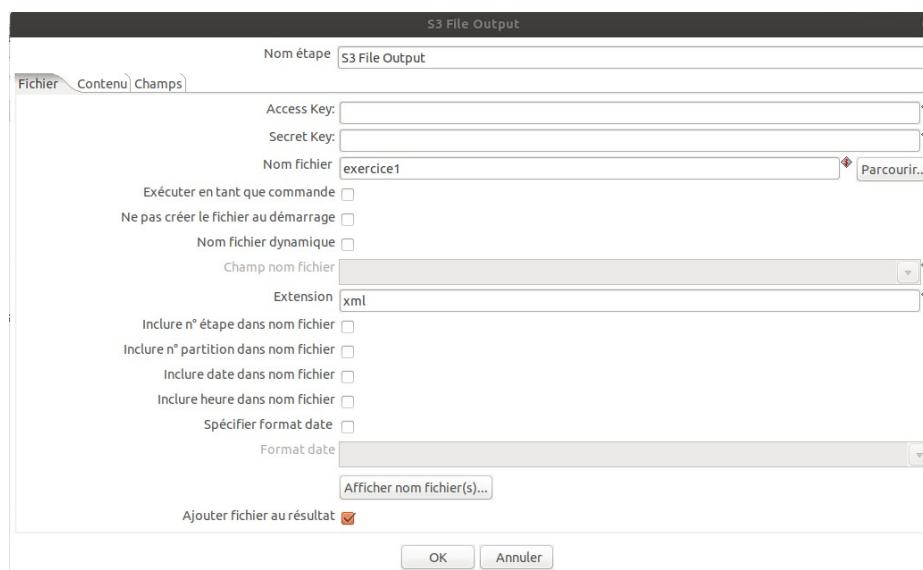
- ✓ L'étape Ajouter une séquence ajoute une séquence au flux.
- ✓ Une séquence est une valeur entière en constante évolution avec une valeur de début et d'incrémentation spécifique. Vous pouvez soit utiliser une séquence de base de données pour déterminer la valeur de la séquence, soit la faire générer.
- ✓ Les valeurs de séquence générées sont uniques lorsqu'elles sont utilisées dans la même transformation et reviennent à la même valeur de départ chaque fois que vous exécutez la transformation.

- Exporter le résultat dans un fichier XML

On export le résultat avec les configuration représentées dans les deux figures suivantes :

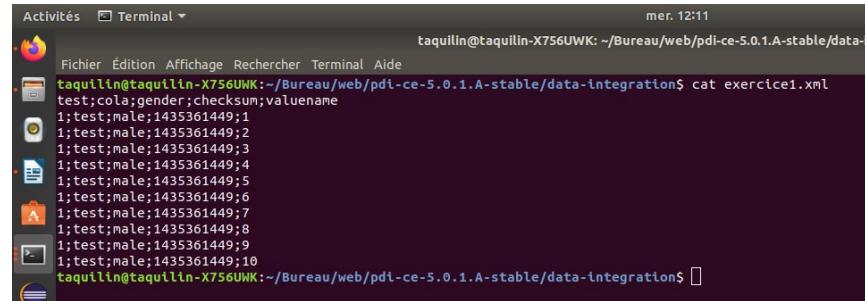


- ➔ format UTF-8
- ➔ Délimiteur ;



## → extension : XML

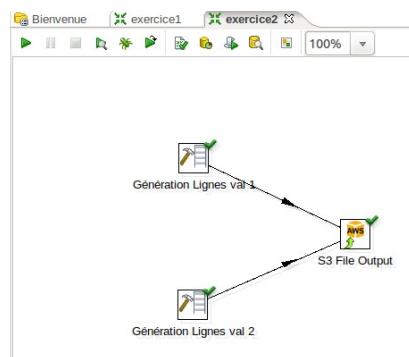
⇒ à la fin on aura le résultat suivant lors de la lecture du fichier XML



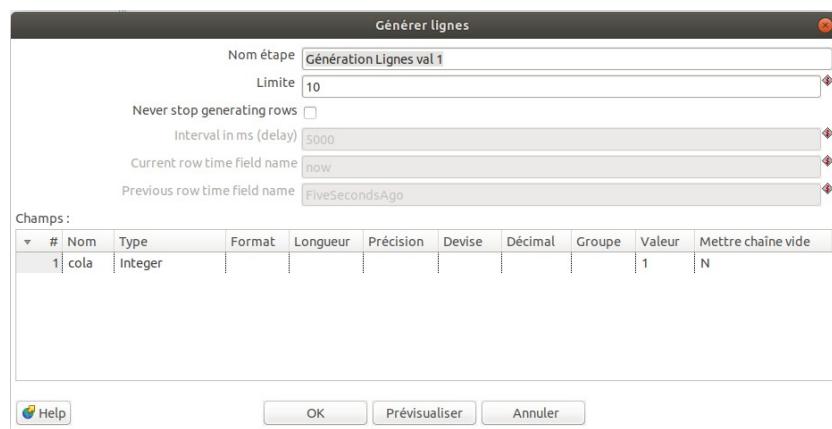
```
Fichier Édition Affichage Rechercher Terminal Aide
taquillin@taquillin-X756UWK:~/Bureau/web/pdi-ce-5.0.1.A-stable/data-integration$ cat exercice1.xml
test;male;1435361449;1
1;test;male;1435361449;2
1;test;male;1435361449;3
1;test;male;1435361449;4
1;test;male;1435361449;5
1;test;male;1435361449;6
1;test;male;1435361449;7
1;test;male;1435361449;8
1;test;male;1435361449;9
1;test;male;1435361449;10
taquillin@taquillin-X756UWK:~/Bureau/web/pdi-ce-5.0.1.A-stable/data-integration$
```

## 2. Exercice 2

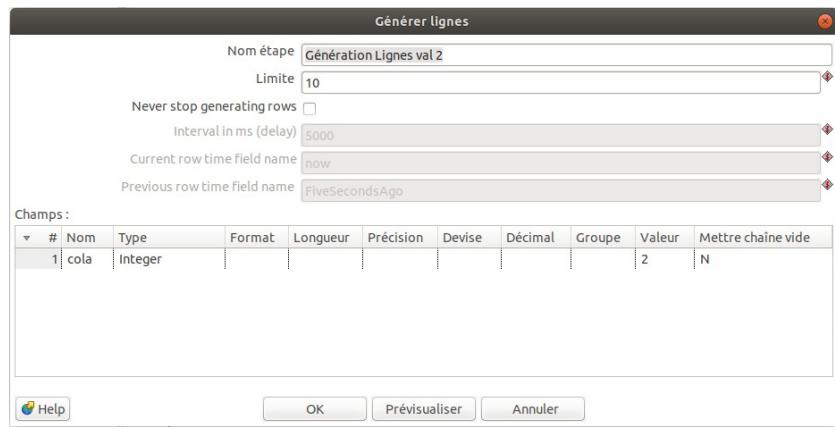
- Créez un fichier CSV contenant une colonne **cola** dont les 10 premières valeurs valent 1 et les 10 suivantes valent 2



Configuration De « génération lignes val 1 » 10 lignes avec une valeur 1



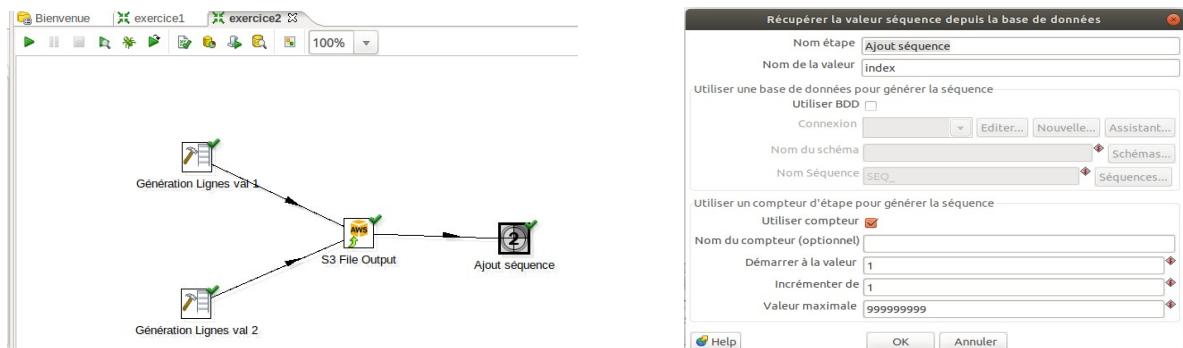
Configuration De « génération lignes val 2 » 10 lignes avec une valeur 2



## Visualisation du résultat

Données prévisualisées		
Lignes de l'étape: S3 File Output (20 lignes)		
#	cola	
6	1	
7	1	
8	1	
9	1	
10	1	
11	2	
12	2	
13	2	
14	2	
15	2	
16	2	
17	2	
18	2	

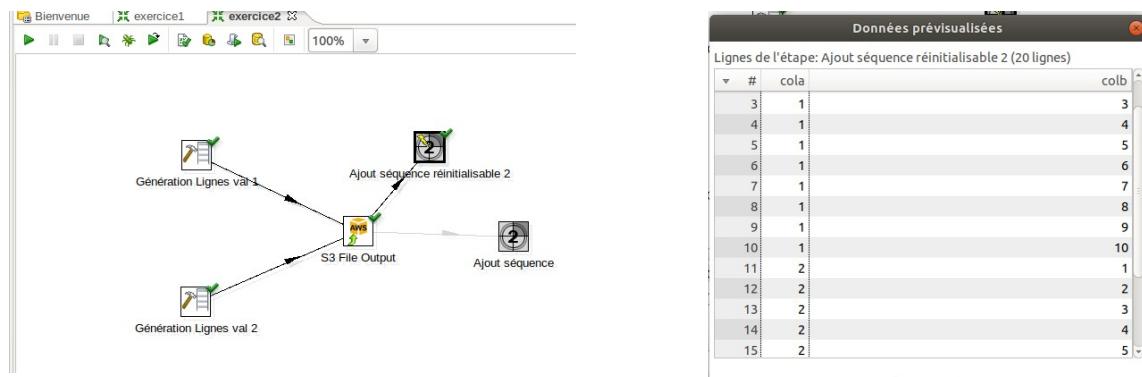
- Numérotez les lignes dans un champs index



## Visualisation du résultat

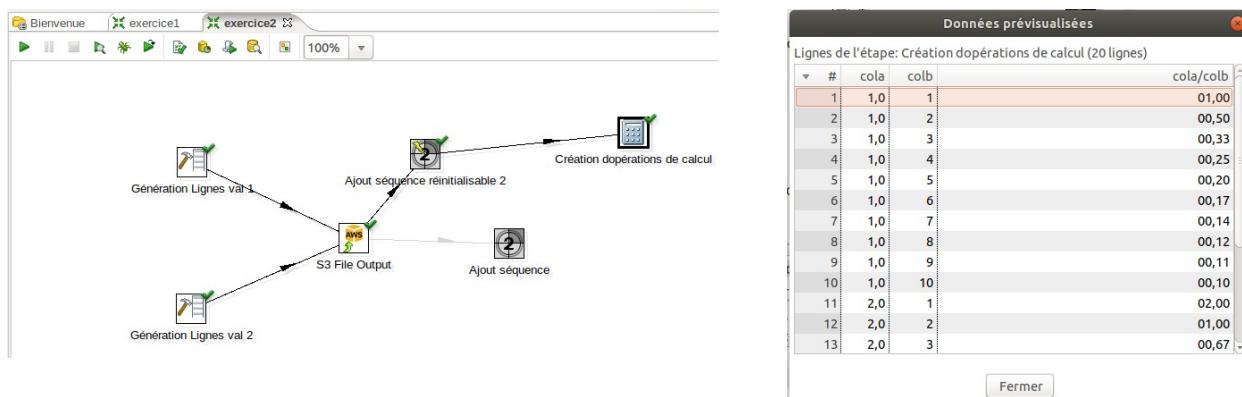
Données prévisualisées		
Lignes de l'étape: Ajout séquence (20 lignes)		
#	cola	index
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
7	1	7
8	1	8
9	1	9
10	1	10
11	2	11
12	2	12
13	2	13

- Numérotez les lignes dont cola vaut 1 et les lignes dont cola vaut 2 indépendamment à l'aide de ajouter séquence ré-initialisation dans un compteur colb.

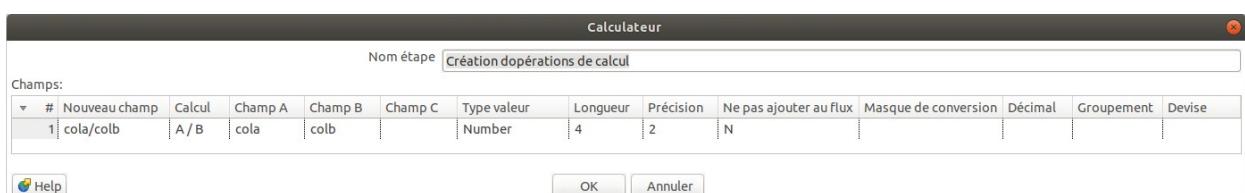


⇒ Dans cette étape a chaque fois que la valeur du champs change on va ré-initialiser le compteur.

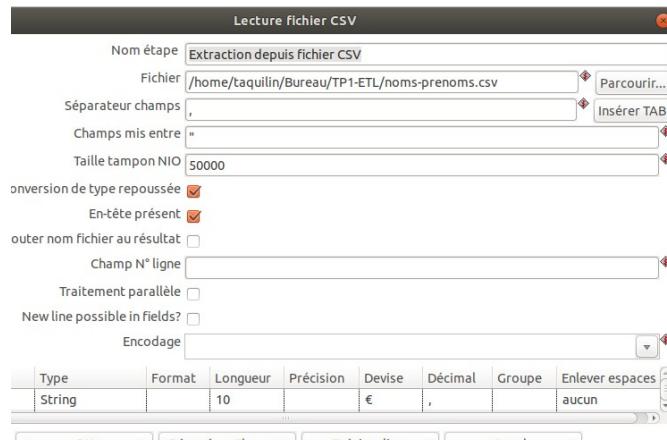
- Rajouter une colonne à l'entrée de valeur cola/colb grâce à un calculateur en utilisant la formule [cola]/[colb]



Ajout de la colonne cola/colb avec la formule



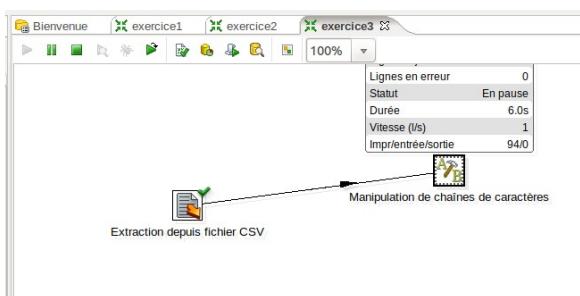
### 3. Exercice 3



- Importer le fichier noms-prenom.csv

**x** Il faut vérifier le séparateur on utilise ici , pour un fichier CSV

- A l'aide de Manipulation de chaînes de caractères, écrire les noms en majuscule, les prénoms en minuscule et les pays avec la première lettre en majuscule.



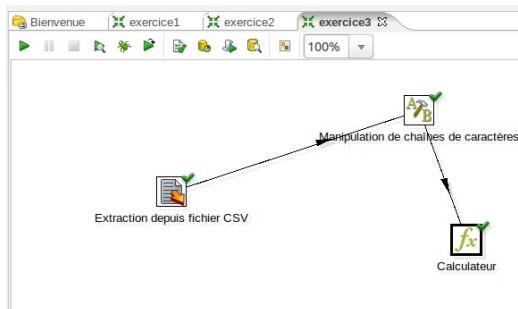
### Configuration de la manipulation chaîne de caractères



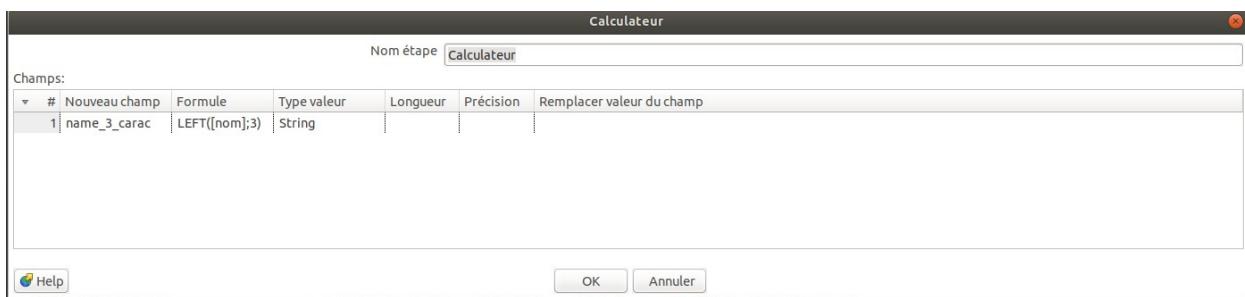
Le résultat est représenté la la figure suivante :

Données prévisualisées								
Lignes de l'étape: Manipulation de chaînes de caractères (100 lignes)								
#	nom	prenom	tel	pays	age	nom_out	prenom_out	pays_out
1	Conley	Pearl	<null>	France	12	CONLEY	pearl	France
2	Head	Paloma	<null>	Luxembourg	34	HEAD	paloma	Luxembourg
3	Simpson	Kitra	<null>	France	56	SIMPSON	kitra	France
4	Monroe	Merritt	<null>	United Arab Emirates	78	MONROE	merritt	United Arab Emirates
5	Buck	Zeus	<null>	France	32	BUCK	zeus	France
6	Velez	Adena	<null>	Turkmenistan	45	VELEZ	adena	Turkmenistan
7	Lancaster	Risa	<null>	Luxembourg	65	LANCASTER	risa	Luxembourg
8	Buchanan	Tamekah	<null>	French Guiana	26	BUCHANAN	tamekah	French Guiana
9	Barr	Cherokee	<null>	Norway	89	BARR	cherokee	Norway
10	Hughes	Warren	<null>	Norway	61	HUGHES	warren	Norway
11	Sutton	Ralph	<null>	Iran	57	SUTTON	ralph	Iran
12	Merritt	Ulric	<null>	France	19	MERRITT	ulric	France
13	Young	Ashely	<null>	France	51	YOUNG	ashely	France

- Créer une colonne qui contient les 3 premières lettres du nom de famille à l'aide de Calculateur ainsi que de Extraction depuis chaînes de caractères.

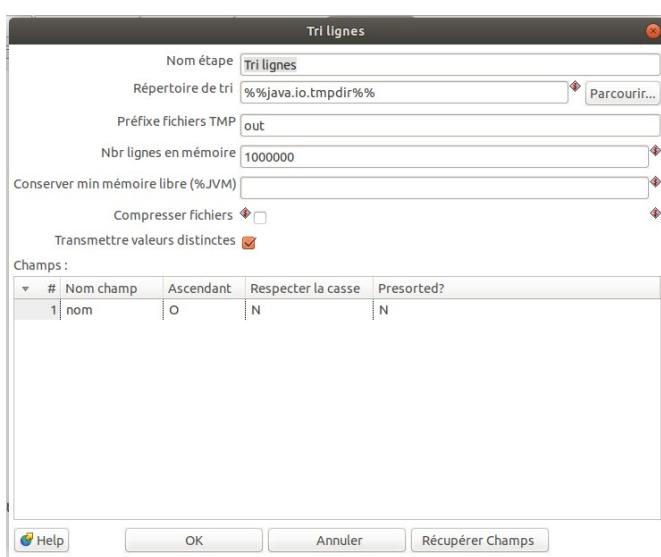


Configuration du calculateur avec l'utilisation de l'opérateur LEFT qui Renvoie un nombre sélectionné de caractères de texte à partir de la gauche.



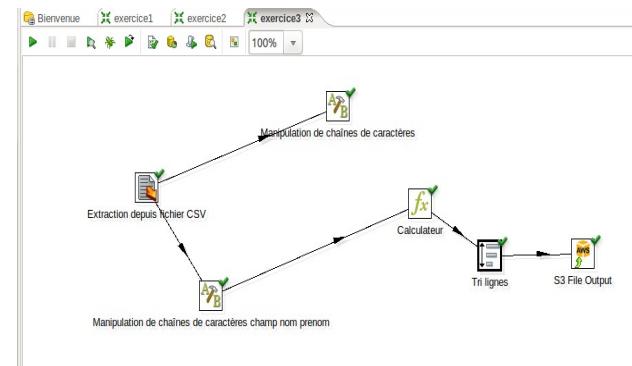
Le résultat est le suivant

Données prévisualisées										
Lignes de l'étape: Calculateur (100 lignes)										
#	nom	prenom	tel	pays	age	nom_out	prenom_out	tel_out	pays_out	name_3_carac
1	Conley	Pearl	(894) 671-5006	France	12	CONLEY	pearl	(894) 671-5006	France	Con
2	Head	Paloma	(316) 784-3364	Luxembourg	34	HEAD	paloma	(316) 784-3364	Luxembourg	Hea
3	Simpson	Kitra	(428) 721-0717	France	56	SIMPSON	kitra	(428) 721-0717	France	Sim
4	Monroe	Merritt	(451) 987-8014	United Arab Emirates	78	MONROE	merritt	(451) 987-8014	United Arab Emirates	Mon
5	Buck	Zeus	(875) 958-4198	France	32	BUCK	zeus	(875) 958-4198	France	Buc
6	Velez	Adena	(481) 280-8350	Turkmenistan	45	VELEZ	adena	(481) 280-8350	Turkmenistan	Vel
7	Lancaster	Risa	(930) 829-8777	Luxembourg	65	LANCASTER	risa	(930) 829-8777	Luxembourg	Lan
8	Buchanan	Tamekah	(283) 956-5634	French Guiana	26	BUCHANAN	tamekah	(283) 956-5634	French Guiana	Buc
9	Barr	Cherokee	(914) 183-7016	Norway	89	BARR	cherokee	(914) 183-7016	Norway	Bar
10	Hughes	Warren	(857) 459-8826	Norway	61	HUGHES	warren	(857) 459-8826	Norway	Hug
11	Sutton	Ralph	(803) 832-5808	Iran	57	SUTTON	ralph	(803) 832-5808	Iran	Sut
12	Merritt	Ulric	(230) 560-5193	France	19	MERRITT	ulric	(230) 560-5193	France	Mer



- Trier par ordre alphabétique sur le nom

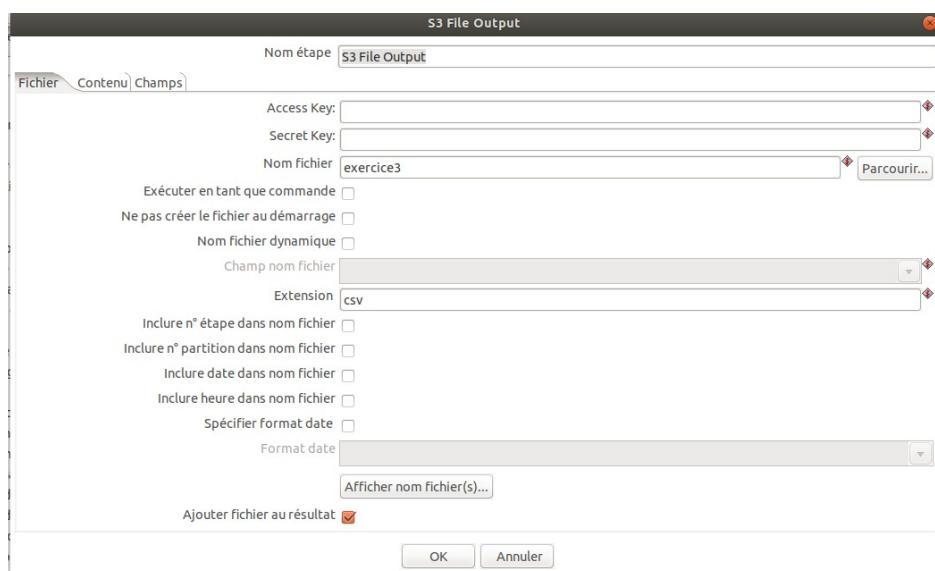
pour trier on va utiliser tri ligne avec cette configuration



pour aboutir a ce résultat

Données prévisualisées								
#	nom	prenom	tel	pays	age	nom_out	prenom_out	name_3_carac
1	Barr	Silas	(607) 106-6814	Macedonia	76	BARR	silas	Bar
2	Beasley	Flavia	(287) 705-8658	Cambodia	19	BEASLEY	flavia	Bea
3	Bender	Amelia	(223) 333-0507	France	62	BENDER	amelia	Ben
4	Boyd	Bethany	(550) 342-9908	Dominican Republic	77	BOYD	bethany	Boy
5	Bradford	Angelica	(756) 366-6843	Sierra Leone	15	BRADFORD	angelica	Bra
6	Buchanan	Tamekah	(283) 956-5634	French Guiana	26	BUCHANAN	tamekah	Buc
7	Burton	Baker	(298) 316-1315	American Samoa	15	BURTON	baker	Bur
8	Caldwell	Rinah	(468) 405-5802	Swaziland	4	CALDWELL	rinah	Cal
9	Cantu	Juliet	(545) 696-7266	Armenia	56	CANTU	juliet	Can
10	Carter	Chester	(799) 687-6381	Norway	56	CARTER	chester	Car
11	Cleveland	Xena	(511) 319-4855	Niue	72	CLEVELAND	xena	Cle
12	Coffey	Vanna	(979) 545-6453	Slovakia	54	COFFEY	vanna	Cof
13	Cohen	Belle	(984) 921-0799	Hungary	55	COHEN	belle	Coh

- Sauvegarder la sorte dans un fichier CSV



- Sauvegarder la sorte dans un fichier Excel et ouvrez le fichier

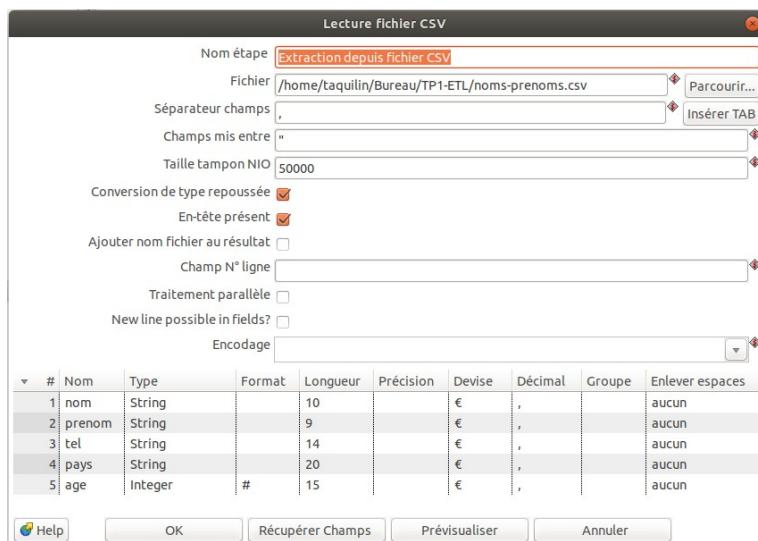
A	B	C	D	E	F
1 nom	prenom	tel	pays	age	=nom_3_carac
2 BARR	silas	(607) 106-6814	Macedonia	76	Bar
3 BEASLEY	flavia	(287) 705-8658	Cambodia	19	Bea
4 BENDER	amelia	(223) 333-0507	France	62	Ben
5 BOYD	bethany	(550) 342-9908	Dominican Republic	77	Boy
6 BRADFORD	angelica	(756) 366-6843	Sierra Leone	15	Bra
7 BUCHANAN	tamekah	(283) 956-5634	French Guiana	26	Buc
8 BURTON	baker	(298) 316-1315	American Samoa	10	Bur
9 CALDWELL	rimah	(468) 405-5802	Swaziland	4	Cal
10 CANTU	juliet	(545) 696-7266	Armenia	56	Can
11 CARTER	chester	(799) 687-6381	Norway	56	Car
12 CLEVELAND	xena	(511) 319-4855	Niue	72	Cle
13 COFFEY	vanna	(979) 545-6453	Slovakia	54	Cof
14 COHEN	belle	(984) 921-0799	Hungary	55	Coh
15 COLON	abraham	(927) 588-7648	Latvia	51	Col
16 DELGADO	catherine	(106) 616-7990	Cambodia	27	Del
17 ENGLAND	veronica	(259) 946-0267	Latvia	89	Eng
18 FROST	clayton	(152) 175-5307	Guatemala	94	Fro
19 FRY	blaine	(269) 935-3999	Norway	30	Fry
20 HEAD	paloma	(316) 784-3364	Luxembourg	34	Hea
21 HILL	elton	(971) 544-4710	France	51	Hil
22 HUFF	camilla	(543) 384-0775	Bahamas	79	Huf
23 HUGHES	warren	(857) 459-8826	Norway	61	Hug
24 JUSTICE	halla	(192) 340-3538	Luxembourg	86	Jus
25 LOVE	adele	(530) 946-6988	Chad	86	Lov
26 LOWE	memill	(153) 570-0148	Moldova	72	Low
27 MANNING	blair	(113) 979-7166	Chad	43	Man
28 MARKS	felicia	(108) 149-4218	Luxembourg	60	Mar
29 MCCLAIN	leoma	(903) 115-2648	Guernsey	63	Mcc
30 MCKAY	olive	(775) 479-7066	Moldova	95	Mck
31 MERRITT	ulric	(230) 560-5193	France	19	Mer
32 MEYER	clarke	(236) 761-7196	Denmark	61	Mey
33 MONROE	gwenith	(454) 907-9014	United Arab Emirates	70	Mon

#### 4. Exercice 4

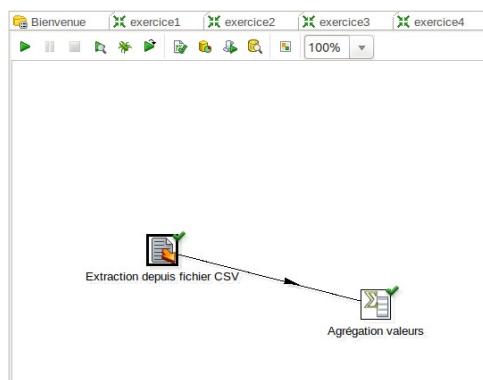
Téléchargez le fichier mobiliers de Stationnement

#### 5. Exercice 7

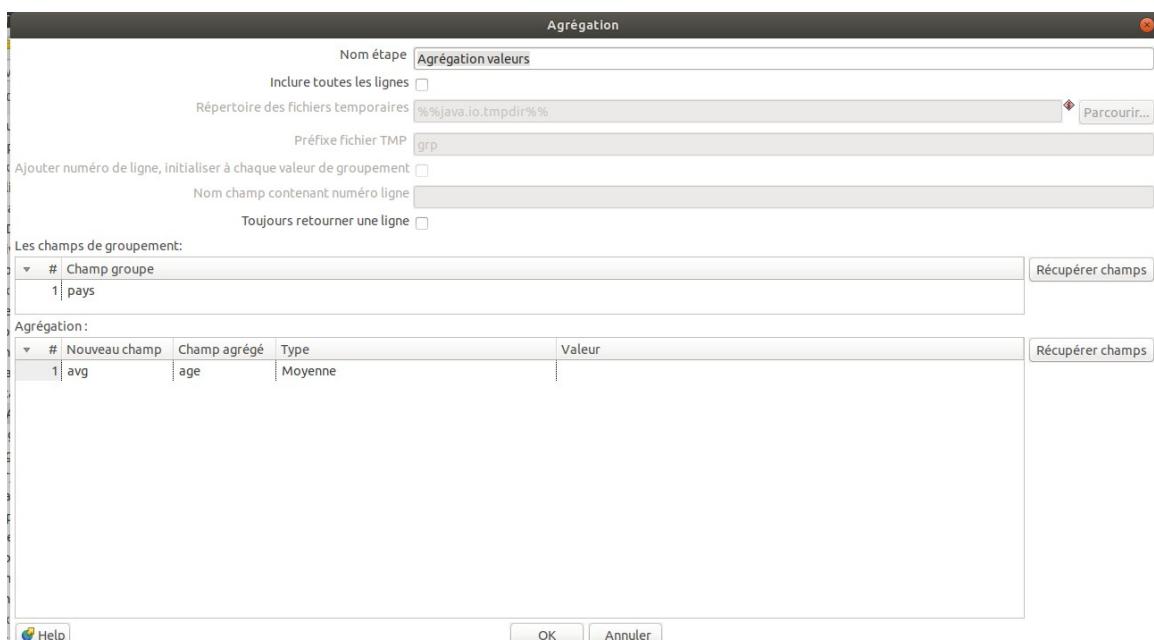
Importer le fichier noms-prenom.csv



## Calculer la moyenne d'âge par pays



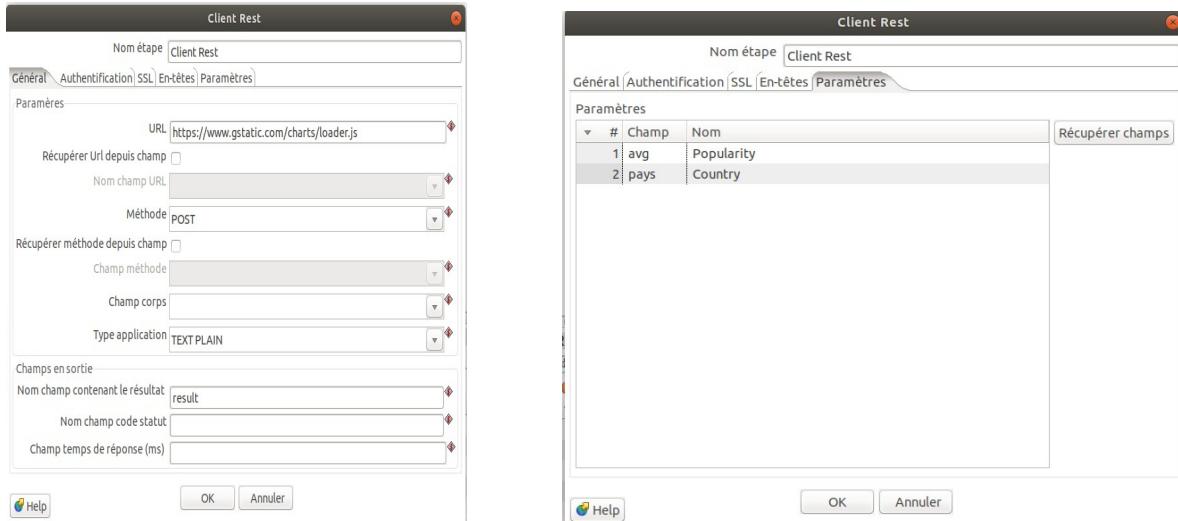
On utilise ici l'agrégation des valeurs avec la configuration suivante :



On aura le résultat suivant :

Données prévisualisées		
Lignes de l'étape: Agrégation valeurs (88 lignes)		
#	pays	avg
1	France	12
2	Luxembourg	34
3	France	56
4	United Arab Emirates	78
5	France	32
6	Turkmenistan	45
7	Luxembourg	65
8	French Guiana	26
9	Norway	75
10	Iran	57
11	France	35
12	Cameroon	75
13	New Caledonia	53

Exportez le fichier afin de pouvoir l'utiliser dans Google configuration



## 6. Exercice 6

