

# Tale of Data

Big Data Preparation for Business People

Guide de démarrage



---

# Table des matières

Introduction .....	4
Tale of Data (TOD) .....	4
Principes fondamentaux .....	4
Définitions .....	5
Description des icônes Tale of Data .....	5
Présentation de l'écran d'accueil .....	6
Ecran d'accueil .....	6
Créer son premier Flow .....	7
Créer un Flow .....	7
Accéder au Flow Designer .....	7
Présentation du Flow Designer .....	8
Barre d'outils .....	8
Canevas .....	8
Zone de configuration .....	8
Zone de prévisualisation .....	9
Ajouter une source de données .....	9
Configurer le nœud source .....	9
Ajouter une cible pour les données .....	12
Configurer le nœud cible .....	12
Relier les nœuds .....	14
Ajouter un nœud filtre .....	15
Configurer le nœud filtre .....	16
Exécuter le Flow .....	18
Télécharger les résultats .....	20
Comparer les fichiers avant et après le traitement .....	20
Pour aller plus loin .....	22
Gérer ses Flows .....	22
Nœuds de traitement de la barre d'outils .....	23
Lien vers le guide de référence .....	23

# Introduction

## Tale of Data (TOD)

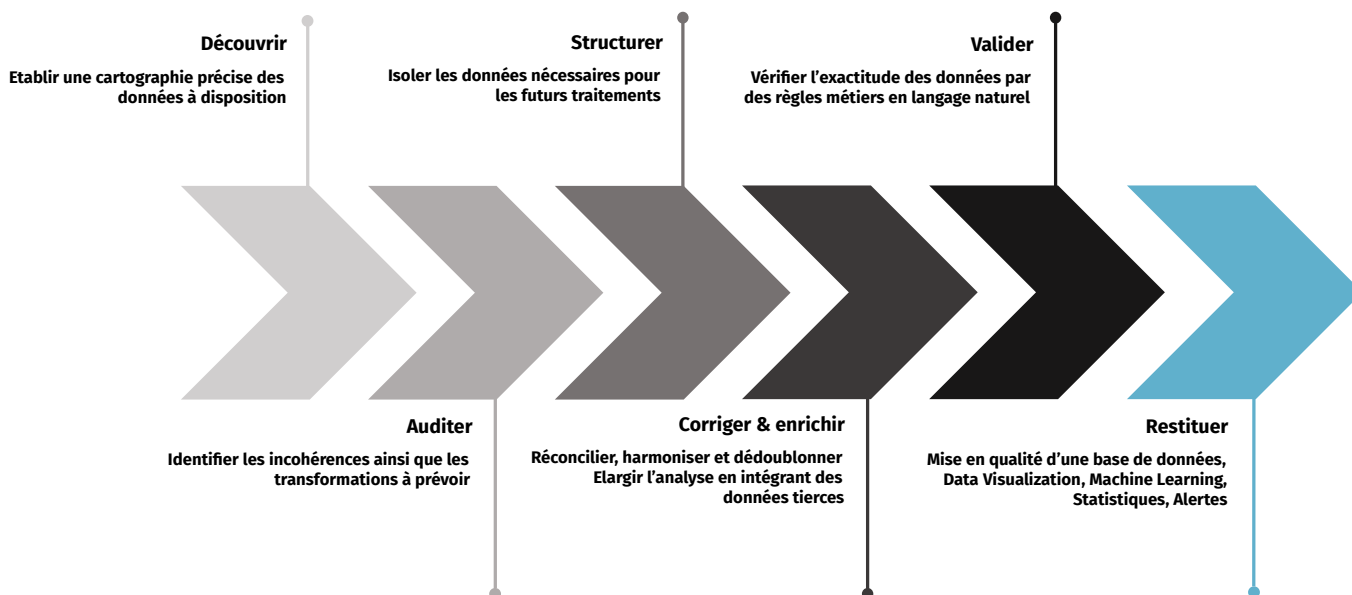
Tale of Data est un logiciel conçu pour les utilisateurs métier, afin de leur permettre de reprendre confiance dans leurs données et donc de prendre de bonnes décisions.

Tale of Data propose les fonctionnalités suivantes, sans avoir besoin d'écrire du code :

- Fiabilisation intelligente des données,
- Contrôle de la conformité des données,
- Découverte des jeux de données disséminés au sein du Système d'Information,
- Cartographie sémantique (nature des données) et cartographie des anomalies,
- Croisement de sources de données hétérogènes, augmentation des données,
- Automatisation des contrôles, des alertes et des opérations de remédiation.

## Principes fondamentaux

Des données auditées, fiabilisées par les métiers pour des décisions avisées afin de maximiser vos résultats et minimiser vos risques !



---



## Définitions



**Flow** Un flow est un traitement construit sous forme graphique. Il est constitué :

- De données en entrée,
- D'un ensemble de transformation,
- De données de sortie.

---



**Nœud** Un nœud est représenté par des icônes dans un flow.  
On retrouve un nœud pour représenter :

-  Les données de départ, appelées 'nœud source',
-  Les données d'arrivée, appelées 'nœud cible',
- Les traitements utilisés pour transformer, corriger ou compléter les données.

On utilisera par exemple un nœud filtre  ou un nœud enrichissement  dans les traitements de notre flow.

## Description des icônes Tale of Data

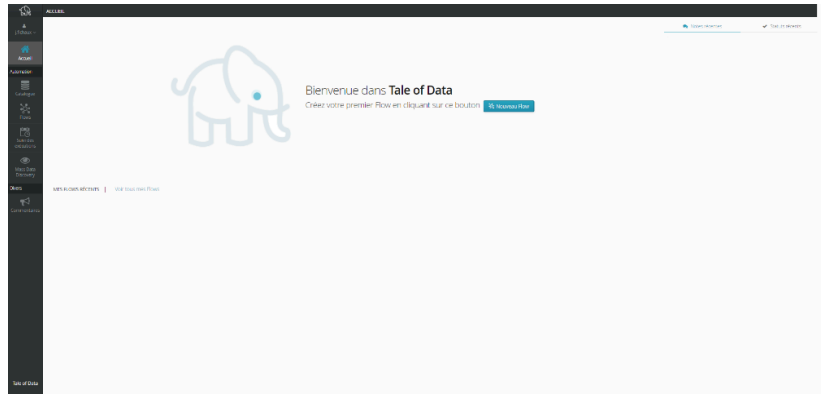
Tale of Data affiche des icônes pour vous guider :

-  Un astérisque indique qu'il est obligatoire de remplir le champ.
-  Une bulle avec un point d'exclamation indique qu'un paramètre est manquant. De plus, cette icône vous fournit une aide pour y remédier lorsque vous déplacez votre souris dessus.

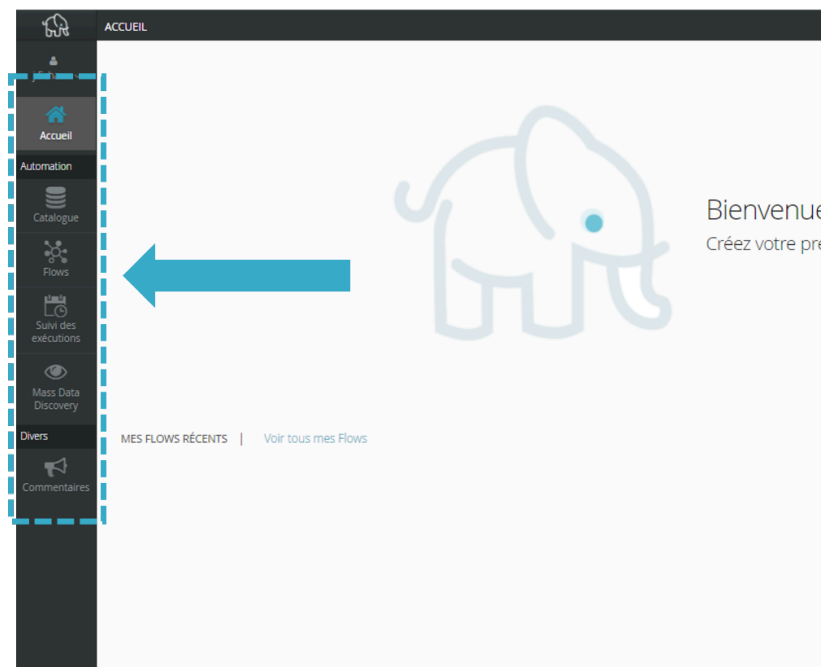
# Présentation de l'écran d'accueil

## Ecran d'accueil

Après avoir saisi votre identifiant et votre mot de passe, l'écran d'accueil apparaît.

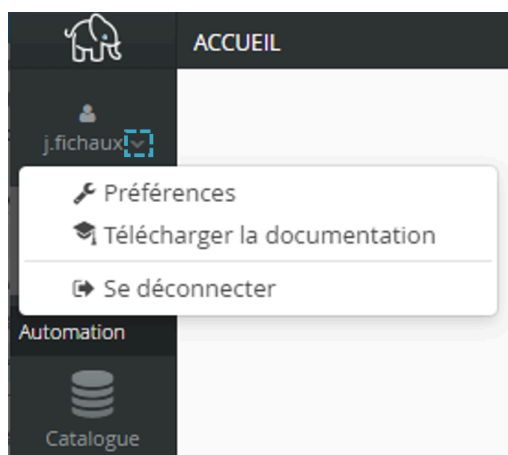


Vous retrouvez à gauche de l'écran le menu principal de Tale of Data qui reste toujours visible.



Le menu déroulant permet de :




- Paramétrer les préférences utilisateur,
- Télécharger la documentation,
- Se déconnecter.



# Créer son premier Flow

Dans le cadre de ce guide de démarrage, nous allons créer un Flow "simple". Ce Flow aura pour objectif d'appliquer un filtre sur un jeu de données afin de récupérer les items concernant la France.

Pour y parvenir, vous devrez créer et paramétrer :

1.  Un nœud source,
2.  Un nœud cible,
3.  Un nœud filtre.



## Important

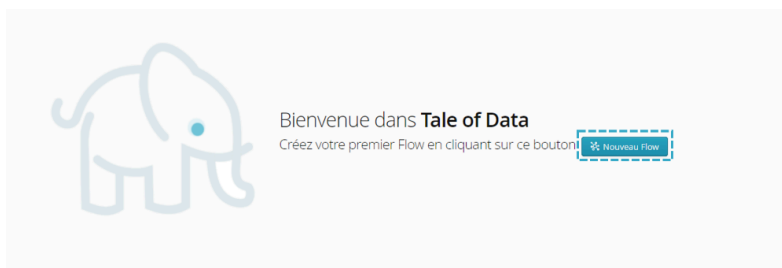
Ce Flow permet à un nouvel utilisateur de se familiariser avec les concepts de Tale of Data. Il s'agit d'un Flow d'exemple qui ne reflète aucunement le plein potentiel de Tale of Data.

## Créer un Flow

### Accéder au Flow Designer

Dans l'écran d'accueil :

1. Appuyez sur le bouton "Nouveau Flow".



2. Nommez le Flow."  
Par exemple : "Mon premier Flow"


3. Appuyez sur "Valider".







## Important

Le nœud préparation  est le seul nœud qui se configure dans une autre interface.

## Zone de prévisualisation

4


Zone de prévisualisation pour visionner les données d'entrée et de sortie du nœud sélectionné dans le Flow.

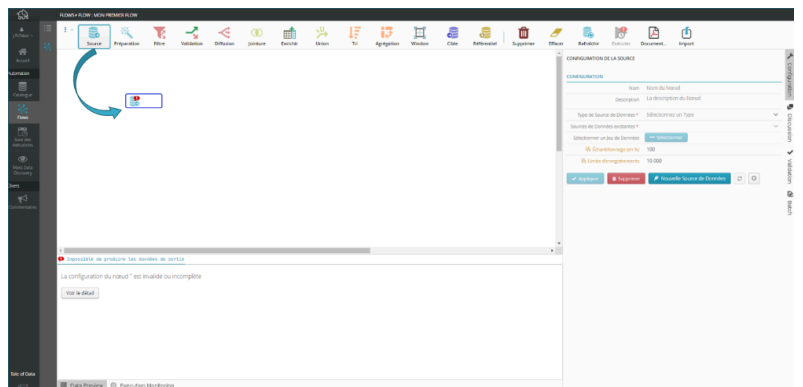



## Attention

Cette prévisualisation est partielle. En effet, elle n'est appliquée qu'à un échantillon de lignes.

## Ajouter une source de données


1. Faites un glisser-déposer de l'icône du nœud source  de la barre d'outils sur le canevas.



Un nœud source  permet de récupérer des enregistrements (en se connectant à des fichiers ou à des bases de données).

## Configurer le nœud source

Sur la partie droite de l'écran, configuration.

1. Vous pouvez éventuellement nommer votre nœud source .

Par exemple : "Ma Source de données"

Si vous ne le faites pas, le nœud prendra automatiquement le nom du fichier ou de la table sélectionné.

2. Sélectionnez dans le champ "Type de Source de Données" : Système de fichiers Tale of Data.

CONFIGURATION	
Nom	Ma source de données
Description	La description du Nœud
Type de Source de Données *	▼
Sources de Données existantes *	Système de fichiers Tale of Data
Sélectionner un Jeu de Données	
Échantillonnage (en %)	
Limite d'enregistrements	
Appliquer	Supprimer



## Note

Le champ "Sources de Données existantes" est complété automatiquement.

3. Appuyez sur le bouton "Sélectionner".

CONFIGURATION	
Nom	Ma source de données
Description	La description du Nœud
Type de Source de Données *	Système de fichiers Tale of Data ▼
Sources de Données existantes *	Mon espace de travail ▼
Sélectionner un Jeu de Données	*** Sélectionner

Une nouvelle fenêtre apparaît.

Sélectionner un Jeu de Données

Mon espace de travail

SÉLECTION D'UN FICHIER

NOUVEAU DOSSIER

Nom du nouveau dossier    Nom du nouveau dossier à créer    Créer le dossier

TÉLÉVERSER DES FICHIERS

Téléverser des fichiers dans le dossier courant    Sélectionner les fichiers

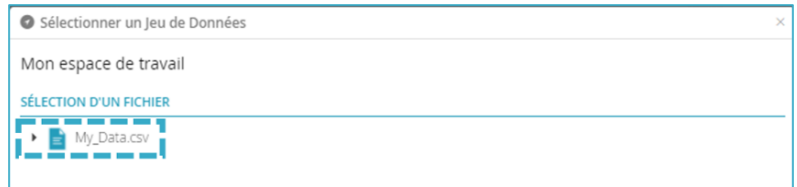
Raccourcis    Précédent    Sélectionner    Annuler    Précédent    Sélectionner    Annuler

4. Dans la partie "Téléverser des fichiers", appuyez sur le bouton "Sélectionner les fichiers".



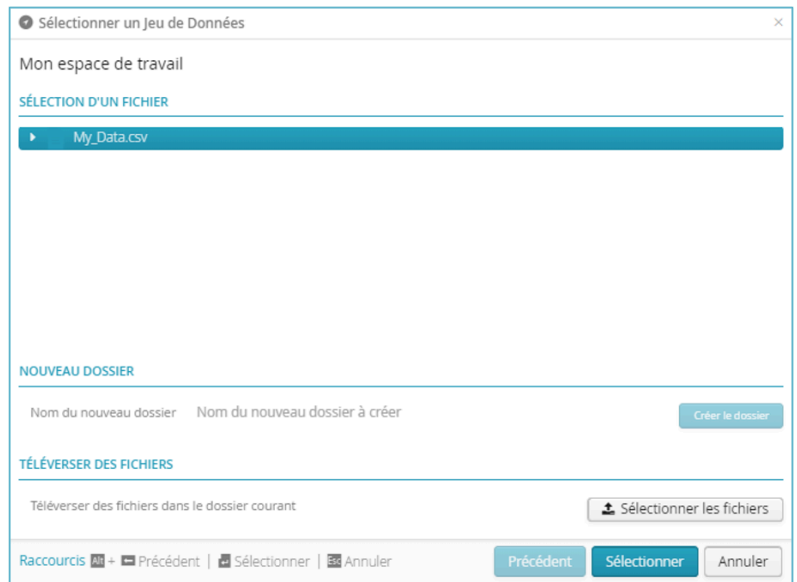
5. Importez le fichier de démonstration fourni avec le guide de démarrage (My\_Data.csv).

Dans la partie "Sélection d'un fichier", le fichier est visible.



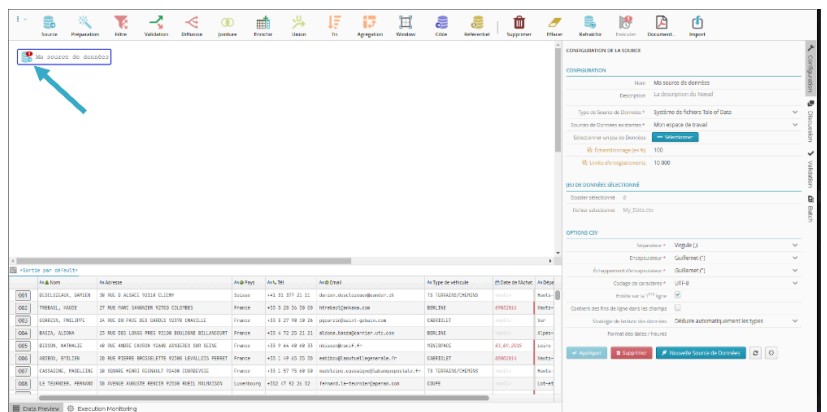
6. Cliquez sur le fichier "My\_Data.csv" pour le sélectionner.

Le bouton "Sélectionner" devient coloré et donc actif.





7. Appuyez sur le bouton "Sélectionner".

L'interface Flow Designer réapparaît avec votre nœud source configuré.





## Note



L'icône  indique que la source n'est pas encore prête à être utilisée. En effet, elle doit être liée à un nœud cible .

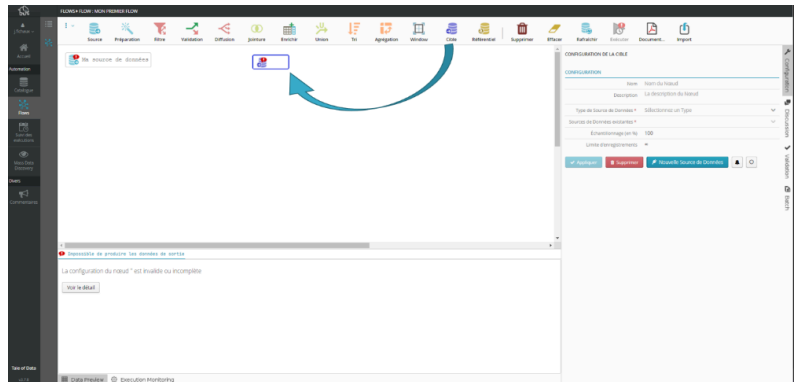


## Note

Vous pouvez observer le contenu du fichier importé dans la zone de prévisualisation en bas de l'écran.

# Ajouter une cible pour les données


1. Faites un glisser-déposer de l'icône du nœud cible  de la barre d'outils dans le canevas à côté du nœud source .



Un nœud cible  permet d'envoyer des enregistrements vers un système de stockage.

## Configurer le nœud cible

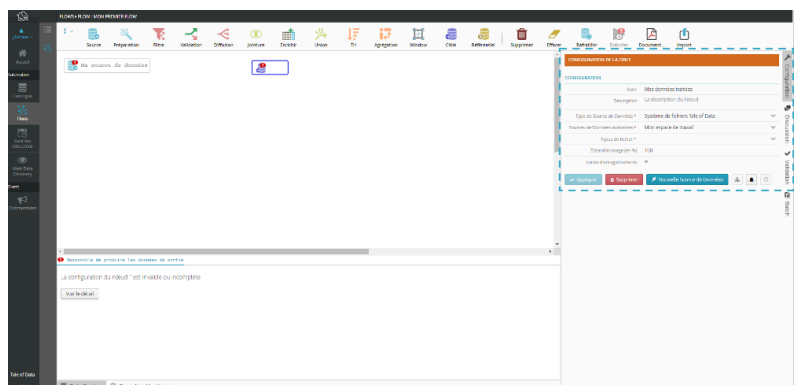
Sur la partie Configuration à droite de l'écran :

1. Vous pouvez éventuellement nommer votre nœud cible .

Par exemple : "Données traitées"

Si vous ne le faites pas, le nœud prendra automatiquement le nom du fichier ou de la table sélectionné.

2. Sélectionnez dans le champ "Type de Source de Données" : Système de fichiers Tale of Data.



## Note

Le champ "Sources de Données existantes" est complété automatiquement.

3. Dans la liste déroulante du champ "Types de fichier", sélectionnez "XLSX (à partir d'Excel 2007)".

**CONFIGURATION DE LA CIBLE**

**CONFIGURATION**

Nom	Mes données traitées
Description	La description du Nœud
Type de Source de Données *	Système de fichiers Tale of Data
Sources de Données existantes *	Mon espace de travail
Types de fichier *	<b>XLSX (à partir d'Excel 2007)</b>
Échantillonnage (en %)	100
Limite d'enregistrements	1 048 576

Une nouvelle partie de la zone de configuration apparaît.

**CONFIGURATION DE LA CIBLE**

**CONFIGURATION**

Nom	Mes données traitées
Description	La description du Nœud
Type de Source de Données *	Système de fichiers Tale of Data
Sources de Données existantes *	Mon espace de travail
Types de fichier *	XLSX (à partir d'Excel 2007)
Échantillonnage (en %)	100
Limite d'enregistrements	1 048 576

**OPTIONS EXCEL (XLSX)**

Mode de Sauvegarde *	Écraser
Politique d'horodatage *	Aucun horodatage
Sélectionner le fichier de sortie	<b>Sélectionner</b>
Format des dates *	yyyy-MM-dd HH:mm:ss
Entête sur la 1 <sup>ère</sup> ligne	<input checked="" type="checkbox"/>

4. Dans la partie "Options Excel (XLSX)", appuyez sur le bouton "Sélectionner".

Une nouvelle fenêtre apparaît.

5. Dans la partie "Fichier de Sortie" remplissez les 2 champs :

- "Nom du Fichier", nom qui sera attribué au fichier Excel,

Par exemple : "Données Traitées"

- "Nom de la feuille", nom de la feuille de calcul où seront stockées les données.

Par exemple : "Feuille\_1"

My\_Data.csv

**FICHER DE SORTIE**

Nom du fichier *	Données Traitées.xlsx
Nom de la feuille *	Feuille_1

**NOUVEAU DOSSIER**

Nom du nouveau dossier	Nom du nouveau dossier à créer	<input type="button" value="Créer le dossier"/>
------------------------	--------------------------------	---

Raccourcis   |  |

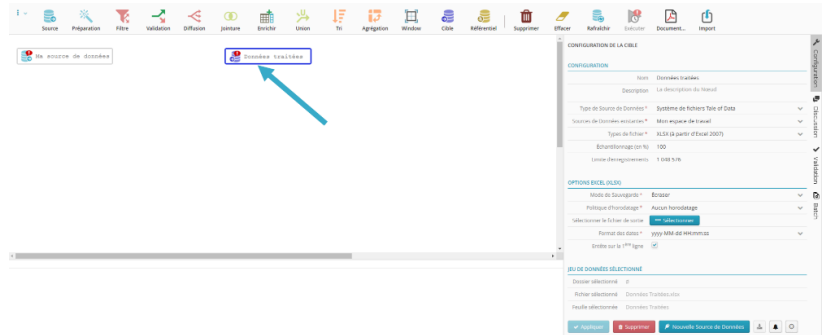


## Important

Il est nécessaire de sortir des champs de saisie pour rendre le bouton "Sélectionner" actif.

Après avoir rempli et être sorti du champ "Nom de la feuille", le bouton se colore et devient actif.

6. Appuyez sur le bouton "Sélectionner".



L'interface Flow Designer réapparaît avec votre nœud cible configuré.



## Note

L'icône indique que la cible n'est pas encore prête à être utilisée. En effet, elle doit être liée à un nœud source

## Relier les nœuds

Pour créer un lien entre deux items du Flow :

1. Cliquez sur le nœud source
2. Faites glisser la souris vers le nœud cible en maintenant le bouton gauche de la souris enfoncé.



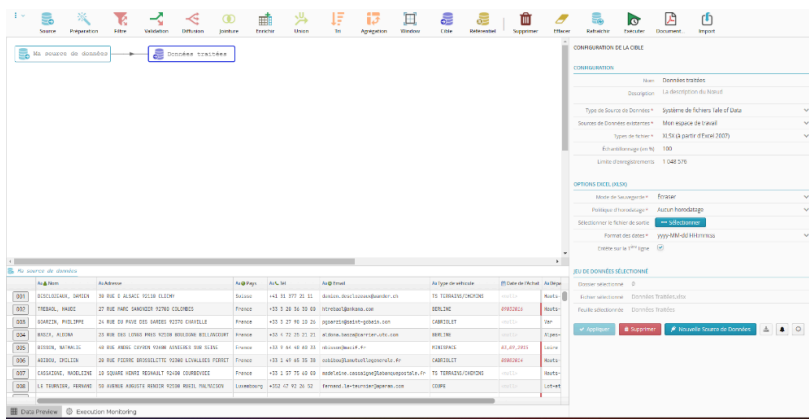
## Note

Une icône de flèche doit apparaître

3. Relâchez le bouton gauche de la souris une fois au-dessus du nœud cible .






Vous avez créé le lien entre les deux nœuds.

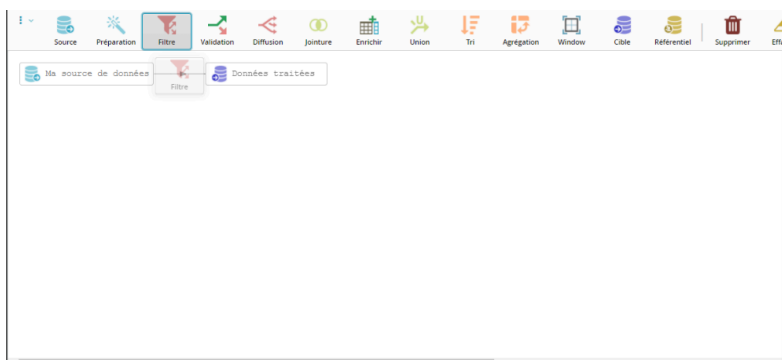



## Note

La bulle avec le point d'exclamation a disparu : le lien est valide et les nœuds sont prêts à être utilisés.

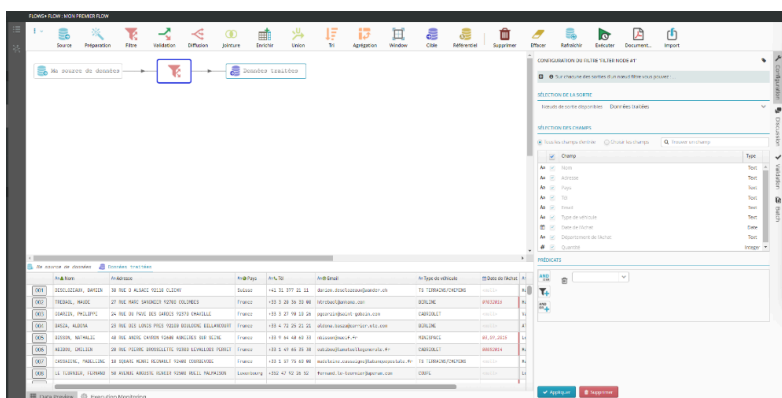
## Ajouter un nœud filtre

- Faites un glisser-déposer de l'icône du nœud filtre  sur le lien qui relie le nœud source  et le nœud cible .






Un nœud filtre  permet de réaliser un filtrage de vos données.


Le nœud filtre  est intégré au Flow.

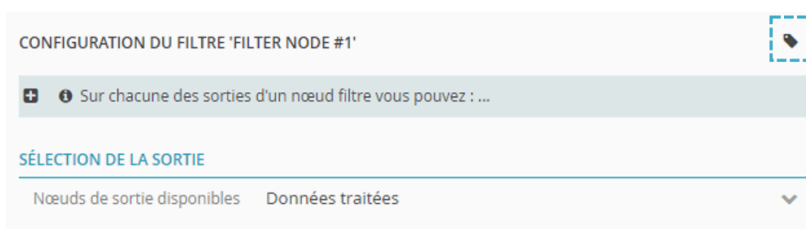


## Note

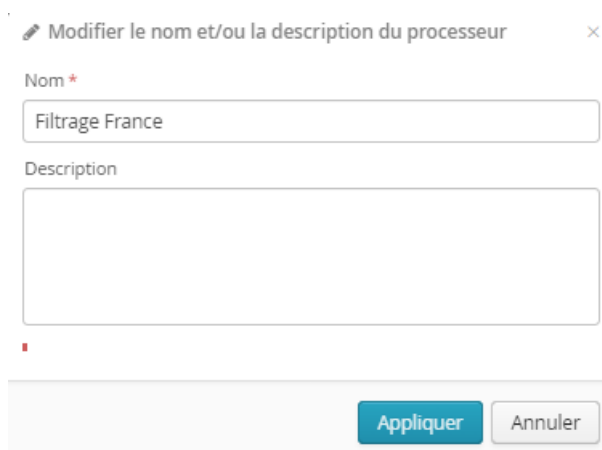
Il est tout à fait possible de construire le Flow de gauche à droite : nœud cible , puis nœud filtre  et enfin nœud source .

## Configurer le nœud filtre

Vous pouvez nommer votre filtre et son action dans la zone de configuration en appuyant sur l'icône  à droite de l'écran.



1. Nommez votre filtre dans le champ "Nom".  
Par exemple : "Filtrage France".  
Appuyez sur le bouton "Appliquer".



### Note

Pour rappel, ce Flow a pour objectif de réaliser un filtre afin de récupérer les données concernant la France.



2. Dans la partie Prédicats, accédez à la liste déroulante en cliquant sur la flèche.

**SÉLECTION DES CHAMPS**

Tous les champs d'entrée  Choisir les champs

<input checked="" type="checkbox"/>	Champ	Type
<input checked="" type="checkbox"/>	Aa Nom	Text
<input checked="" type="checkbox"/>	Aa Adresse	Text
<input checked="" type="checkbox"/>	Aa Pays	Text
<input checked="" type="checkbox"/>	Aa Tél	Text
<input checked="" type="checkbox"/>	Aa Email	Text
<input checked="" type="checkbox"/>	Aa Type de véhicule	Text
<input checked="" type="checkbox"/>	Date de l'Achat	Date
<input checked="" type="checkbox"/>	Aa Département de l'Achat	Text
<input checked="" type="checkbox"/>	# Quantité	Integer

**PRÉDICATS**

AND

AND OR

- Aa Nom
- Aa Adresse
- Aa Pays**
- Aa Tél
- Aa Email
- Aa Type de véhicule
- Date de l'Achat
- Aa Département de l'Achat
- # Quantité



## Note

Le contenu de la liste déroulante est généré en fonction du fichier chargé.

3. Sélectionnez "Pays".  
Deux nouveaux champs apparaissent.

**PRÉDICATS**

AND

AND OR

Aa Pays equal to

4. Dans le champ vide, saisissez la valeur "France".  
La valeur saisie est sensible à la casse (minuscule, majuscule).

**PRÉDICATS**

AND

AND OR

Aa Pays equal to



## Note

La condition "equal to" signifie "égale à".

Elle permet de chercher dans la colonne "Pays" une valeur "égale à" "France" dans le but de récupérer les données concernant la France.

5. Appuyez sur le bouton "Appliquer".


Un aperçu des données filtrées s'affiche dans la zone de prévisualisation en bas de l'écran.

The screenshot shows a data tool interface with a filter configuration panel on the right and a data preview table at the bottom. The filter configuration is set to 'Sélection de la source' and 'Sélection des champs'. The data preview table shows a list of records with columns for ID, Nom, Adresse, Age, Sexe, and Email.

ID	Nom	Adresse	Age	Sexe	Email
001	TREMBLÉ, MARIE	27 RUE MARC SANDREY NOTRE DAME	France	+33 9 28 36 38 88	mtrebl@orange.com
002	LAUREN, PHILIPPE	24 RUE DU PAPE DES JARDIS NOTRE DAME	France	+33 9 27 96 28 26	lphlaur@orange.com
003	ROUSSEAU, ALAIN	28 RUE DES ERMS RUE DES ERMS RUE DES ERMS	France	+33 9 27 28 28 28	alain.rousseau@orange.com
004	BOISSON, MARCELLE	48 RUE MARC SANDREY NOTRE DAME	France	+33 9 24 48 48 33	mboisson@orange.com
005	MAISON, JULIEN	38 RUE PIERRE BROUSSELET NOTRE DAME	France	+33 1 47 47 33 33	jmaison@orange.com
006	COSSIGNA, NOELLE	38 SQUARE HENRI RENAUULT NOTRE DAME	France	+33 1 27 27 27 27	noelle.cossigna@orange.com
007	ROUSSEAU, ALAIN	38 RUE MARC SANDREY NOTRE DAME	France	+33 9 28 28 28 28	alain.rousseau@orange.com
008	TRANCHES, ALAIN	18 RUE PIERRE BROUSSELET NOTRE DAME	France	+33 9 24 48 48 33	alain.tranches@orange.com



## Attention

Après avoir cliqué sur le nœud cible , ce dernier semble contenir les résultats escomptés mais ce n'est qu'un échantillon.

Il faudra exécuter le Flow pour que le filtrage soit bien appliqué sur tout le jeu de données.

## Exécuter le Flow

1. Appuyez sur le bouton "Exécuter" situé à droite dans la barre d'outils.

The screenshot shows a toolbar with five buttons: 'Effacer', 'Rafraichir', 'Exécuter', 'Document...', and 'Import'. The 'Exécuter' button is highlighted with a dashed blue box.

2. La fenêtre de planification apparaît.


Appuyez sur le bouton "Exécuter maintenant".


Vous pouvez observer un message de réussite de la planification 1 et de l'exécution 2 du Flow.



Félicitations, vous venez de créer et d'exécuter votre premier Flow !

## Télécharger les résultats

1. Cliquez **sur** le nœud cible  pour le sélectionner.

Le bouton "Télécharger"  est devenu actif.



JEU DE DONNÉES SÉLECTIONNÉ

Dossier sélectionné	Ø
Fichier sélectionné	Données Traitées.xlsx
Feuille sélectionnée	Données Traitées

Appliquer Supprimer Nouvelle Source de Données Télécharger

2. Cliquez sur le bouton  pour télécharger le fichier.

## Comparer les fichiers avant et après le traitement

Vous pouvez effectuer un comparatif du fichier source et du fichier cible.

Fichier source "My\_Data"

80 RUE ROUGET DE LISLE 92000 NANTERRE	Belgique
22 RUE JULES HETZEL 92310 SEVRES	Suisse
42 RUE DE PARIS 92100 BOULOGNE BILLANCOURT	Luxembourg
9 RUE DE LA FERME 92350 LE PLESSIS ROBINSON	Suisse
11 RUE PIERRE BROSSOLETTE 92400 COURBEVOIE	Allemagne
44 RUE DE L ALMA 92400 COURBEVOIE	France
7 RUE MAURICE THOREZ 92000 NANTERRE	France
310 AVENUE ROGER SALENGRO 92370 CHAVILLE	France
29 RUE EDMOND ABOUT 92350 LE PLESSIS ROBINSON	France
7 BOULEVARD ANATOLE FRANCE 92190 MEUDON	Belgique
43 AVENUE DE LA DIVISION LECLERC 92160 ANTONY	Suisse
1 RUE LOUIS POUET 92800 PUTEAUX	France

Fichier cible "Mes données traitées"

54 RUE DES MOULINEAUX 92150 SURESNES	France
4 RUE ANDRE RIVOIRE 92240 MALAKOFF	France
67 AVENUE CHANDON 92230 GENNEVILLIERS	France
38 RUE DE VARSOVIE 92700 COLOMBES	France
46 RUE FESSART 92100 BOULOGNE BILLANCOURT	France
50 BD VICTOR HUGO 92110 CLICHY	France
16 RUE D'ALSACE 92110 CLICHY	France
16 AVENUE CAMBERWELL 92000 SCEAUX	France
5 RUE CHARCOT 92200 NEUILLY SUR SEINE	France
105 RUE JEAN MARIN NAUDIN 92220 BAGNEUX	France
7 RUE CLAUDE DALSEME 92190 MEUDON	France
3 RUE DE LA MARE ADAM 92370 CHAVILLE	France
150 RTE DE LA REINE 92100 BOULOGNE BILLANCOURT	France
62 RUE ARAGO 92320 CHATILLON	France
4 RUE TELLES DE LA POTERIE 92130 ISSY LES MOULINEAUX	France
3 RUE VOLTAIRE 92160 ANTONY	France



### Note

Ce premier Flow est un exemple d'introduction : des traitements beaucoup plus élaborés (mais simples d'utilisation) sont disponibles.

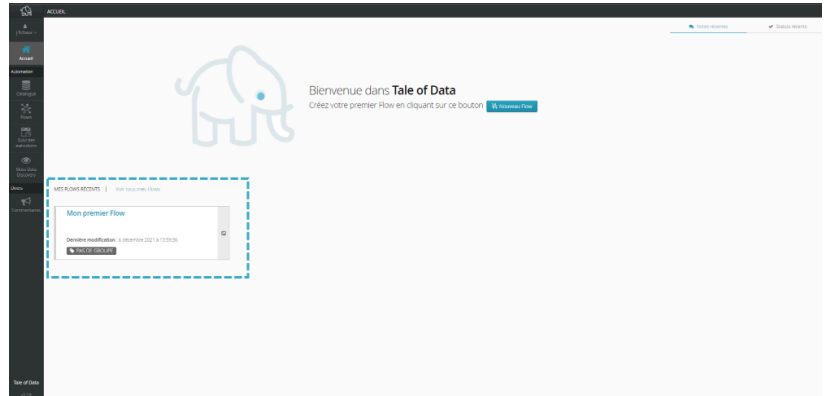
Il est par exemple possible de :

- 
- Exécuter le même traitement sur d'autres données sans refaire le Flow (= réutilisation des Flows),
  - Planifier les exécutions (quotidiennes, hebdomadaires, etc.),
  - Traiter de très grosses volumétries de données (milliards de lignes),
  - Lancer des alertes paramétrables sur tous types d'anomalies présentes dans les données.
-

# Pour aller plus loin

## Gérer ses Flows

Vous pouvez observer que votre Flow apparaît maintenant dans la partie "Mes Flows récents" de l'écran d'accueil.



1. Appuyez sur "Voir tous mes Flows".



2. Sélectionnez le Flow que vous souhaitez modifier (en cochant la case).

- 1 Ouvrir le Flow
- 2 Editer le Flow
- 3 Dupliquer le Flow
- 4 Supprimer le Flow
- 5 Partager le Flow



### Astuce

Vous pouvez, par exemple, renommer ou supprimer vos Flows depuis cette interface grâce à la barre latérale.

---

# Nœuds de traitement de la barre d'outils

---



## Note

Cette liste n'est pas exhaustive.

---

## Diffusion

Un nœud de diffusion permet de dupliquer chaque enregistrement (entrées) sur toutes les sorties.

## Union

Un nœud d'union permet d'ajouter plusieurs jeux de données en entrée (empilement), correspondant à un ajout de lignes (union au sens SQL).

## Jointure

Un nœud de jointure permet d'ajouter des informations à une donnée, correspondant à un ajout de colonnes (jointure au sens SQL).

## Enrichissement

Un nœud d'enrichissement permet, notamment en utilisant du matching flou, d'ajouter de nouveaux champs à un jeu de données (dit jeu de données à enrichir ou jeu de données n°1) à partir d'un jeu de données d'enrichissement (= jeu de données n°2, connecté par un lien bleu).

## Agrégation

Un nœud d'agrégation permet de créer des tableaux croisés.

## Tri

Un nœud de tri permet de trier les enregistrements en entrée selon différents critères.

## Validation

Un nœud de validation envoie les enregistrements valides sur sa 1ère sortie et les enregistrements invalides sur sa 2ème sortie, s'il en a une.

## Lien vers le guide de référence

Pour télécharger le guide référence :

[la section intitulée « Ecran d'accueil »](#)