

Apprendre SSIS

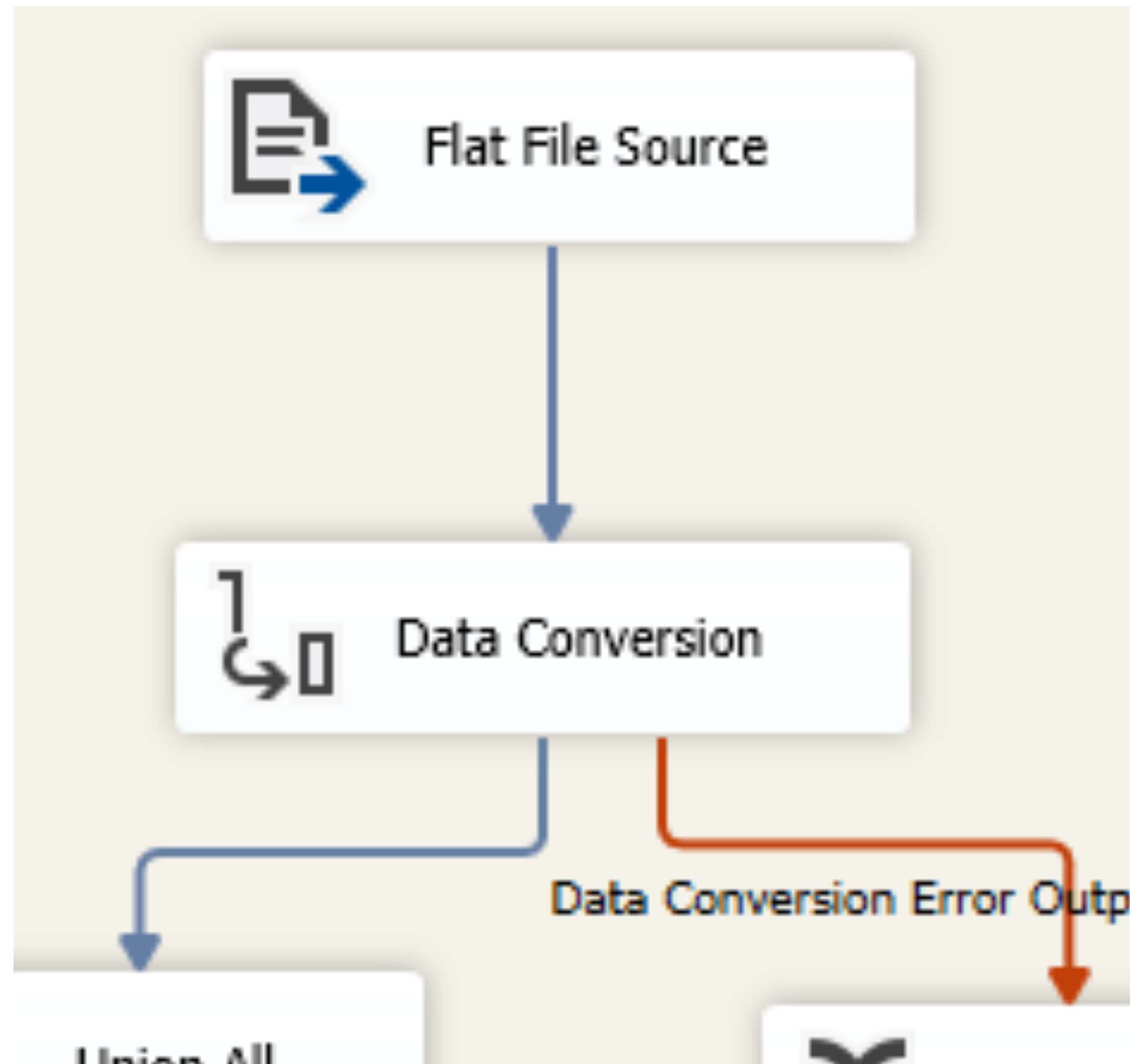
Les outils de transformations

bstorm

Something different
Something better




Conversion de données (et sortie d'erreur)



Elle convertit les données d'une colonne d'entrée en un autre type de données puis les copie dans une nouvelle colonne de sortie

Vous pouvez appliquer plusieurs conversions à une seule colonne d'entrée.

Conversion de données

 Data Conversion Transformation Editor

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the is converted, set the length, precision, scale, and code page of the column.

Available Input Columns

| | |
|-------------------------------------|---------|
| <input checked="" type="checkbox"/> | Name |
| <input type="checkbox"/> | best_of |
| <input type="checkbox"/> | round |
| <input checked="" type="checkbox"/> | minutes |
| <input checked="" type="checkbox"/> | w_ace |
| <input type="checkbox"/> | w_df |
| <input type="checkbox"/> | w_svpt |
| <input type="checkbox"/> | w_1stIn |

| Input Column | Output Alias | Data Type | Length | Precision | Scale |
|--------------|-----------------|-----------------------------------|--------|-----------|-------|
| minutes | Copy of minutes | eight-byte signed integer [DT_I8] | | | |
| w_ace | Copy of w_ace | string [DT_STR] | 50 | | |
| | | | | | |

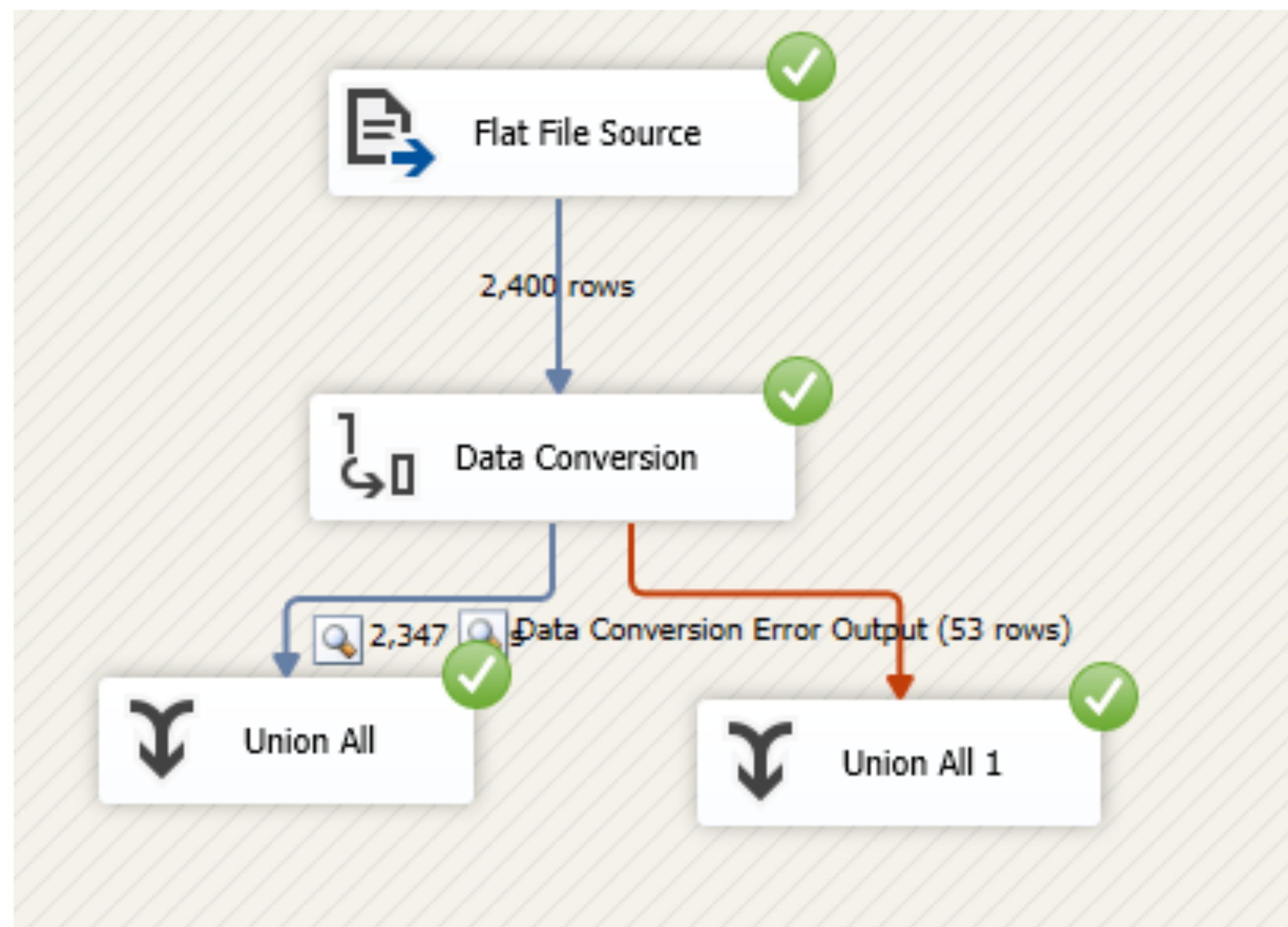
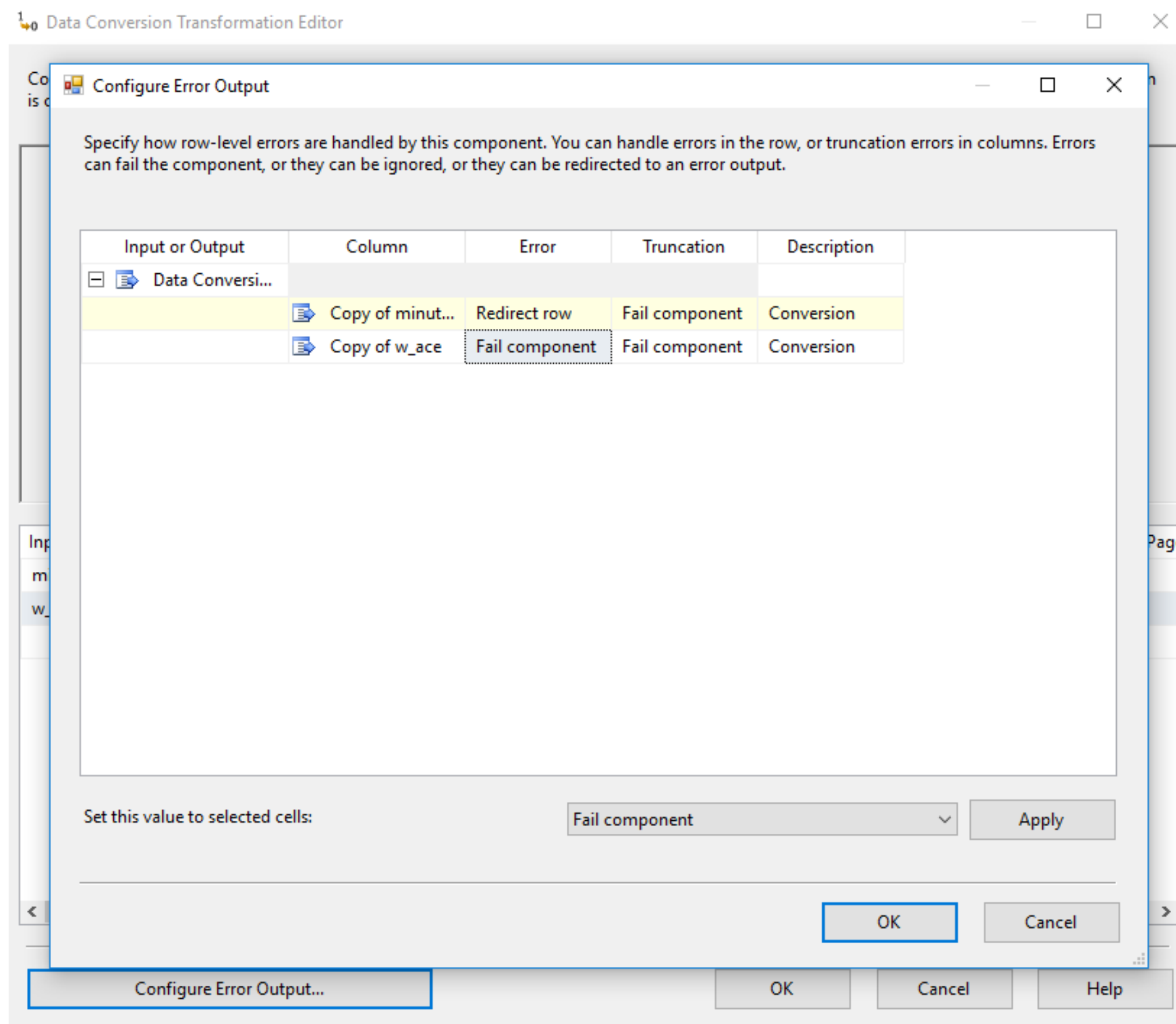
Configure Error Output...

OK

Cancel

Elle convertit les données d'une colonne d'entrée en un autre type de données puis les copie dans une nouvelle colonne de sortie

Vous pouvez appliquer plusieurs conversions à une seule colonne d'entrée.



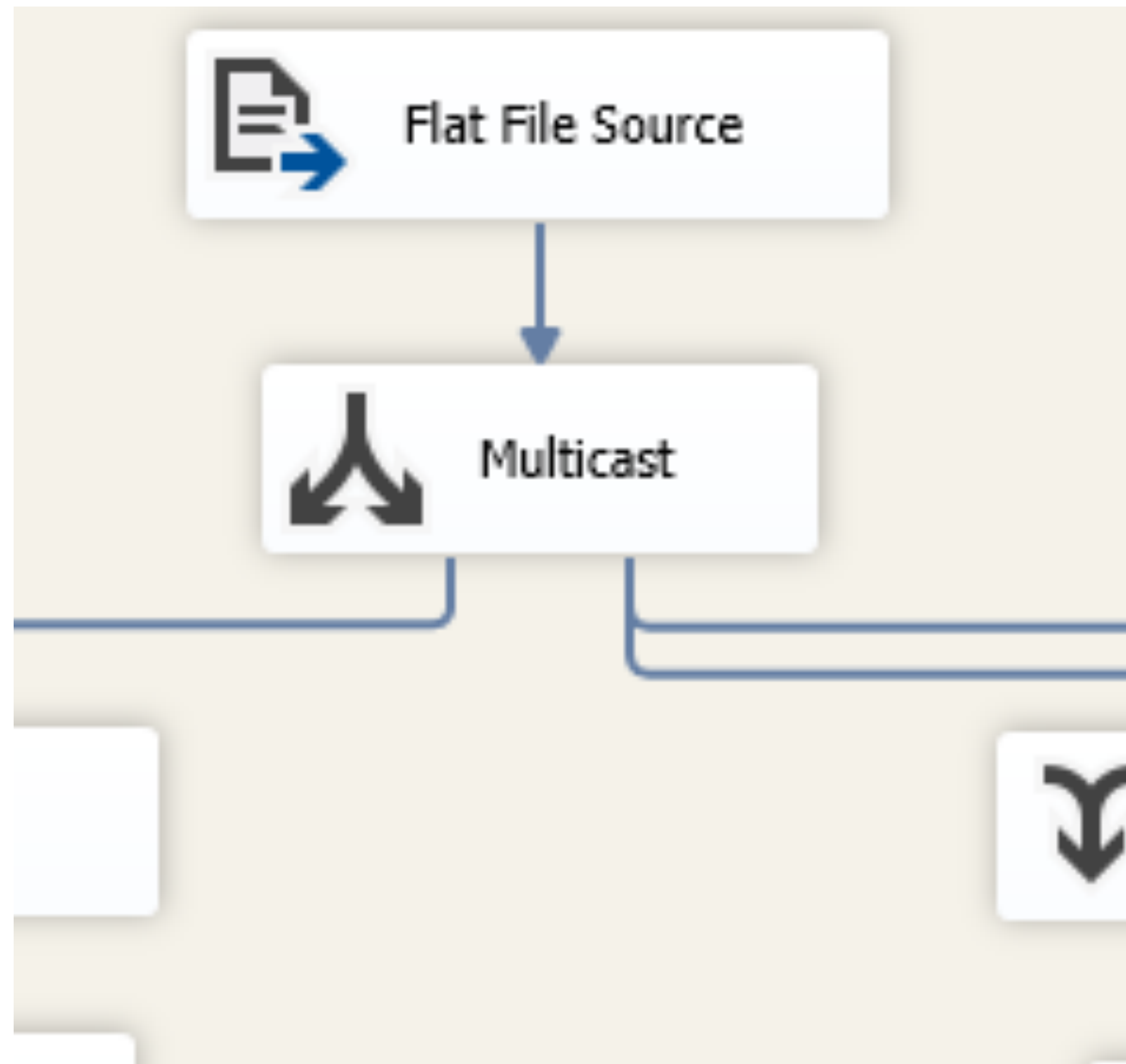
Conversion de données

| SSIS DATA TYPE | SQL SERVER DATA TYPE | .NET DATA TYPE |
|----------------|----------------------|----------------|
| DT_NUMERIC | numeric | System.Decimal |
| DT_DECIMAL | Decimal | |
| DT_CY | numeric, decimal | |
| DT_I1 | | System.Sbyte |
| DT_I2 | smallint | System.Int16 |
| DT_I4 | int | System.Int32 |
| DT_BOOL | Bit | System.Boolean |
| DT_I8 | bigint | System.Int64 |
| DT_R4 | real | System.Single |
| DT_R8 | float | System.Double |
| DT_U1 | tinyint | System.Byte |
| DT_U2 | | System.UInt16 |
| DT_U4 | | System.UInt32 |
| DT_U8 | | System.UInt64 |
| DT_GUID | Uniqueidentifier | System.Guid |

Conversion de données

| SSIS DATA TYPE | SQL SERVER DATA TYPE | .NET DATA TYPE |
|-----------------------|--------------------------------|------------------------|
| DT_WSTR | nvarchar, nchar | System.String |
| DT_STR | varchar, char | |
| DT_TEXT | text | |
| DT_NTEXT | ntext, sql_variant, xml | |
| DT_BYTES | binary, varbinary | System.Byte() |
| DT_IMAGE | timestamp, image | |
| DT_DBTIMESTAMP | smalldatetime, datetime | System.DateTime |
| DT_DBTIMESTAMP2 | datetime | |
| DT_DBDATE | Date | |
| DT_DATE | | |
| DT_FILETIME | | |
| DT_DBDATETIMEOFFSET | datetimeoffset | |
| DT_DBTIME2 | time | System.TimeSpan |
| DT_DBTIME | | |

Multidiffusion

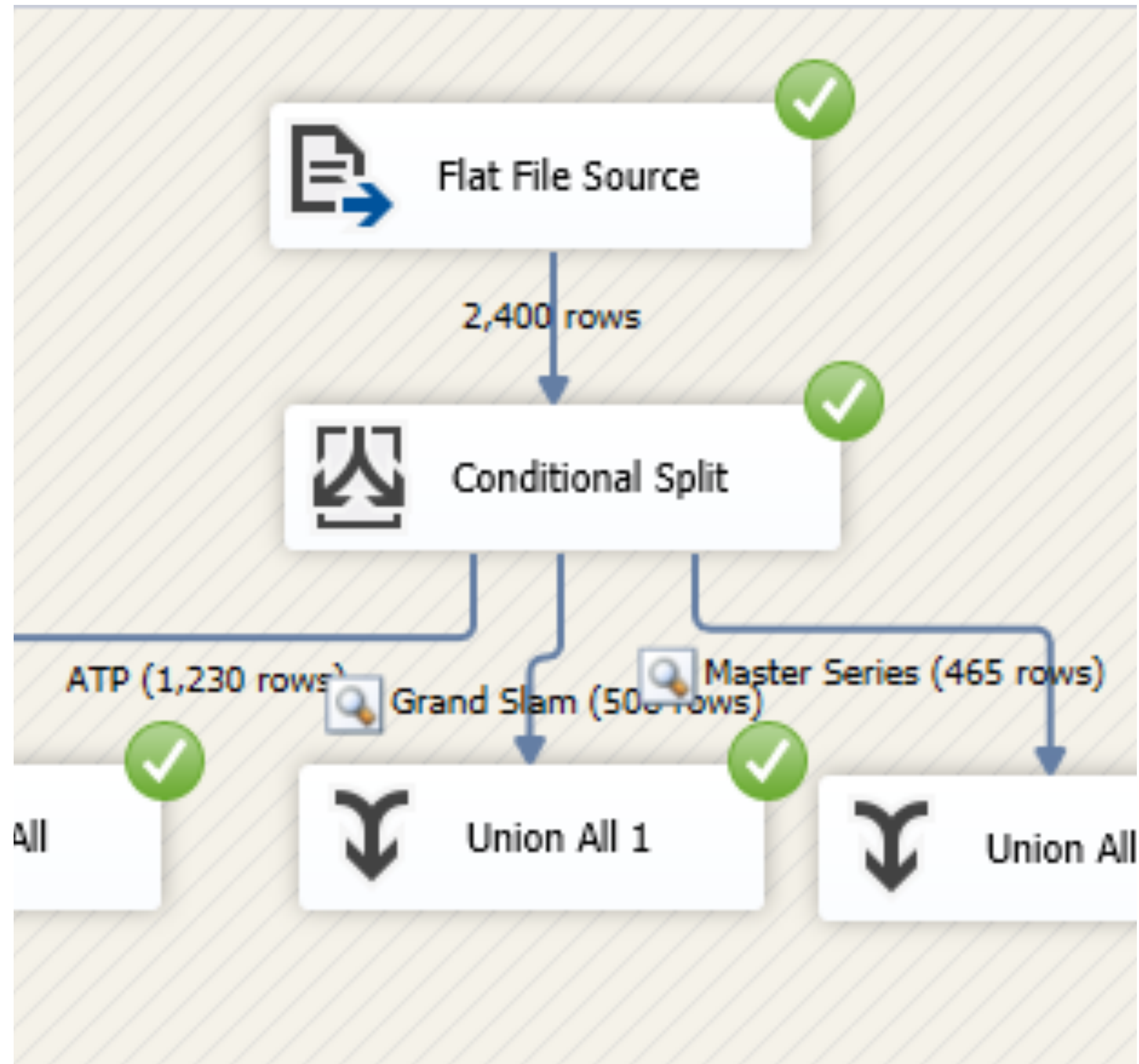


Elle distribue son entrée sur une ou plusieurs sorties.

Cette fonctionnalité est utile lorsque le package doit appliquer plusieurs ensembles de transformations aux mêmes données.

Par exemple,
une copie des données est agrégée et seules cet agrégat sera chargé dans la destination,
une autre copie des données est étendue avec des colonnes dérivées avant d'être chargée dans la destination.

Division conditionnelle



Cette transformation évalue les expressions et oriente la ligne de données vers la sortie spécifiée en fonction des résultats.

- Exemple : je souhaite répartir les matchs en fonction des types de tournois: ATP, Master Series, Grand Chelem
- Cette transformation fournit également une sortie par défaut

Division conditionnelle

Conditional Split Transformation Editor

Specify the conditions used to direct input rows to specific outputs. If an input row matches no condition, the row is directed to a default output.

Variables and Parameters

Columns

Mathematical Functions

String Functions

Date/Time Functions

NULL Functions

Type Casts

Operators

Description:

| Order | Output Name | Condition |
|-------|---------------|----------------------|
| 1 | ATP | tourney_level == "A" |
| 2 | Master Series | tourney_level == "M" |
| 3 | Grand Slam | tourney_level == "G" |
| | | |

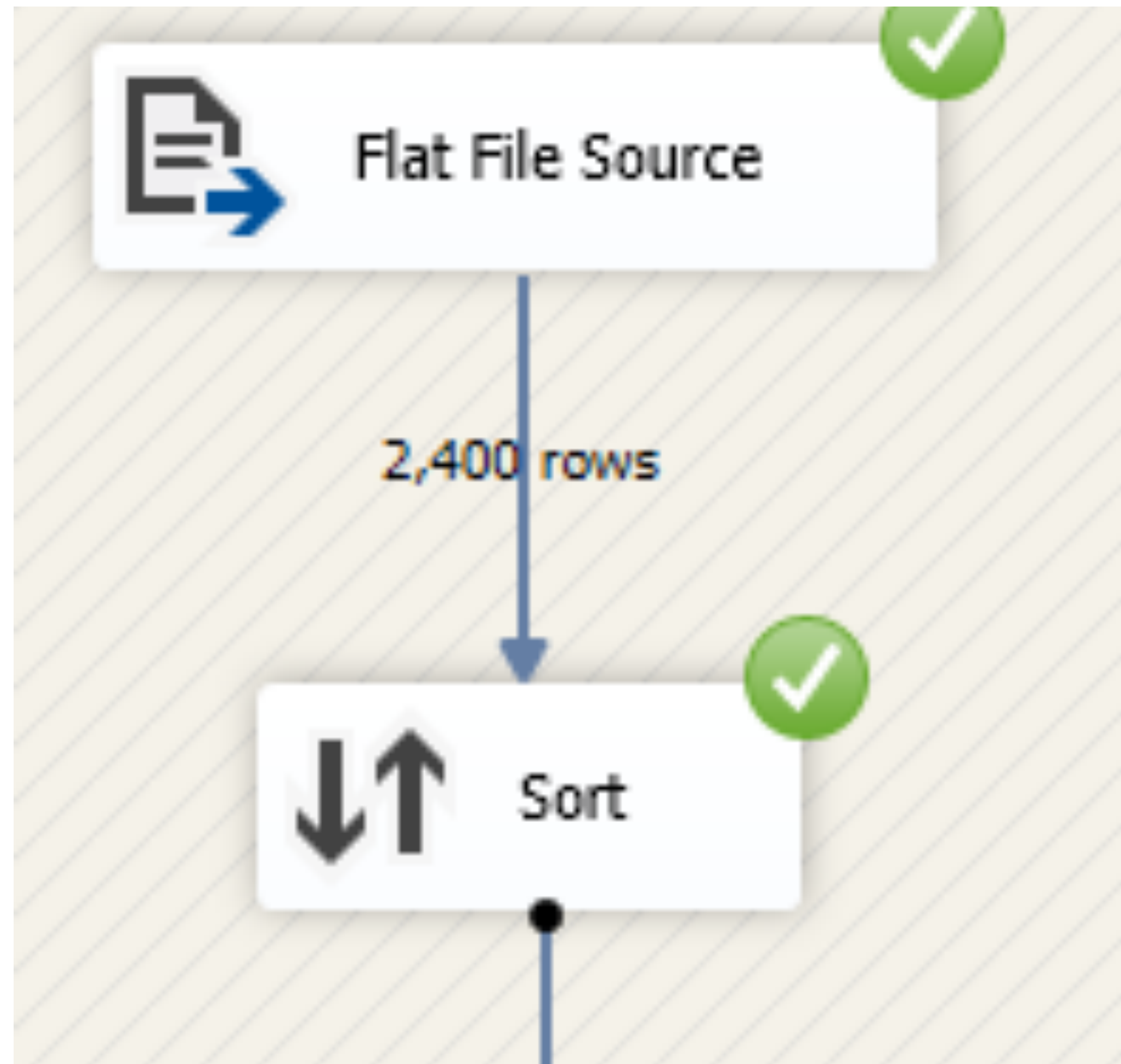
Default output name: Else

Configure Error Output... OK Cancel Help

Cette transformation évalue les expressions et oriente la ligne de données vers la sortie spécifiée en fonction des résultats.

- Exemple : je souhaite répartir les matchs en fonction des types de tournois: ATP, Master Series, Grand Chelem
- Cette transformation fournit également une sortie par défaut

Trier



Elle trie les données d'entrée dans l'ordre croissant ou décroissant.

- Vous pouvez appliquer plusieurs tris à une entrée
- chaque tri est identifié par un chiffre qui détermine l'ordre de tri

Trier

Sort Transformation Editor

Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Available Input Columns

| <input checked="" type="checkbox"/> | Name | Pass Through |
|-------------------------------------|--------------|-------------------------------------|
| <input type="checkbox"/> | tourney_name | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> | surface | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> | winner_name | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> | loser_name | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> | score | <input checked="" type="checkbox"/> |

| Input Column | Output Alias | Sort Type | Sort Order | Comp |
|--------------|--------------|-----------|------------|------|
| winner_name | winner_name | ascending | 1 | |
| | | | | |

☐ Remove rows with duplicate sort values

OK Cancel Help

Elle trie les données d'entrée dans l'ordre croissant ou décroissant.

- Vous pouvez appliquer plusieurs tris à une entrée
- chaque tri est identifié par un chiffre qui détermine l'ordre de tri
- vous pouvez supprimer les doublons en cochant la case du dessous

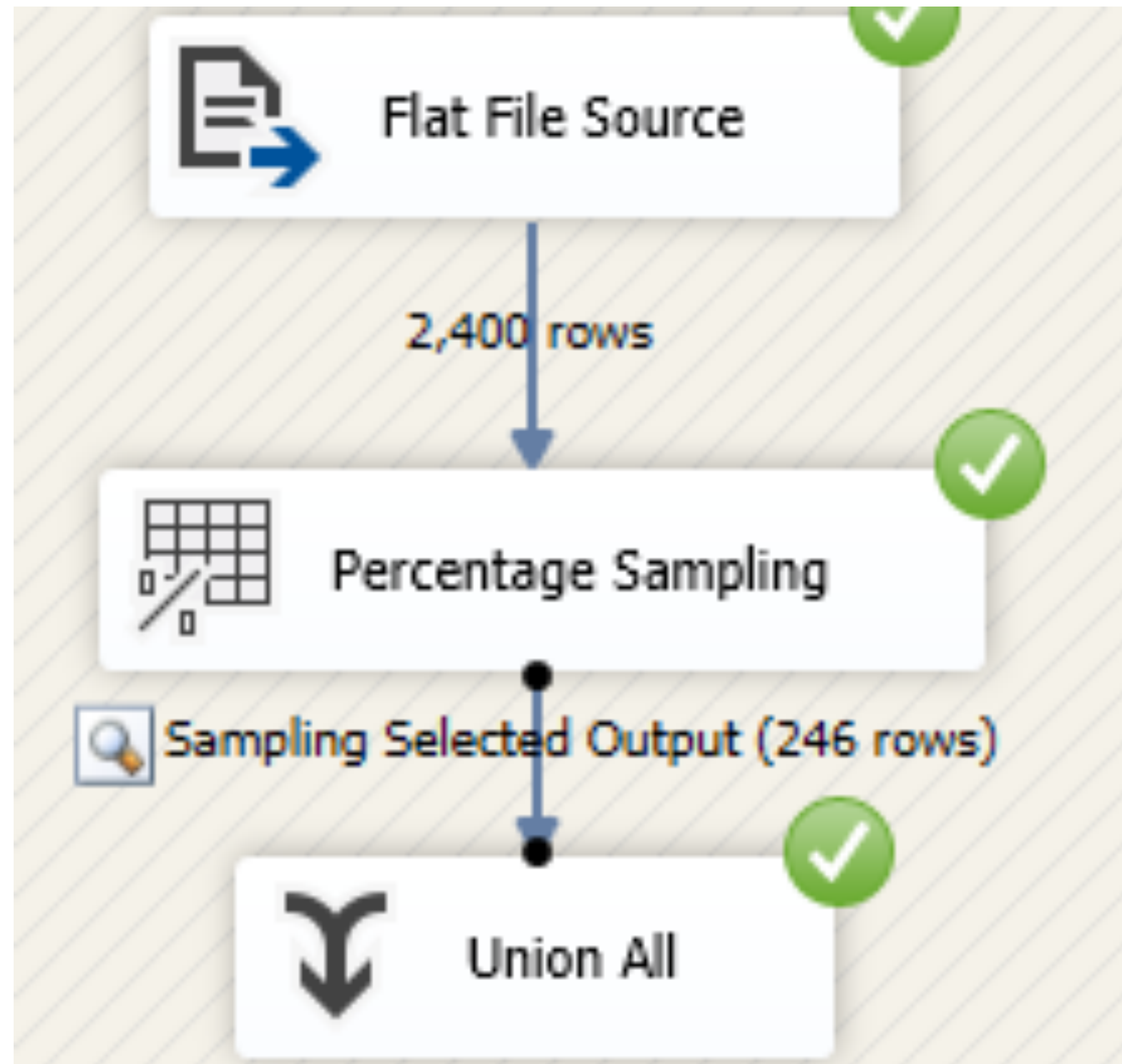
Trier

| Port Output Data Viewer at Data Flow Task | | | | |
|---|---------|---------------------------|------------------------|------------------------|
| <div><div></div><div>Detach</div><div>Copy Data</div></div> | | | | |
| toumey_name | surface | winner_name | loser_name | score |
| Wimbledon | Grass | Adam Pavlasek | Ernesto Escobedo | 6-7(7) 6-1 6-3 6-1 |
| Canada Masters | Hard | Adrian Mannarino | Daniil Medvedev | 6-3 6-1 |
| Antalya | Grass | Adrian Mannarino | Mohamed Safwat | 6-4 6-4 |
| Wimbledon | Grass | Adrian Mannarino | Gael Monfilis | 7-6(3) 4-6 5-7 6-3 6-2 |
| Antalya | Grass | Adrian Mannarino | Borna Coric | 7-5 6-4 |
| Wimbledon | Grass | Adrian Mannarino | Feliciano Lopez | 5-7 6-1 6-1 4-3 RET |
| Miami Masters | Hard | Adrian Mannarino | Paolo Lorenzi | 6-4 3-6 6-2 |
| Indian Wells Masters | Hard | Adrian Mannarino | Juan Monaco | 6-3 6-7(5) 6-1 |
| Monte Carlo | Clay | Adrian Mannarino | Guillermo Garcia Lopez | 1-6 6-3 6-4 |
| Cincinnati Masters | Hard | Adrian Mannarino | Robin Haase | 2-6 7-6(8) 6-2 |
| Monte Carlo | Clay | Adrian Mannarino | Jo Wilfried Tsonga | 6-7(3) 6-2 6-3 |
| Cincinnati Masters | Hard | Adrian Mannarino | Sam Querrey | 6-2 7-6(4) |
| Miami Masters | Hard | Adrian Mannarino | Benjamin Becker | 7-6(6) 6-3 |
| Canada Masters | Hard | Adrian Mannarino | Milos Raonic | 6-4 6-4 |
| Us Open | Hard | Adrian Mannarino | Bjorn Fratangelo | 6-3 6-7(4) 6-1 6-2 |
| 'S-Hertogenbosch | Grass | Adrian Mannarino | Jordan Thompson | 6-4 6-1 |
| Antalya | Grass | Adrian Mannarino | Andreas Seppi | 6-4 6-4 |
| Miami Masters | Hard | Adrian Mannarino | Borna Coric | 6-4 2-6 7-6(3) |
| Los Cabos | Hard | Adrian Mannarino | Bjorn Fratangelo | 6-1 6-1 |
| Canada Masters | Hard | Adrian Mannarino | Hyeon Chung | 6-3 6-3 |
| Antalya | Grass | Adrian Mannarino | Fernando Verdasco | 2-6 7-5 6-2 |
| Acapulco | Hard | Adrian Mannarino | Taylor Harry Fritz | 3-6 7-6(5) 6-4 |
| Wimbledon | Grass | Adrian Mannarino | Yuichi Sugita | 6-1 5-7 4-6 7-6(2) 6-2 |
| Us Open | Hard | Adrian Mannarino | Ricardas Berankis | 6-2 6-4 6-2 |
| Los Cabos | Hard | Adrian Mannarino | Jason Jung | 6-3 6-3 |
| Us Open | Hard | Adrian Menendez Maceiras | Patrick Kypson | 6-4 7-6(9) 6-1 |
| Newport | Grass | Adrian Menendez Maceiras | Stefan Kozlov | 6-2 7-6(5) |
| Davis Cup G1 R1: ROU vs... | | Adrian Ungur | Ilya Ivashka | 6-2 5-7 7-5 6-4 |
| Davis Cup G2 R1: PAK vs ... | | Aisam Ul Haq Qureshi | Anoosha Shahgholi | 6-1 6-2 6-2 |
| Los Cabos | Hard | Akira Santillan | Konstantin Kravchuk | 6-0 6-1 |
| Newport | Grass | Akira Santillan | Michael Mmoh | 6-1 6-2 |
| Attached | | Total rows: 0, buffers: 0 | | Rows displayed = 2400 |

Elle trie les données d'entrée dans l'ordre croissant ou décroissant.

- Vous pouvez appliquer plusieurs tris à une entrée
- chaque tri est identifié par un chiffre qui détermine l'ordre de tri
- vous pouvez supprimer les doublons en cochant la case du dessous

Echantillonnage en pourcentage



Elle crée un échantillon de données en sélectionnant un pourcentage des lignes d'entrée de transformation via une sélection aléatoire.

- particulièrement utile pour l'exploration de données:
- vous pouvez diviser au hasard un ensemble de données en deux ensembles de données: un pour l'apprentissage du modèle d'exploration de données et un pour le test du modèle

Echantillonnage en pourcentage

Percentage Sampling Transformation Editor

Configure the properties to divide the transformation input into a sample and a leftovers output, using a percentage to specify the sample size.

Percentage of rows:

10

Sample output name:

Sampling Selected Output

Unselected output name:

Sampling Unselected Output

☐ Use the following random seed:

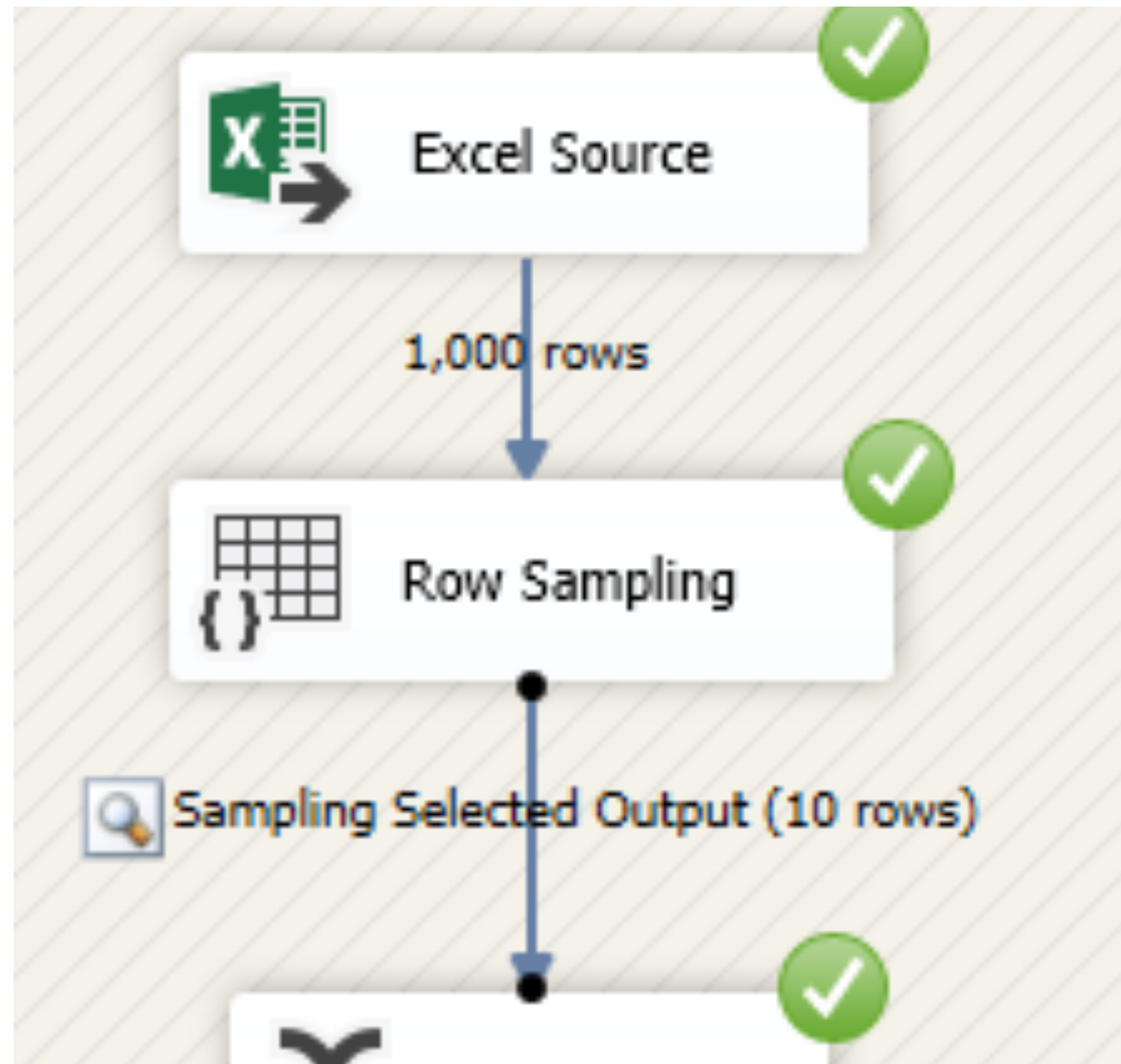
1

OK Cancel Help

Elle crée un échantillon de données en sélectionnant un pourcentage des lignes d'entrée de transformation via une sélection aléatoire.

- particulièrement utile pour l'exploration de données:
- vous pouvez diviser au hasard un ensemble de données en deux ensembles de données: un pour l'apprentissage du modèle d'exploration de données et un pour le test du modèle

Echantillonnage de lignes



Elle est utilisée pour obtenir un sous-ensemble sélectionné au hasard.

- Vous pouvez spécifier la taille exacte de l'échantillon de sortie
- Par exemple, une entreprise qui souhaitait sélectionner au hasard 50 employés pour recevoir des prix dans une loterie pourrait utiliser la transformation d'échantillonnage de lignes dans la base de données des employés pour générer le nombre exact de gagnants.

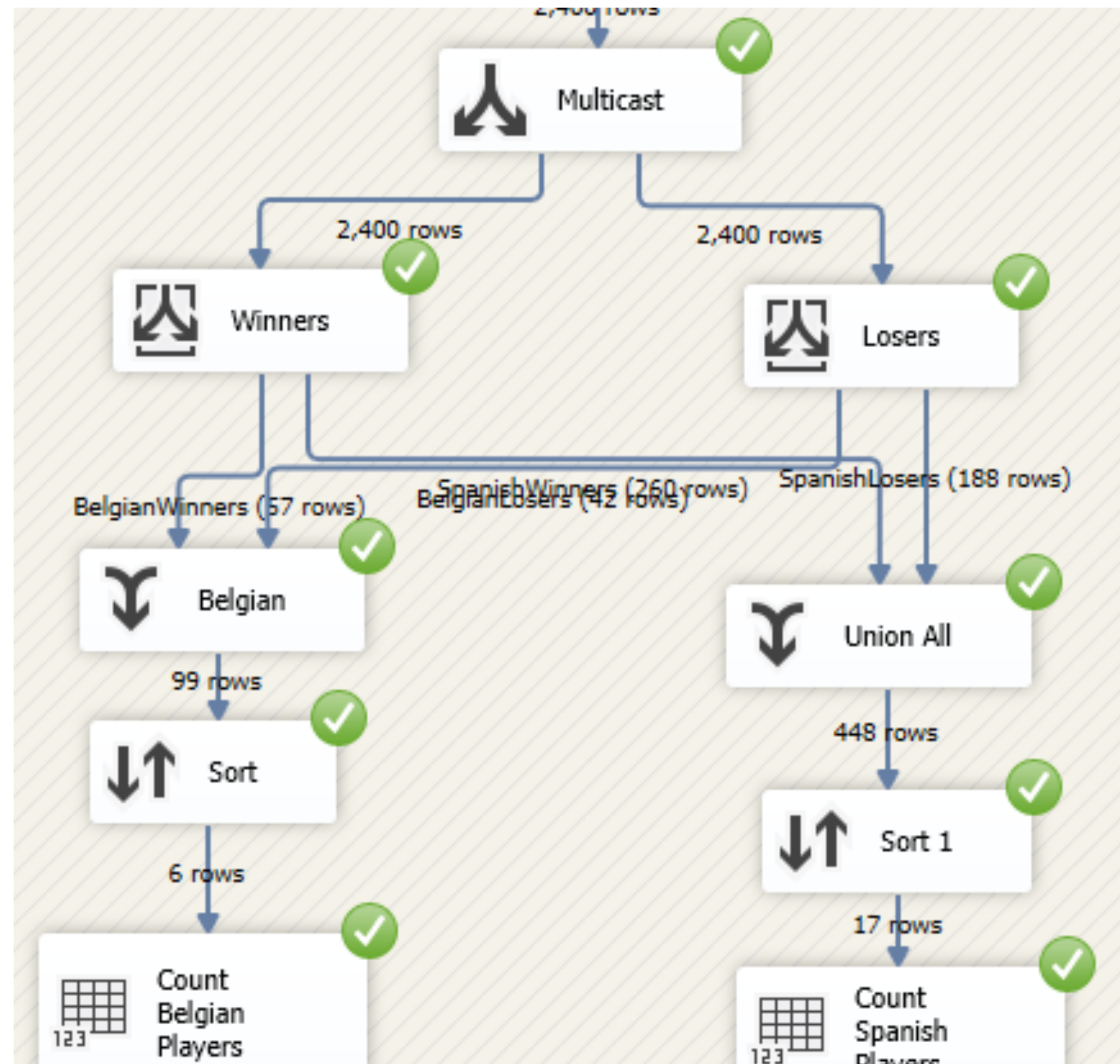
Echantillonnage de lignes

The screenshot shows the 'Row Sampling Transformation Editor' window. The title bar reads 'Row Sampling Transformation Editor'. Below the title bar, a subtitle states: 'Configure the properties used to divide the transformation input into a sample and a leftovers output, using a number to specify the sample size.' The main area is divided into two sections: 'Sampling' (selected) and 'Columns'. The 'Sampling' section contains the following fields: 'Number of rows:' with a spinner box set to 10; 'Sample output name:' with a text box containing 'Sampling Selected Output'; 'Unselected output name:' with a text box containing 'Sampling Unselected Output'; and a checkbox labeled 'Use the following random seed:' which is unchecked, with a spinner box set to 1 below it. At the bottom right, there are three buttons: 'OK', 'Cancel', and 'Help'.

Elle est utilisée pour obtenir un sous-ensemble sélectionné au hasard.

- Vous pouvez spécifier la taille exacte de l'échantillon de sortie
- Par exemple, une entreprise qui souhaitait sélectionner au hasard 50 employés pour recevoir des prix dans une loterie pourrait utiliser la transformation d'échantillonnage de lignes dans la base de données des employés pour générer le nombre exact de gagnants.



Nombre de lignes

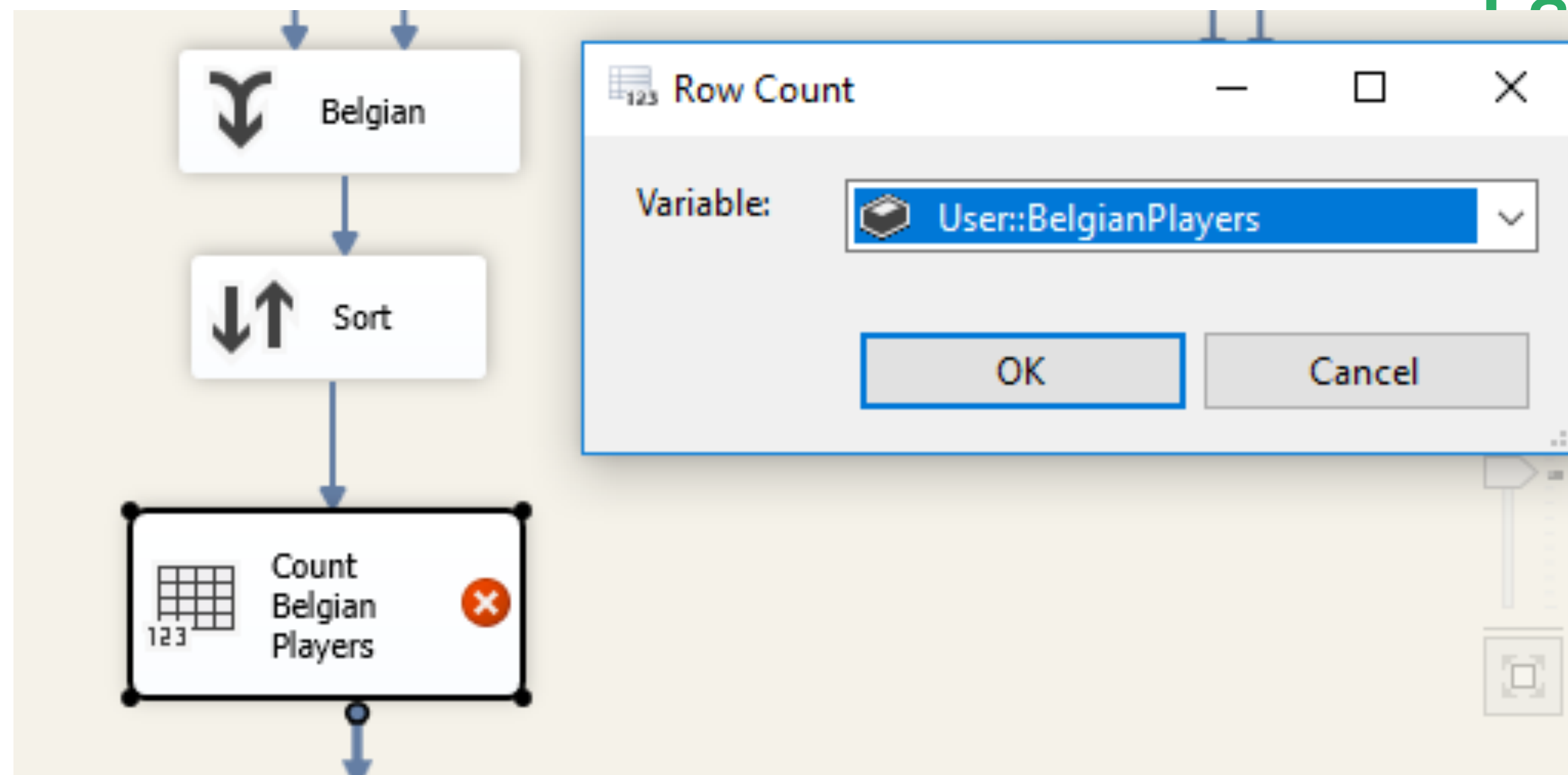


Elle compte les lignes lors de leur passage dans un flux de données et stocke le résultat final dans une variable pré-existante.

- permet de mettre à jour les variables utilisées dans les scripts, les expressions et les expressions de propriété
- permet de mettre à jour le texte du message dans un message électronique pour inclure le nombre de lignes
- Exemple: Stocke le nombre de joueurs belges dans une variable et le nombre de joueurs espagnols dans une autre



Nombre de lignes

| Variables | | | | | |
|--|---------|-----------|-------|------------|-----|
| Name | Scope | Data type | Value | Expression | |
|  BelgianPlayers | Package | Int32 | 0 | | ... |
|  SpanishPlayers | Package | Int32 | 0 | | ... |












La variable utilisée par la transformation n'existe pas déjà exister

Nombre de lignes

| Watch 1 | | | ▼ | 📌 | ✕ |
|--|-------|-------|---|---|---|
| Name | Value | Type | | | |
| ▶  User::BelgianPlayers | {6} | Int32 | | | |
| ▶  User::SpanishPlayers | {17} | Int32 | | | |
| | | | | | |

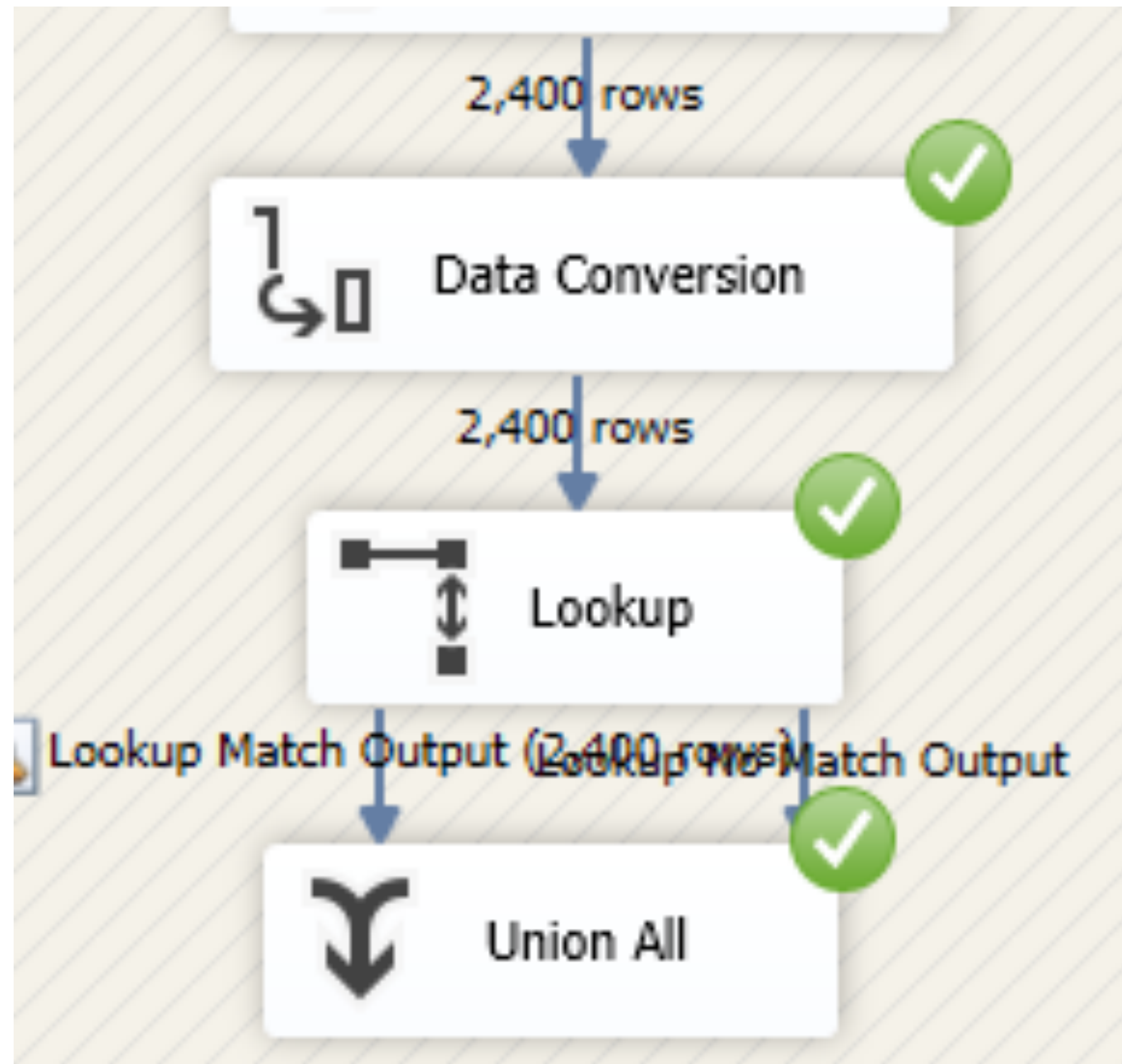
Set Breakpoints - Row Count

Select the breakpoints in the task, For Loop, Foreach Loop, or Sequence to enable. Optionally, select the number of times a breakpoint is ignored before execution is suspended on the breakpoint.

| Enabl... | Break Condition | Hit Count Type | Hit Count |
|-------------------------------------|---|----------------|-----------|
| <input type="checkbox"/> |  Break when the container receives the OnPreExecute event | Always | 0 |
| <input checked="" type="checkbox"/> |  Break when the container receives the OnPostExecute event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnError event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnWarning event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnInformation event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnTaskFailed event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnProgress event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnQueryCancel event | Always | 0 |
| <input type="checkbox"/> |  Break when the container receives the OnVariableValueChanged... | Always | 0 |

✕ Pour voir la valeur de la variable, vous devez définir un point d'arrêt PostExecute

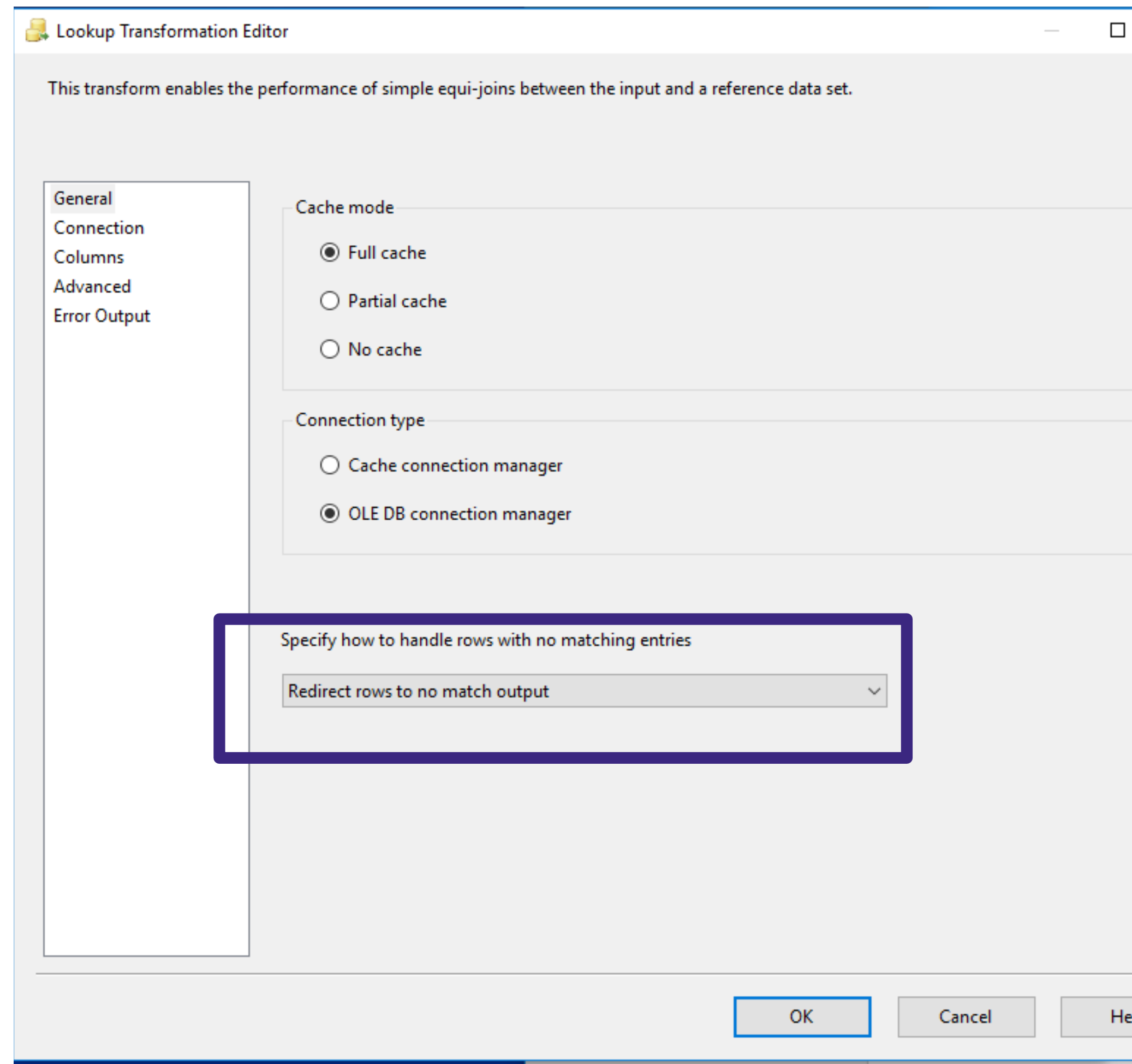
Recherche



Elle effectue des recherches en joignant les données des colonnes d'entrée aux colonnes d'un jeu de données de référence

- Permet d'accéder à des **informations supplémentaires** dans une table associée basée sur des valeurs dans des colonnes communes.
- effectue une **équijointure** entre les valeurs dans l'entrée de transformation et les valeurs dans l'ensemble de données de référence

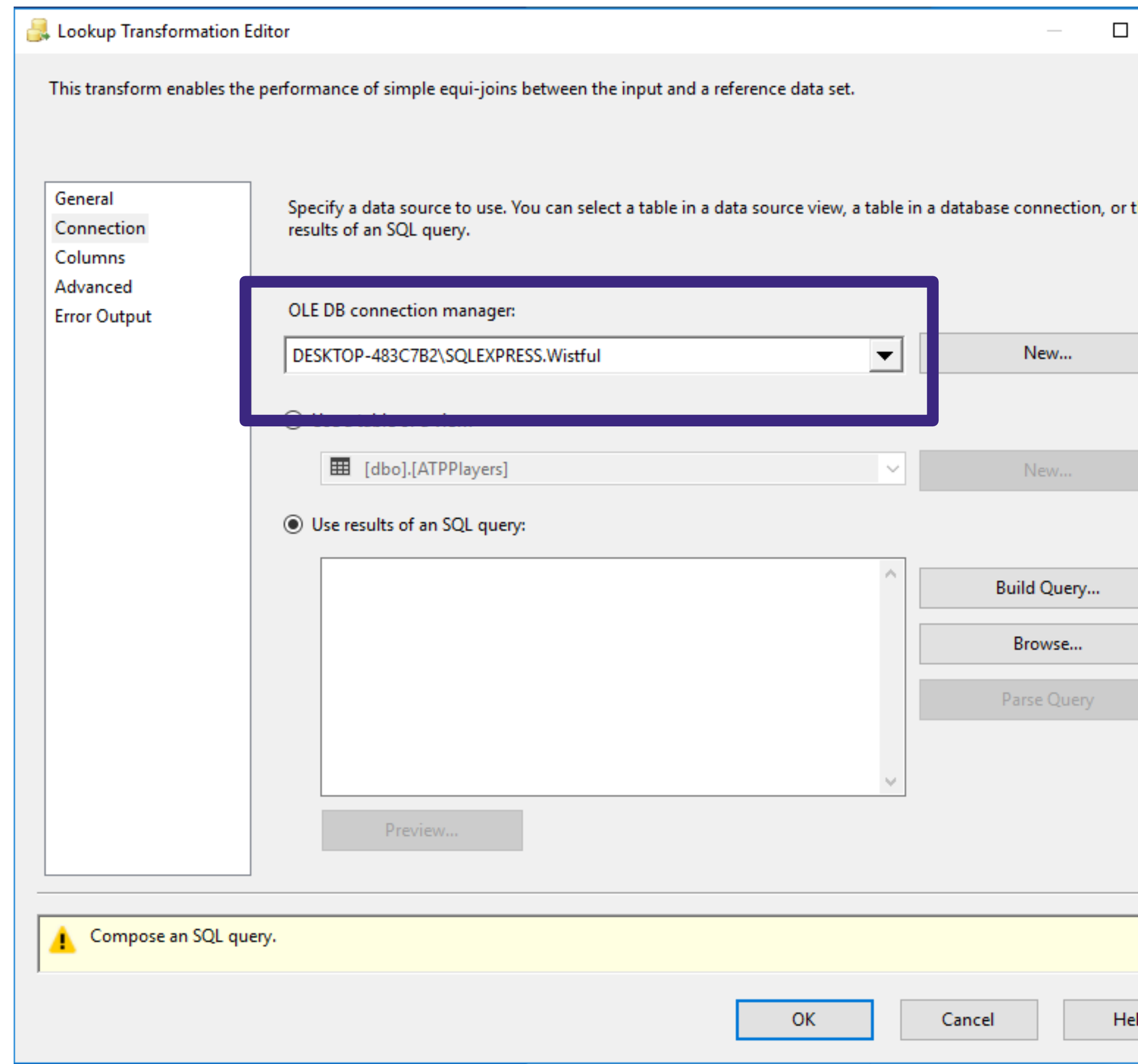
Recherche



Elle effectue des recherches en joignant les données des colonnes d'entrée aux colonnes d'un jeu de données de référence

- Permet d'accéder à des **informations supplémentaires** dans une table associée basée sur des valeurs dans des colonnes communes.
- effectue une **équijointure** entre les valeurs dans l'entrée de transformation et les valeurs dans l'ensemble de données de référence

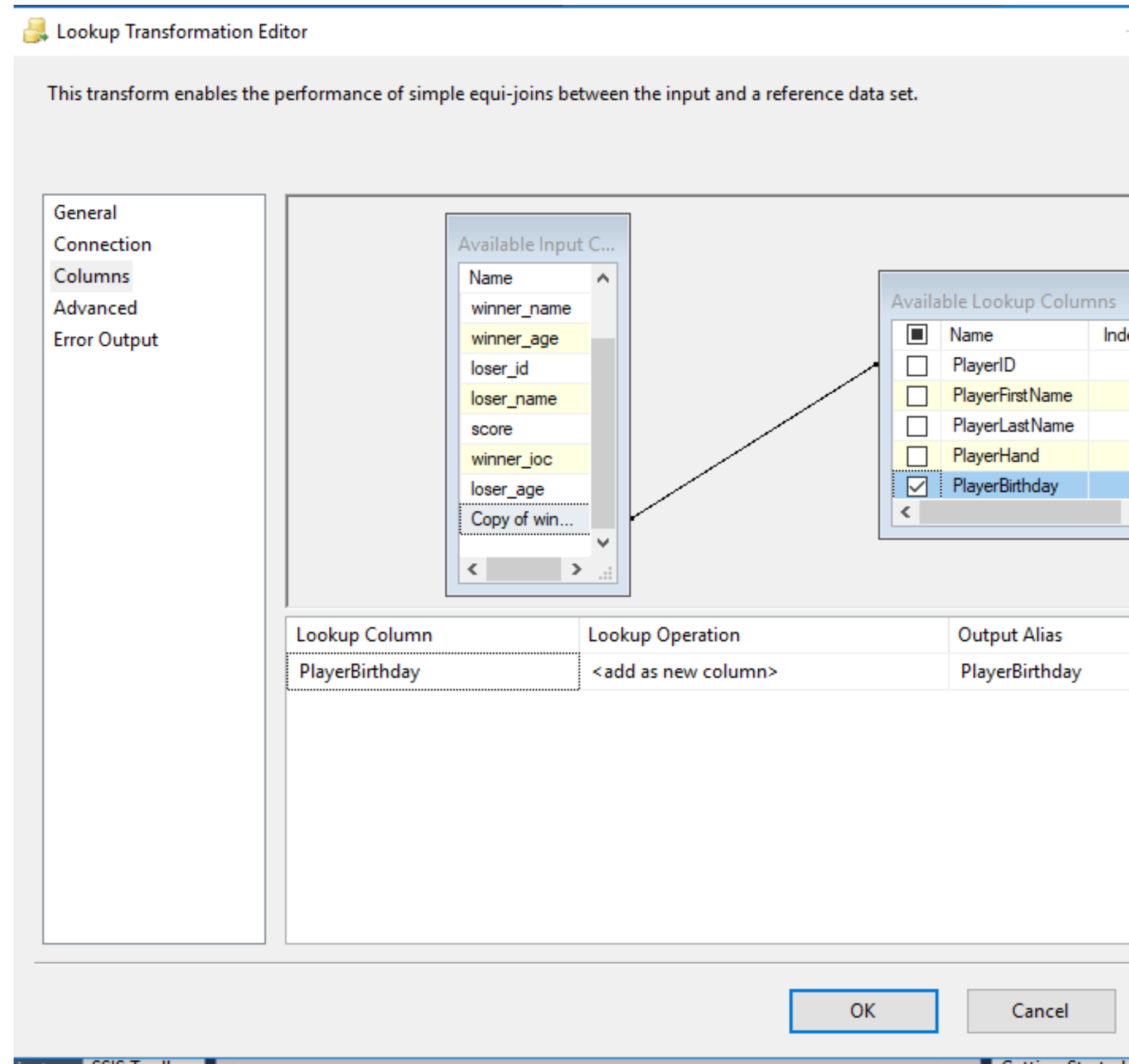
Recherche



Elle effectue des recherches en joignant les données des colonnes d'entrée aux colonnes d'un jeu de données de référence

- Permet d'accéder à des **informations supplémentaires** dans une table associée basée sur des valeurs dans des colonnes communes.
- effectue une **équi-jointure** entre les valeurs dans l'entrée de transformation et les valeurs dans l'ensemble de données de référence

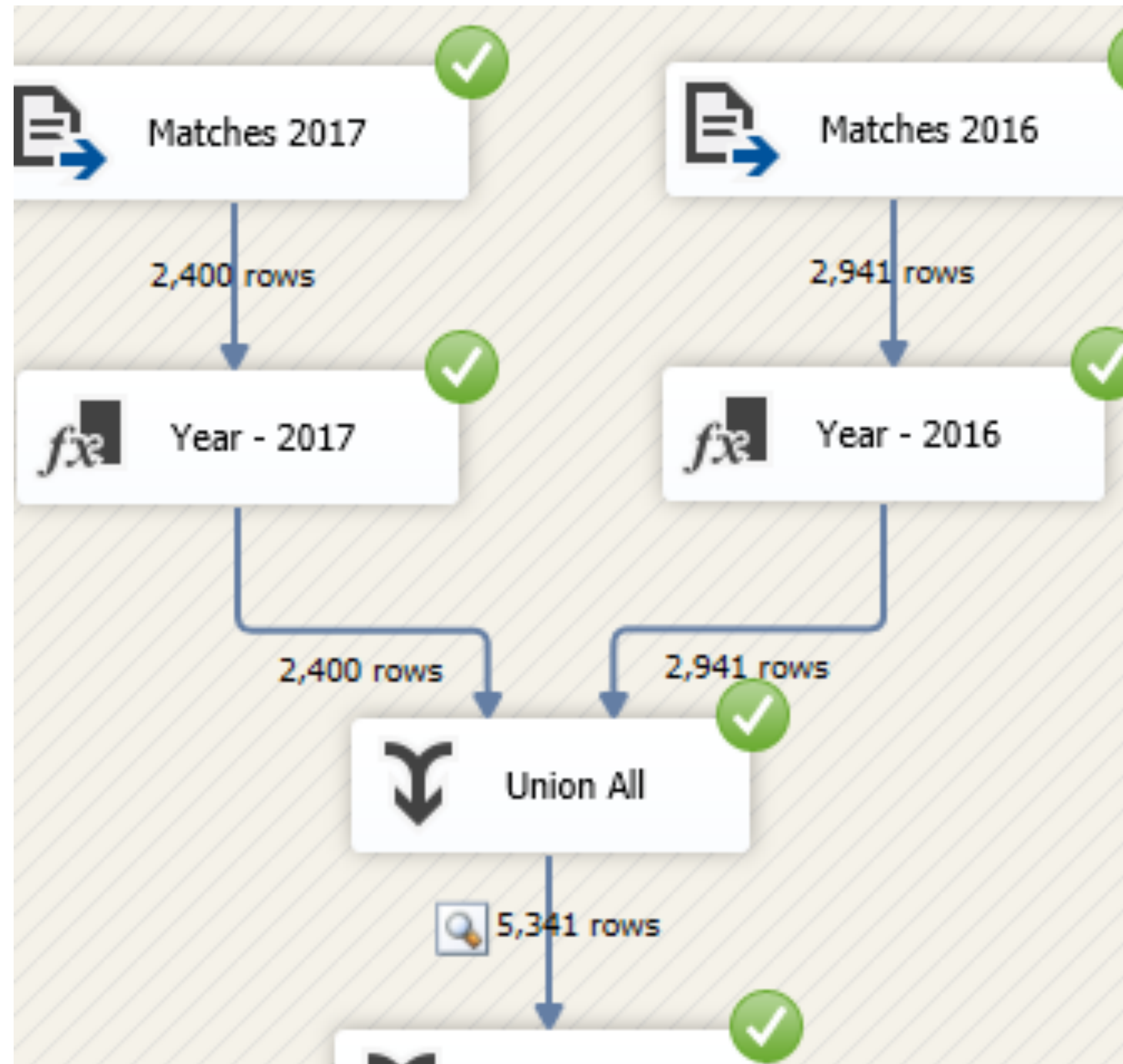
Recherche



Elle effectue des recherches en joignant les données des colonnes d'entrée aux colonnes d'un jeu de données de référence

- Permet d'accéder à des **informations supplémentaires** dans une table associée basée sur des valeurs dans des colonnes communes.
- effectue une **équi-jointure** entre les valeurs dans l'entrée de transformation et les valeurs dans l'ensemble de données de référence

Unir tout



Elle combine plusieurs entrées en une seule sortie sans tri

- Par exemple, les sorties de cinq sources de fichiers plats différentes peuvent être des entrées pour la transformation Union All et combinées en une seule sortie
- aucune réorganisation des lignes ne se produit : **si le package nécessite une sortie triée, vous devez plutôt utiliser la transformation Fusionner**

Unir tout

Union All Transformation Editor

Configure the properties used to merge multiple inputs into one output by creating mappings between columns.

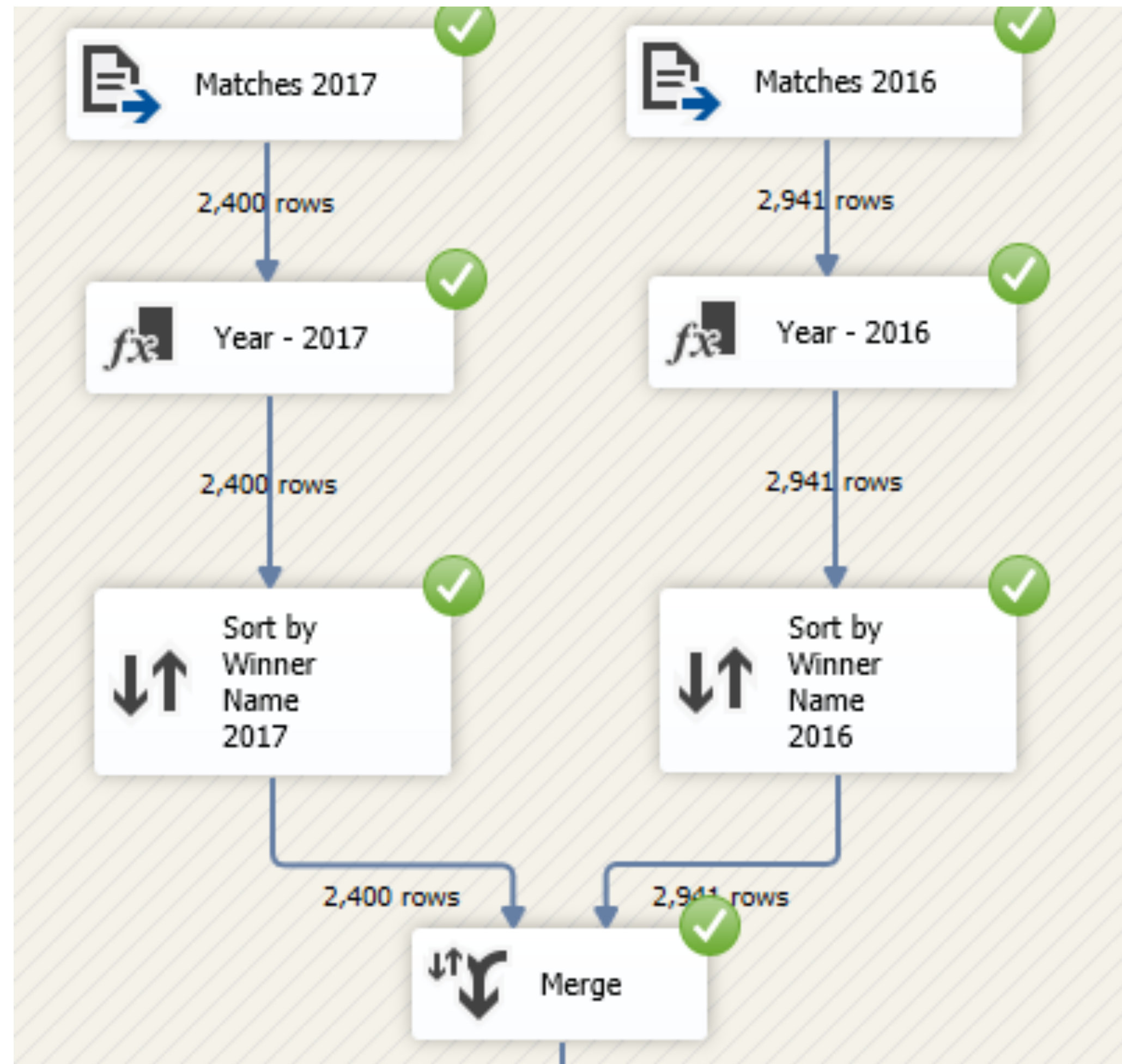
| Output Column Name | Union All Input 1 | Union All Input 2 |
|--------------------|-------------------|-------------------|
| tourney_id | tourney_id | tourney_id |
| tourney_name | tourney_name | tourney_name |
| surface | surface | surface |
| draw_size | draw_size | draw_size |
| tourney_level | tourney_level | tourney_level |
| tourney_date | tourney_date | tourney_date |
| winner_id | winner_id | winner_id |
| winner_name | winner_name | winner_name |
| loser_id | loser_id | loser_id |
| loser_name | loser_name | loser_name |
| score | score | score |
| Year | Year | Year |

OK Cancel

Elle combine plusieurs entrées en une seule sortie sans tri

- Par exemple, les sorties de cinq sources de fichiers plats différentes peuvent être des entrées pour la transformation Union All et combinées en une seule sortie
- aucune réorganisation des lignes ne se produit : **si le package nécessite une sortie triée, vous devez plutôt utiliser la transformation Fusionner**

Fusionner



Elle combine deux jeux de données triés en un seul jeu de donnée

- permet de fusionner les données de deux sources de données, telles que des tables et des fichiers
- Les lignes sont triées en fonction des valeurs de leurs colonnes clés
- En incluant la transformation de fusion dans un flux de données, vous pouvez effectuer les tâches suivantes:
- Créer des jeux de données complexes en imbriquant des transformations de fusion
- Remerge des lignes après correction des erreurs dans les données
- Exemple : Je veux que la sortie soit triée par gagnant

Fusionner

Merge Transformation Editor

Configure the properties used to merge two sorted inputs into one output by creating mappings between columns.

| Output Column Name | Merge Input 1 | Merge Input 2 |
|---------------------------|---------------------------|---------------------------|
| tourney_name | tourney_name | tourney_name |
| surface | surface | surface |
| winner_name (Sort key: 1) | winner_name (Sort key: 1) | winner_name (Sort key: 1) |
| loser_name | loser_name | loser_name |
| score | score | score |
| Year | Year | Year |

OK Cancel

Elle combine deux jeux de données triés en un seul jeu de donnée

- permet de fusionner les données de deux sources de données, telles que des tables et des fichiers
- Les lignes sont triées en fonction des valeurs de leurs colonnes clés
- En incluant la transformation de fusion dans un flux de données, vous pouvez effectuer les tâches suivantes:
- Créer des jeux de données complexes en imbriquant des transformations de fusion
- Remerge des lignes après correction des erreurs dans les données
- Exemple : Je veux que la sortie soit triée par gagnant

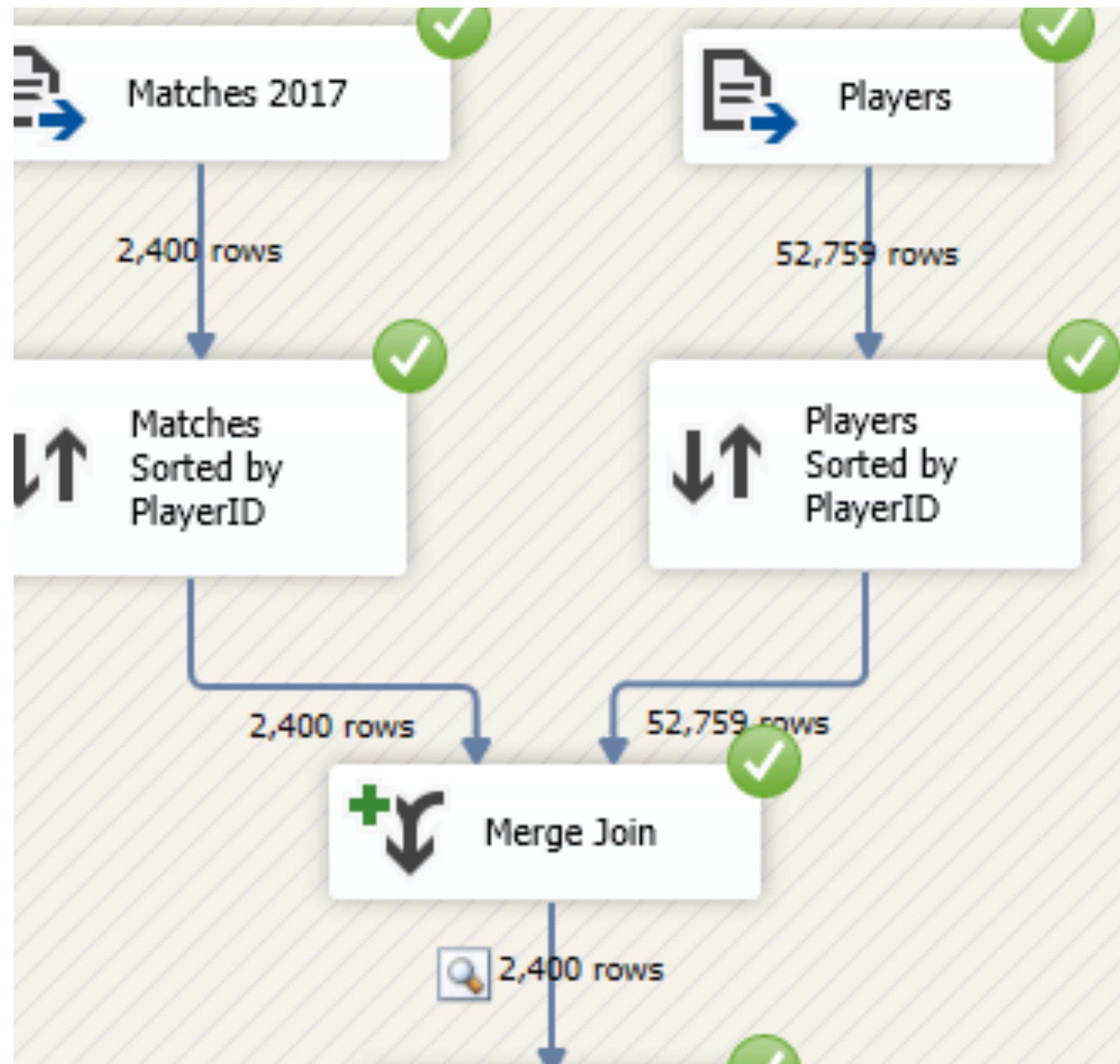
Fusionner

| ▶ Detach Copy Data | | | | |
|---|------------------|-------------------------|------------------------|------|
| surface | winner_name | loser_name | score | year |
| Grass | Adam Pavlasek | Ernesto Escobedo | 6-7(7) 6-1 6-3 6-1 | 2017 |
| Clay | Adam Pavlasek | Maximo Gonzalez | 6-3 6-4 | 2016 |
| Clay | Adam Pavlasek | Marcel Granollers | 6-2 6-4 | 2016 |
| Clay | Adam Pavlasek | Roberto Carballes Baena | 6-2 4-6 6-3 1-6 6-1 | 2016 |
| Hard | Adrian Mannarino | Daniil Medvedev | 6-3 6-1 | 2017 |
| Grass | Adrian Mannarino | Mohamed Safwat | 6-4 6-4 | 2017 |
| Grass | Adrian Mannarino | Gael Monfils | 7-6(3) 4-6 5-7 6-3 6-2 | 2017 |
| Grass | Adrian Mannarino | Borna Coric | 7-5 6-4 | 2017 |
| Grass | Adrian Mannarino | Feliciano Lopez | 5-7 6-1 6-1 4-3 RET | 2017 |
| Hard | Adrian Mannarino | Paolo Lorenzi | 6-4 3-6 6-2 | 2017 |
| Hard | Adrian Mannarino | Juan Monaco | 6-3 6-7(5) 6-1 | 2017 |
| Clay | Adrian Mannarino | Guillermo Garcia Lopez | 1-6 6-3 6-4 | 2017 |
| Hard | Adrian Mannarino | Robin Haase | 2-6 7-6(8) 6-2 | 2017 |
| Clay | Adrian Mannarino | Jo Wilfried Tsonga | 6-7(3) 6-2 6-3 | 2017 |
| Hard | Adrian Mannarino | Sam Querrey | 6-2 7-6(4) | 2017 |
| Hard | Adrian Mannarino | Benjamin Becker | 7-6(6) 6-3 | 2017 |
| Hard | Adrian Mannarino | Milos Raonic | 6-4 6-4 | 2017 |
| Hard | Adrian Mannarino | Bjorn Fratangelo | 6-3 6-7(4) 6-1 6-2 | 2017 |
| Grass | Adrian Mannarino | Jordan Thompson | 6-4 6-1 | 2017 |
| Grass | Adrian Mannarino | Andreas Seppi | 6-4 6-4 | 2017 |
| Hard | Adrian Mannarino | Borna Coric | 6-4 2-6 7-6(3) | 2017 |
| Hard | Adrian Mannarino | Bjorn Fratangelo | 6-1 6-1 | 2017 |
| Hard | Adrian Mannarino | Hyeon Chung | 6-3 6-3 | 2017 |
| Grass | Adrian Mannarino | Fernando Verdasco | 2-6 7-5 6-2 | 2017 |
| Hard | Adrian Mannarino | Taylor Harry Fritz | 3-6 7-6(5) 6-4 | 2017 |
| Grass | Adrian Mannarino | Yuichi Sugita | 6-1 5-7 4-6 7-6(2) 6-2 | 2017 |
| Hard | Adrian Mannarino | Ricardas Berankis | 6-2 6-4 6-2 | 2017 |
| Hard | Adrian Mannarino | Jason Jung | 6-3 6-3 | 2017 |
| Clay | Adrian Mannarino | Robin Haase | 6-7(4) 6-3 6-4 | 2016 |
| Grass | Adrian Mannarino | Guillermo Garcia Lopez | 6-4 6-2 | 2016 |

Elle combine deux jeux de données triés en un seul jeu de donnée

- permet de fusionner les données de deux sources de données, telles que des tables et des fichiers
- Les lignes sont triées en fonction des valeurs de leurs colonnes clés
- En incluant la transformation de fusion dans un flux de données, vous pouvez effectuer les tâches suivantes:
- Créer des jeux de données complexes en imbriquant des transformations de fusion
- Remerge des lignes après correction des erreurs dans les données
- Exemple : Je veux que la sortie soit triée par gagnant

Jointure de fusion



Elle combine deux jeux de données préalablement triés en rajoutant à une table les données d'une autre table à travers une jointure.

- Se combine toujours avec la transformation « Tri » qui survient en amont
- Exemple : Je veux tous les matchs de chaque tournoi et je veux que la date de naissance du vainqueur soit ajoutée à la sortie

Jointure de fusion

Merge Join Transformation Editor

Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.

Join type: Left outer join Swap Inputs

Matches Sorted by PlayerID

| <input checked="" type="checkbox"/> | Name | Order | Join Key |
|-------------------------------------|--------------|-------|-------------------------------------|
| <input checked="" type="checkbox"/> | tourney_id | 0 | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> | tourney_name | 0 | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> | surface | 0 | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> | winner_id | 1 | <input checked="" type="checkbox"/> |
| <input checked="" type="checkbox"/> | winner_name | 0 | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> | loser_name | 0 | <input type="checkbox"/> |

Players Sorted by PlayerID

| <input checked="" type="checkbox"/> | Name | Order | Join Key |
|-------------------------------------|-----------|-------|-------------------------------------|
| <input type="checkbox"/> | PlayerID | 1 | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> | FirstName | 0 | <input type="checkbox"/> |
| <input type="checkbox"/> | LastName | 0 | <input type="checkbox"/> |
| <input type="checkbox"/> | Hand | 0 | <input type="checkbox"/> |
| <input checked="" type="checkbox"/> | Birthdate | 0 | <input type="checkbox"/> |
| <input type="checkbox"/> | Country | 0 | <input type="checkbox"/> |

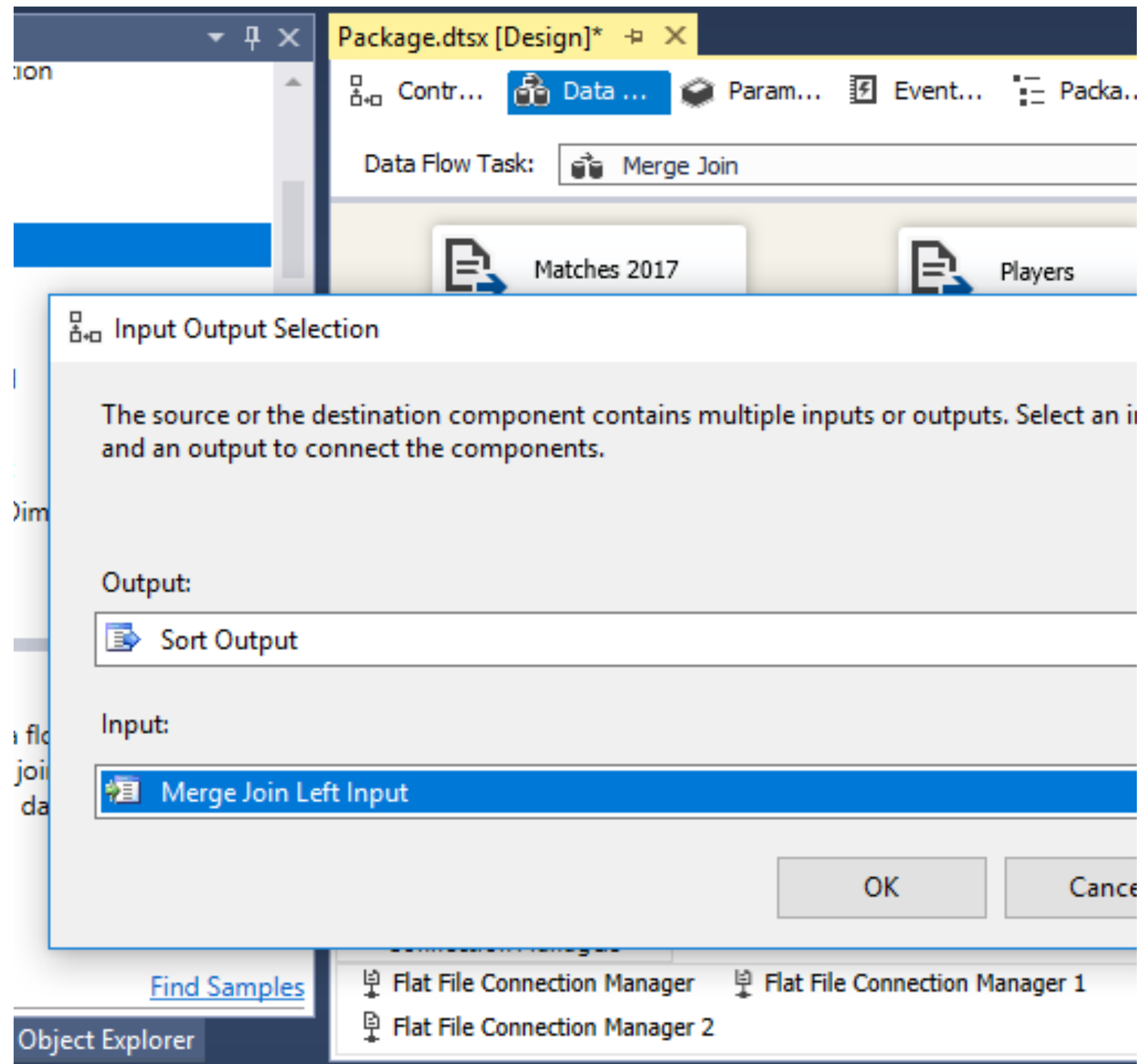
| Input | Input Column | Output Alias |
|---------------------|--------------|--------------|
| Matches Sorted ... | tourney_id | tourney_id |
| Matches Sorted ... | tourney_name | tourney_name |
| Matches Sorted ... | surface | surface |
| Matches Sorted ... | winner_id | winner_id |
| Matches Sorted ... | winner_name | winner_name |
| Matches Sorted ... | loser_name | loser_name |
| Players Sorted b... | Birthdate | Birthdate |

OK Cancel Help

Elle combine deux jeux de données préalablement triés en rajoutant à une table les données d'une autre table à travers une jointure.

- Se combine toujours avec la transformation « Tri » qui survient en amont
- Exemple : Je veux tous les matchs de chaque tournoi et je veux que la date de naissance du vainqueur soit ajoutée à la sortie

Jointure de fusion



Elle combine deux jeux de données préalablement triés en rajoutant à une table les données d'une autre table à travers une jointure.

- Se combine toujours avec la transformation « Tri » qui survient en amont
- Exemple : Je veux tous les matchs de chaque tournoi et je veux que la date de naissance du vainqueur soit ajoutée à la sortie

Jointure de fusion

| Viewer at Merge Join | | | | |
|----------------------|-----------|------------------|-----------------------|-----------|
| Copy Data | | | | |
| surface | winner_id | winner_name | loser_name | Birthdate |
| Clay | 100644 | Alexander Zverev | Fernando Verdasco | 19970420 |
| Grass | 100644 | Alexander Zverev | Philipp Kohlschreiber | 19970420 |
| Clay | 100644 | Alexander Zverev | Marin Cilic | 19970420 |
| Hard | 100644 | Alexander Zverev | Kevin Anderson | 19970420 |
| Grass | 100644 | Alexander Zverev | Roberto Bautista Agut | 19970420 |
| Grass | 100644 | Alexander Zverev | Richard Gasquet | 19970420 |
| Hard | 100644 | Alexander Zverev | Kei Nishikori | 19970420 |
| Grass | 100644 | Alexander Zverev | Evgeny Donskoy | 19970420 |
| Hard | 100644 | Alexander Zverev | Kevin Anderson | 19970420 |
| Hard | 100644 | Alexander Zverev | John Isner | 19970420 |
| Hard | 100644 | Alexander Zverev | Stanislas Wawrinka | 19970420 |
| Grass | 100644 | Alexander Zverev | Adrian Mannarino | 19970420 |
| Clay | 100644 | Alexander Zverev | Novak Djokovic | 19970420 |
| Hard | 100644 | Alexander Zverev | Yen Hsun Lu | 19970420 |
| Clay | 100644 | Alexander Zverev | Tomas Berdych | 19970420 |
| Clay | 100644 | Alexander Zverev | Feliciano Lopez | 19970420 |
| Hard | 100644 | Alexander Zverev | Daniil Medvedev | 19970420 |
| Hard | 100644 | Alexander Zverev | Denis Shapovalov | 19970420 |
| .. | 100644 | Alexander Zverev | Arthur De Greef | 19970420 |
| Hard | 100644 | Alexander Zverev | Roger Federer | 19970420 |
| Hard | 100644 | Alexander Zverev | Richard Gasquet | 19970420 |
| Grass | 100644 | Alexander Zverev | Paolo Lorenzi | 19970420 |
| Hard | 100644 | Alexander Zverev | Jordan Thompson | 19970420 |
| Clay | 100644 | Alexander Zverev | Guido Pella | 19970420 |
| Clay | 100644 | Alexander Zverev | Nicolas Almagro | 19970420 |
| Clay | 100644 | Alexander Zverev | Roberto Bautista Agut | 19970420 |
| Clay | 100644 | Alexander Zverev | Jan Lennard Struff | 19970420 |
| Clay | 100644 | Alexander Zverev | Jeremy Chardy | 19970420 |
| Clay | 100644 | Alexander Zverev | Andreas Seppi | 19970420 |
| Hard | 100644 | Alexander Zverev | Tennys Sandgren | 19970420 |

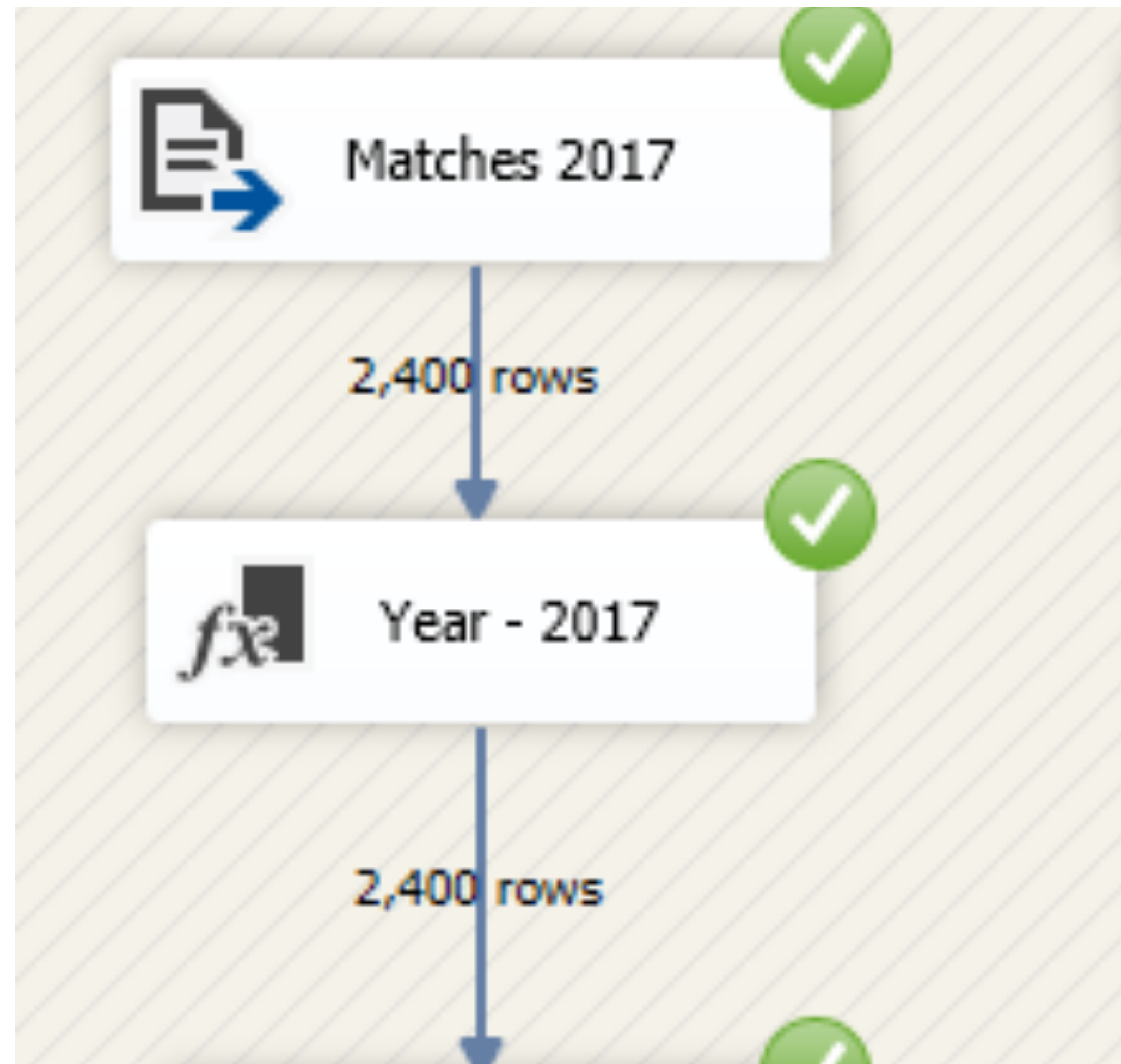
Elle combine deux jeux de données préalablement triés en rajoutant à une table les données d'une autre table à travers une jointure.

- Se combine toujours avec la transformation « Tri » qui survient en amont
- Exemple : Je veux tous les matchs de chaque tournoi et je veux que la date de naissance du vainqueur soit ajoutée à la sortie

Comparatif

| Critères | Union Tout (Union all) | Fusionner (Merge) | Jointure de fusion (Merge Join) |
|----------------------------|---------------------------|----------------------|------------------------------------|
| Nombre de flux d'entrée | ≥ 1 | 2 | 2 |
| Trier | Non | Oui | Oui |
| Type de jointure | Verticale | Verticale | Full, left, inner |

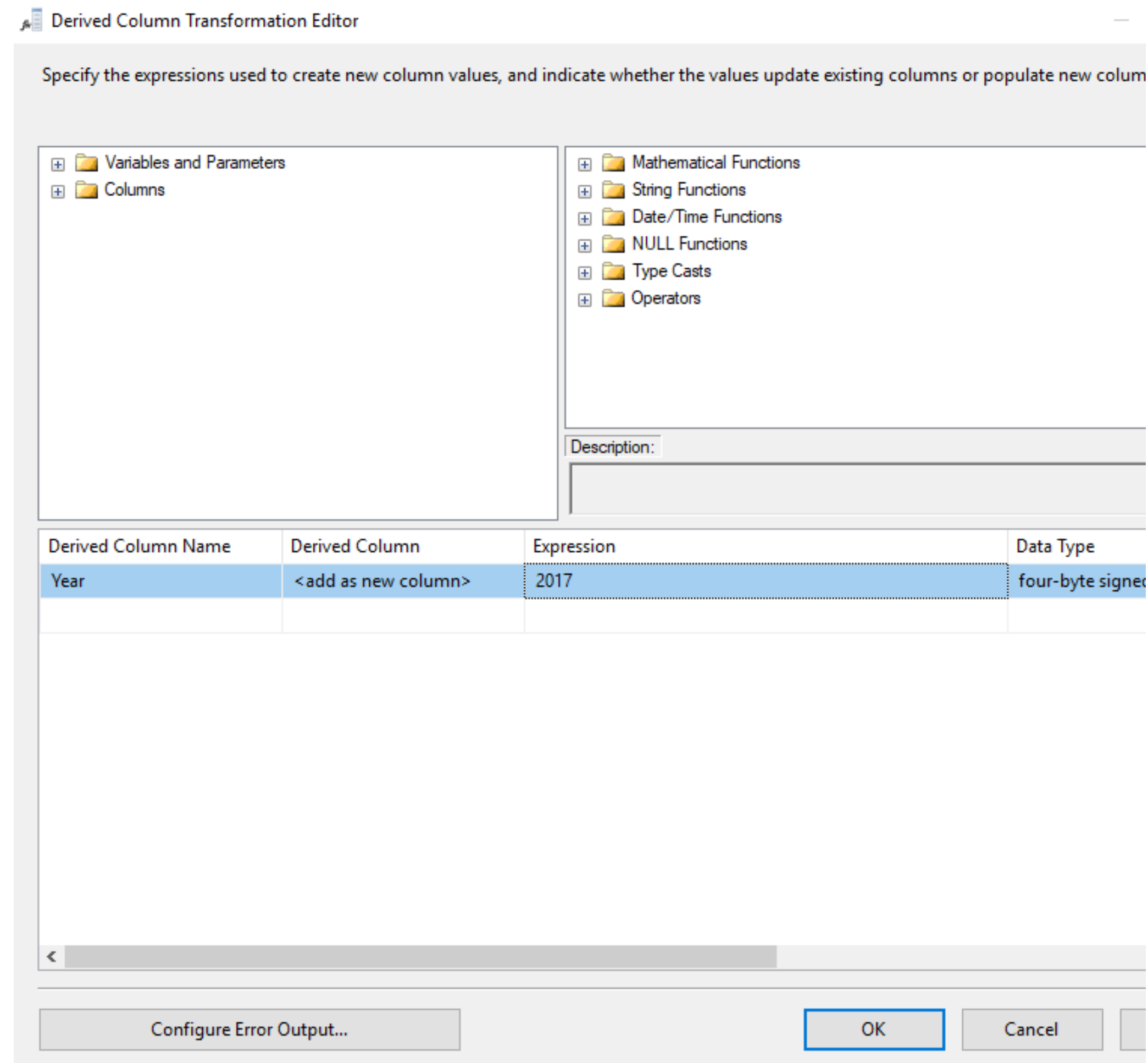
Colonne dérivée



Elle crée de nouvelles valeurs en appliquant des expressions (n'importe quelle combinaison de variables, fonctions, opérateur) aux colonnes d'entrée.

- Le résultat peut être ajouté en tant que nouvelle colonne ou remplacer une colonne existante
- Exemples: concaténer des données de différentes colonnes, appliquer des fonctions mathématiques aux données numériques, extraire des parties d'une valeur datetime...

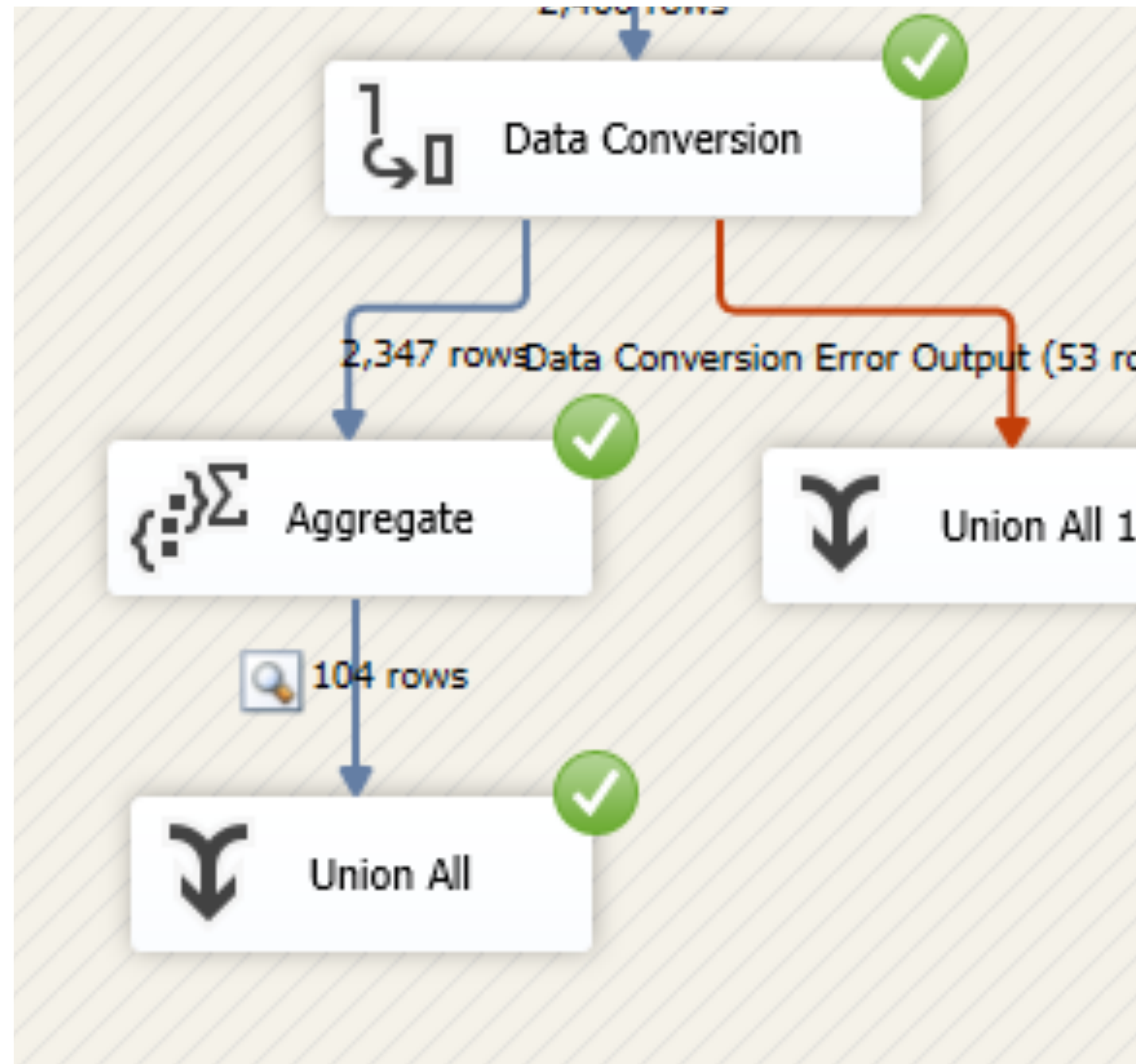
Colonne dérivée



Elle crée de nouvelles valeurs en appliquant des expressions (n'importe quelle combinaison de variables, fonctions, opérateur) aux colonnes d'entrée.

- Le résultat peut être ajouté en tant que nouvelle colonne ou remplacer une colonne existante
- Exemples: concaténer des données de différentes colonnes, appliquer des fonctions mathématiques aux données numériques, extraire des parties d'une valeur datetime...

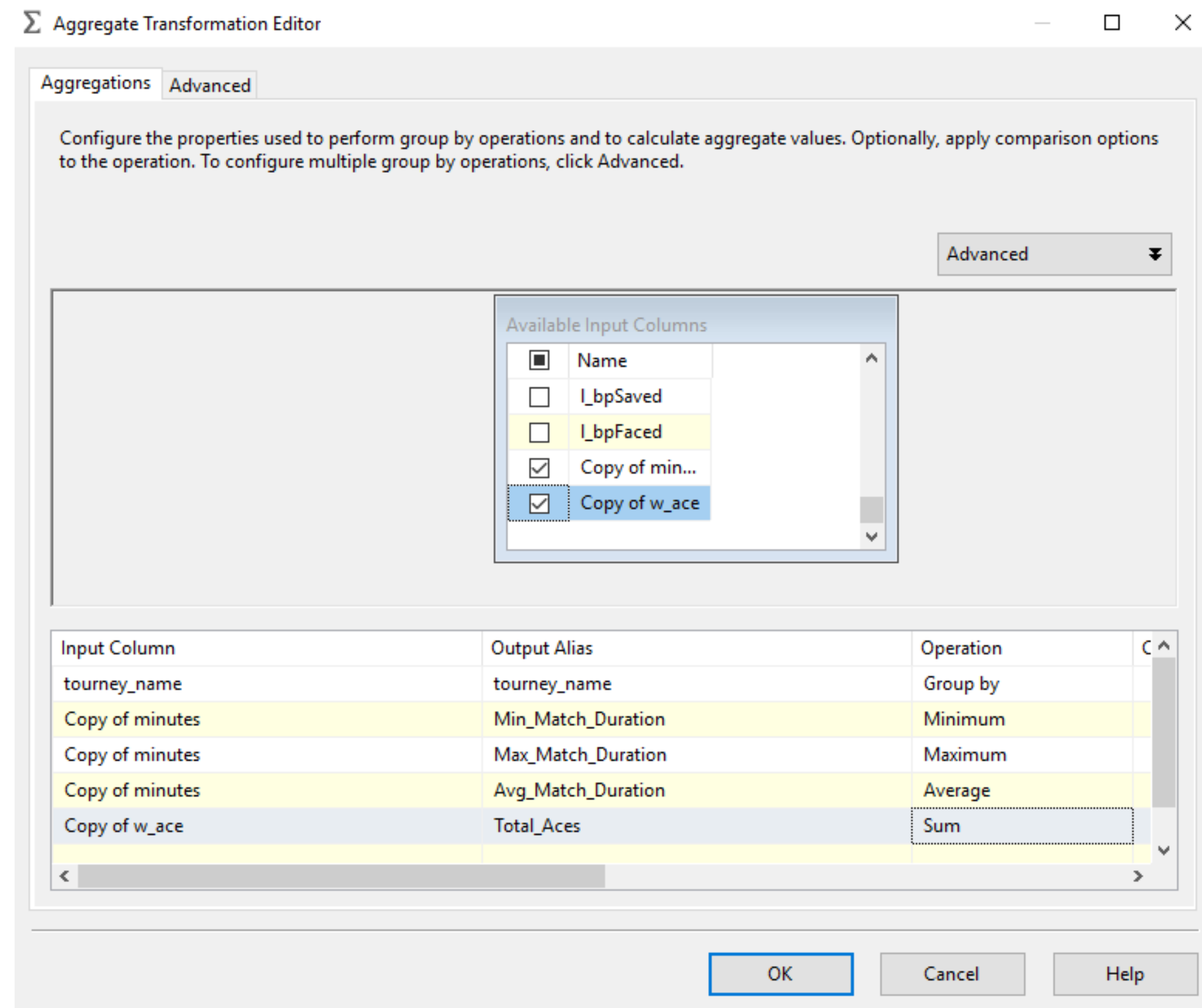
Agrégation



Elle applique des fonctions d'agrégation aux valeurs et copie les résultats dans la sortie de transformation

- Ex : somme, moyenne, nombre, nombre distinct, minimum, maximum
- Elle fournit la clause **group by** que vous pouvez utiliser pour spécifier les groupes à agréger

Agrégation



Elle applique des fonctions d'agrégation aux valeurs et copie les résultats dans la sortie de transformation

- Ex : somme, moyenne, nombre, nombre distinct, minimum, maximum
- Elle fournit la clause **group by** que vous pouvez utiliser pour spécifier les groupes à agréger

Agrégation

er at Data Flow Task

Copy Data

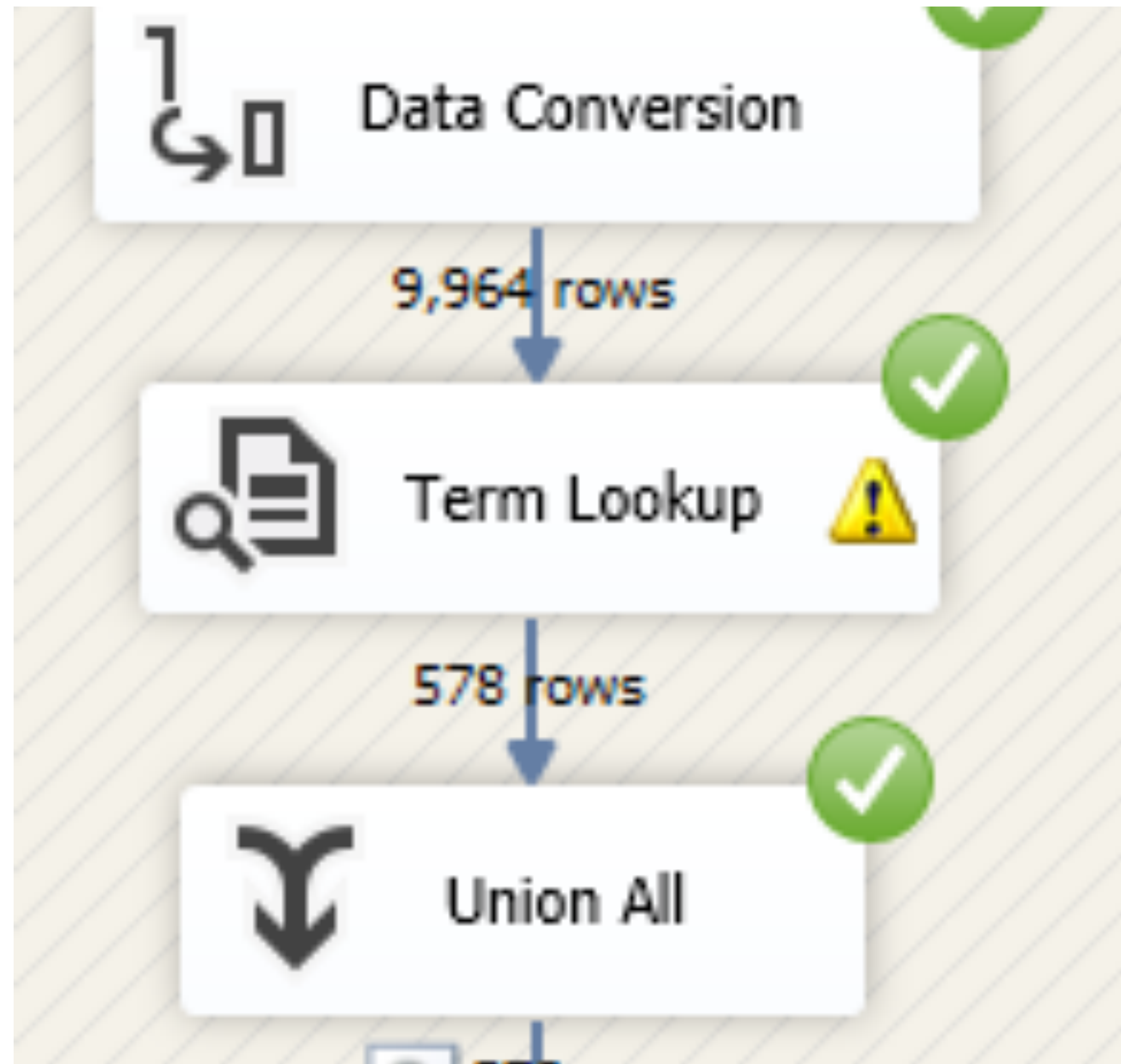
| Min_Match_Duration | Max_Match_Duration | Avg_Match_Duration | Total_Aces |
|--------------------|--------------------|--------------------|------------|
| 91 | 126 | 104.75 | 44 |
| 51 | 157 | 98.4408602150538 | 557 |
| 33 | 122 | 83.5 | 19 |
| 32 | 116 | 96 | 48 |
| 50 | 233 | 142 | 37 |
| 37 | 154 | 101.62962962963 | 162 |
| 39 | 173 | 104 | 149 |
| 115 | 134 | 124.5 | 10 |
| 57 | 71 | 64 | 4 |
| 110 | 117 | 113 | 7 |
| 39 | 174 | 86.4193548387097 | 150 |
| 57 | 150 | 112 | 17 |
| 55 | 158 | 97.1111111111111 | 193 |
| 21 | 139 | 88.6666666666667 | 207 |
| 54 | 163 | 89.3703703703704 | 289 |
| 23 | 314 | 148.629921259843 | 1590 |
| 40 | 272 | 134.25 | 42 |
| 44 | 115 | 94.25 | 35 |
| 34 | 156 | 91.8518518518518 | 314 |

uffers: 0 Rows displayed = 104

Elle applique des fonctions d'agrégation aux valeurs et copie les résultats dans la sortie de transformation

- Ex : somme, moyenne, nombre, nombre distinct, minimum, maximum
- Elle fournit la clause **group by** que vous pouvez utiliser pour spécifier les groupes à agréger

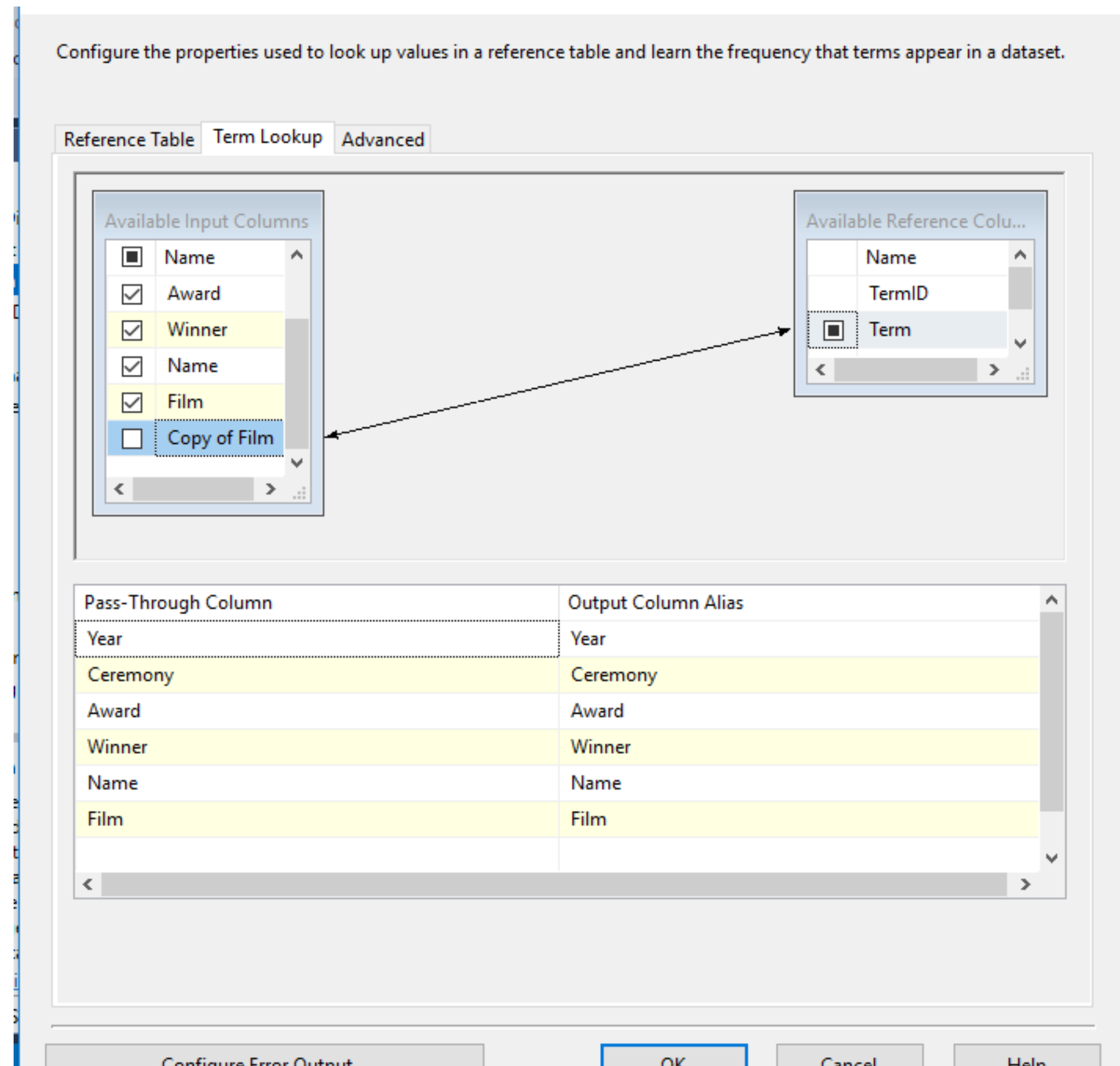
Recherche de terme



Elle recherche et compte les correspondances entre les termes de la colonne d'entrée et ceux d'une table de référence.

- le nombre est inscrit dans les colonnes de la sortie de transformation
- Peut est utile pour créer une liste de mots personnalisée basée sur le texte d'entrée, avec des statistiques de fréquence de mots

Recherche de terme



Elle recherche et compte les correspondances entre les termes de la colonne d'entrée et ceux d'une table de référence.

- le nombre est inscrit dans les colonnes de la sortie de transformation
- Peut est utile pour créer une liste de mots personnalisée basée sur le texte d'entrée, avec des statistiques de fréquence de mots

Transformation de recherche de terme

Term Lookup Transformation Editor

Configure the properties used to look up values in a reference table and learn the frequency that terms appear in a dataset.

Reference Table Term Lookup Advanced

Specify the connection manager to the reference table and the options for the index that the transformation uses.

OLE DB connection manager:
DESKTOP-483C7B2\SQLEXPRESS.Wistful

New...

Reference table name:
[dbo].[TermExtracted]

Map an input column to a reference column and select the pass-through columns on the Term Lookup page.

Configure Error Output... OK Cancel Help

Term Lookup Transformation Editor

Configure the properties used to look up values in a reference table and learn the frequency that terms appear in a dataset.

Reference Table Term Lookup Advanced

Available Input Columns

| | |
|-------------------------------------|--------------|
| <input checked="" type="checkbox"/> | Name |
| <input checked="" type="checkbox"/> | Award |
| <input checked="" type="checkbox"/> | Winner |
| <input checked="" type="checkbox"/> | Name |
| <input checked="" type="checkbox"/> | Film |
| <input type="checkbox"/> | Copy of Film |

Available Reference Columns

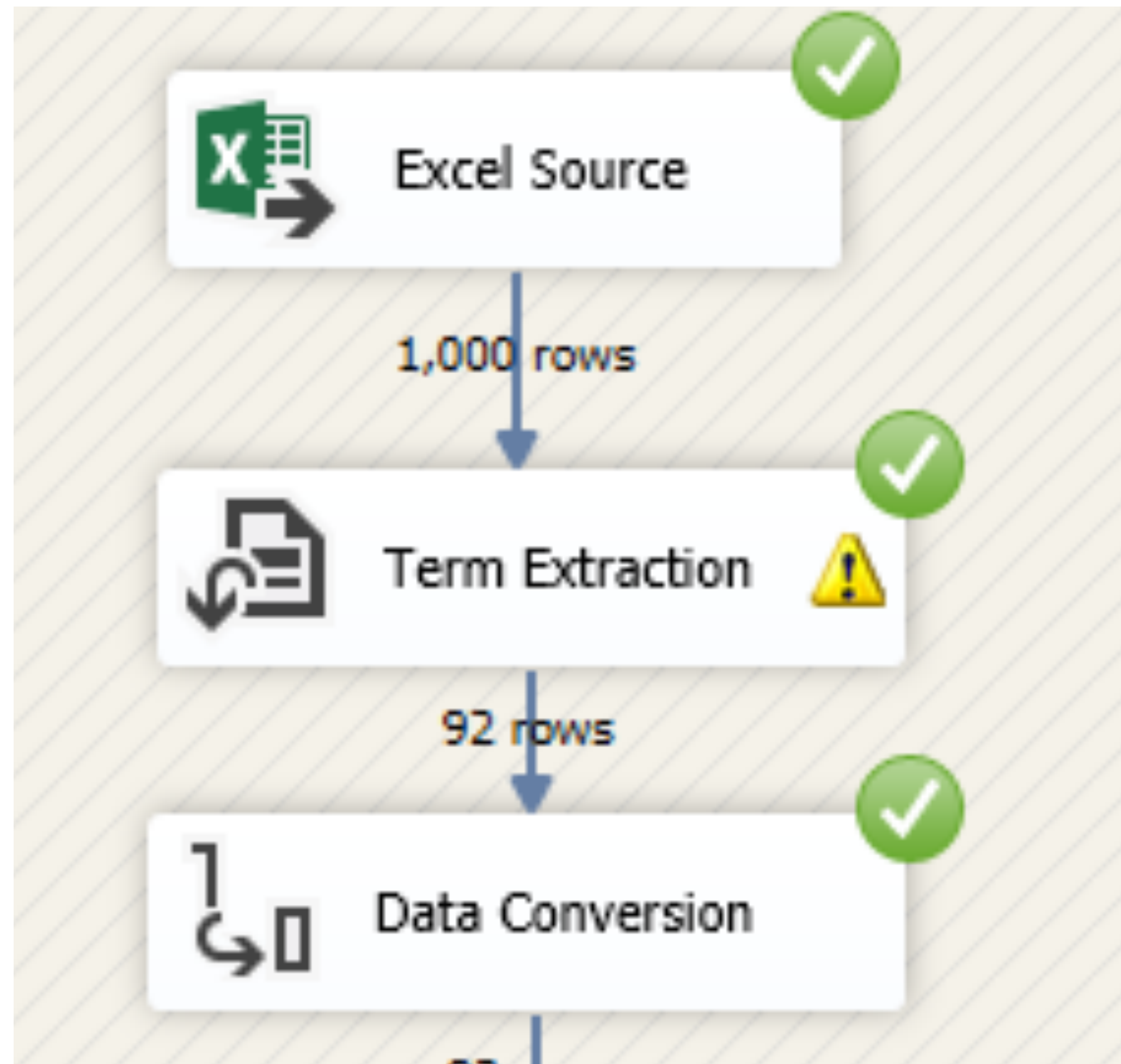
| | |
|-------------------------------------|--------|
| <input type="checkbox"/> | Name |
| <input type="checkbox"/> | TermID |
| <input checked="" type="checkbox"/> | Term |

Pass-Through Column Output Column Alias

| | |
|----------|----------|
| Year | Year |
| Ceremony | Ceremony |
| Award | Award |
| Winner | Winner |
| Name | Name |
| Film | Film |

Configure Error Output... OK Cancel Help

Extraction de termes



Elle extrait les termes du texte dans une colonne d'entrée de transformation, puis écrit les termes dans une colonne de sortie de transformation

- ne peut fonctionner qu'avec du texte dans une colonne qui a le type de données DT_WSTR ou DT_NTEXT
- utile pour découvrir le contenu d'un ensemble de données
- peut extraire des noms, des expressions nominales ou des noms et des expressions nominales
- génère un score pour chaque terme qu'elle extrait

Extraction de termes

The screenshot shows the 'Term Extraction Transformation Editor' window. It has a title bar with a close button. Below the title bar is a descriptive text: 'Configure the properties used to extract terms from an input column by identifying nouns and noun phrases that appear in column values. Optionally, use exclusion terms, stored in a database table, to identify terms to skip.' There are three tabs: 'Term Extraction' (selected), 'Exclusion', and 'Advanced'. The main area is titled 'Specify one text column as input.' and contains a large empty box. Overlaid on this box is a 'Available Input Columns' dialog. This dialog has a table with the following content:

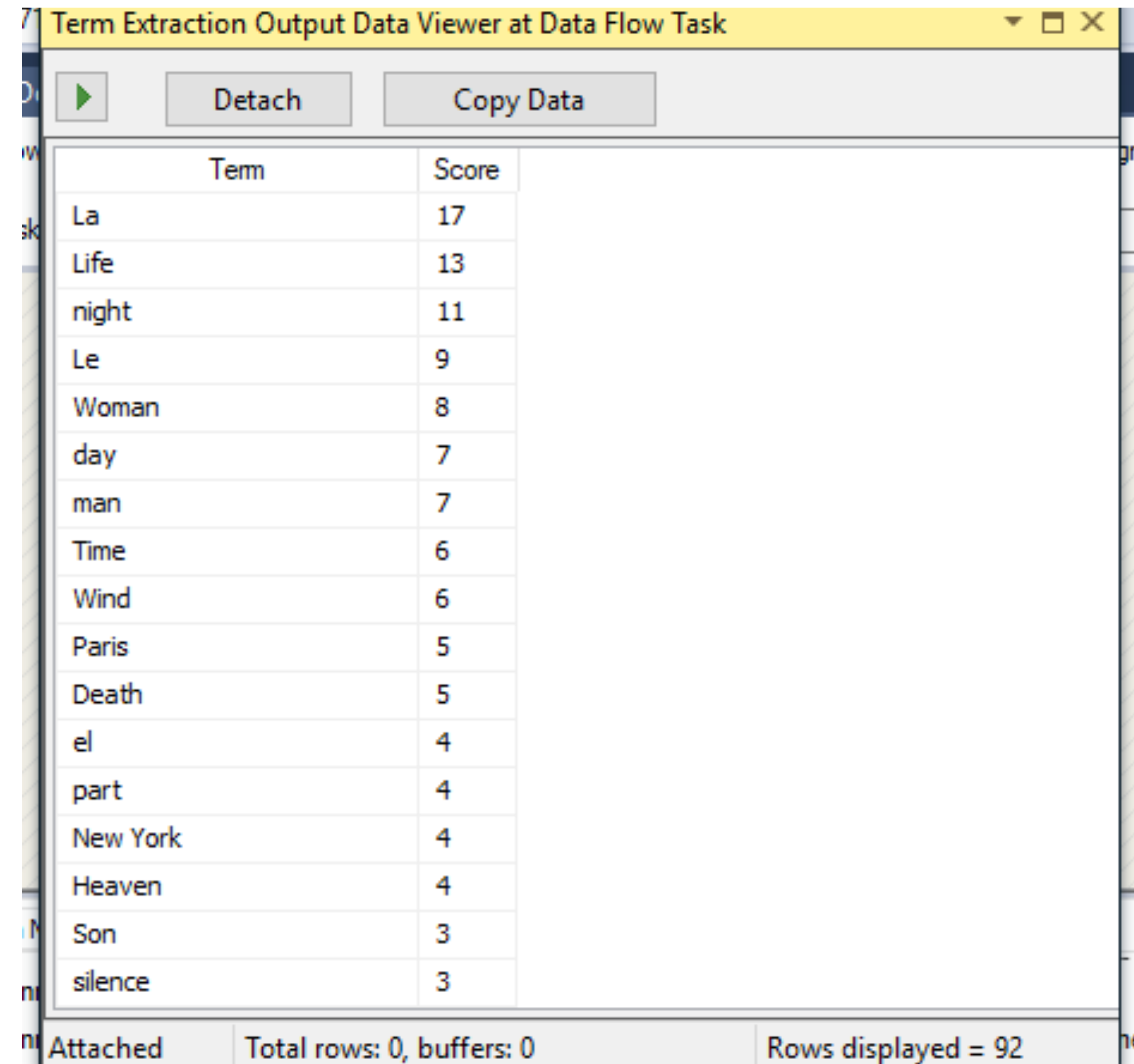
| | Name |
|-------------------------------------|----------|
| <input type="checkbox"/> | Pos |
| <input type="checkbox"/> | 2015 |
| <input checked="" type="checkbox"/> | Title |
| <input type="checkbox"/> | Director |
| <input type="checkbox"/> | Year |
| <input type="checkbox"/> | Country |

Below the table are navigation arrows. At the bottom of the main window, there is a section for 'Output column name' with two input fields: 'Term:' with the value 'Term' and 'Score:' with the value 'Score'. At the very bottom are four buttons: 'Configure Error Output...', 'OK', 'Cancel', and 'Help'.

Elle extrait les termes du texte dans une colonne d'entrée de transformation, puis écrit les termes dans une colonne de sortie de transformation

- ne peut fonctionner qu'avec du texte dans une colonne qui a le type de données DT_WSTR ou DT_NTEXT
- utile pour découvrir le contenu d'un ensemble de données
- peut extraire des noms, des expressions nominales ou des noms et des expressions nominales
- génère un score pour chaque terme qu'elle extrait

Extraction de termes



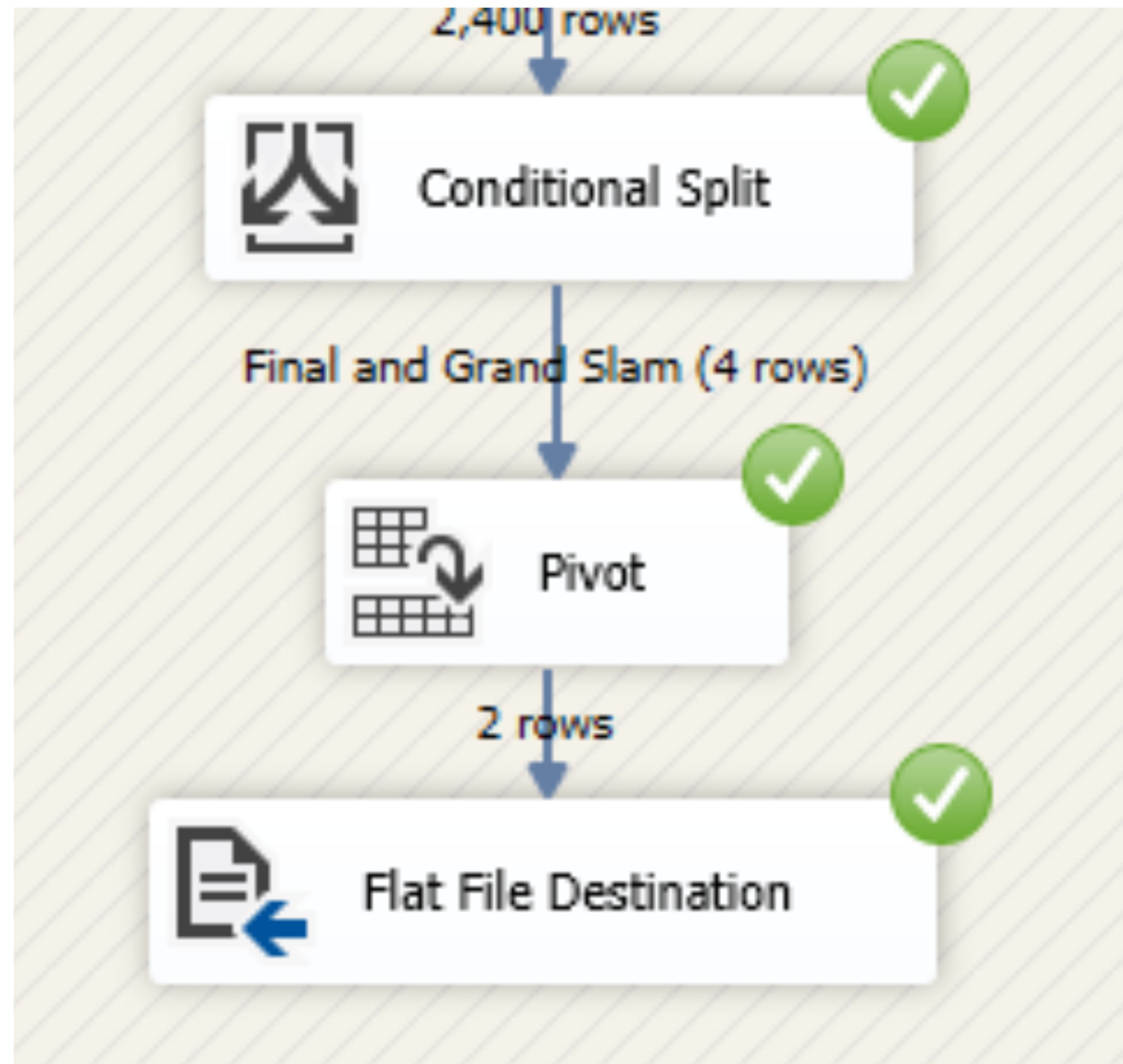
| Term | Score |
|----------|-------|
| La | 17 |
| Life | 13 |
| night | 11 |
| Le | 9 |
| Woman | 8 |
| day | 7 |
| man | 7 |
| Time | 6 |
| Wind | 6 |
| Paris | 5 |
| Death | 5 |
| el | 4 |
| part | 4 |
| New York | 4 |
| Heaven | 4 |
| Son | 3 |
| silence | 3 |

Attached Total rows: 0, buffers: 0 Rows displayed = 92

Elle extrait les termes du texte dans une colonne d'entrée de transformation, puis écrit les termes dans une colonne de sortie de transformation

- ne peut fonctionner qu'avec du texte dans une colonne qui a le type de données DT_WSTR ou DT_NTEXT
- utile pour découvrir le contenu d'un ensemble de données
- peut extraire des noms, des expressions nominales ou des noms et des expressions nominales
- génère un score pour chaque terme qu'elle extrait

Transformation de pivot



Elle transforme un ensemble de données normalisé en une version moins normalisée et plus compacte en faisant pivoter les données d'entrée sur une valeur de colonne.

Remarque : les données d'entrée doivent être triées dans la colonne pivot car un pivot se produit chaque fois que la valeur de la ligne change

Les colonnes d'entrée jouent différents rôles dans le processus de pivotement:

- La colonne est transmise inchangée à la sortie
- La colonne agit comme la clé
- La colonne définit le pivot
- La colonne contient des valeurs placées dans les colonnes créées par le pivot

Transformation de pivot

The screenshot shows the 'Pivot' configuration window. It has a title bar with a 'Pivot' icon and label. The main area is divided into sections for 'Pivot Key', 'Set Key', and 'Pivot Value'. The 'Pivot Key' section has a dropdown menu set to 'tournament_name'. The 'Set Key' section has a dropdown menu set to 'winner_name'. The 'Pivot Value' section has a dropdown menu set to 'score'. Below these sections, there is a checkbox labeled 'Ignore un-matched Pivot Key values and report them after DataFlow execution'. At the bottom, there is a section titled 'Generate pivot output columns from values:' with a hint: 'Hint: choose to 'Ignore' un-matched Pivot Key values, execute this DataFlow in the debugger and copy the value list reported in the debugger's Output Window'. A text box below this contains the list '[Australian Open],[Roland Garros],[Us Open],[Wimbledon]'. To the right of this text box is a section titled 'Existing pivoted out'. At the bottom of the window are three buttons: 'Generate Columns Now', 'OK', and 'Cancel'.

Pivot

Pivot Key:
Values in the input data from this column will become new column names in the output

Set Key:
Identifies a group of input rows that will get pivoted into one output row. The input data must be sorted on this column

Pivot Value:
Values from this column will be mapped into the pivot output columns

☐ Ignore un-matched Pivot Key values and report them after DataFlow execution

Generate pivot output columns from values:
Hint: choose to 'Ignore' un-matched Pivot Key values, execute this DataFlow in the debugger and copy the value list reported in the debugger's Output Window

[Australian Open],[Roland Garros],[Us Open],[Wimbledon]

Generate Columns Now

OK Cancel

Elle transforme un ensemble de données normalisé en une version moins normalisée et plus compacte en faisant pivoter les données d'entrée sur une valeur de colonne.

Remarque : les données d'entrée doivent être triées dans la colonne pivot car un pivot se produit chaque fois que la valeur de la ligne change

Les colonnes d'entrée jouent différents rôles dans le processus de pivotement:

- La colonne est transmise inchangée à la sortie
- La colonne agit comme la clé
- La colonne définit le pivot
- La colonne contient des valeurs placées dans les colonnes créées par le pivot

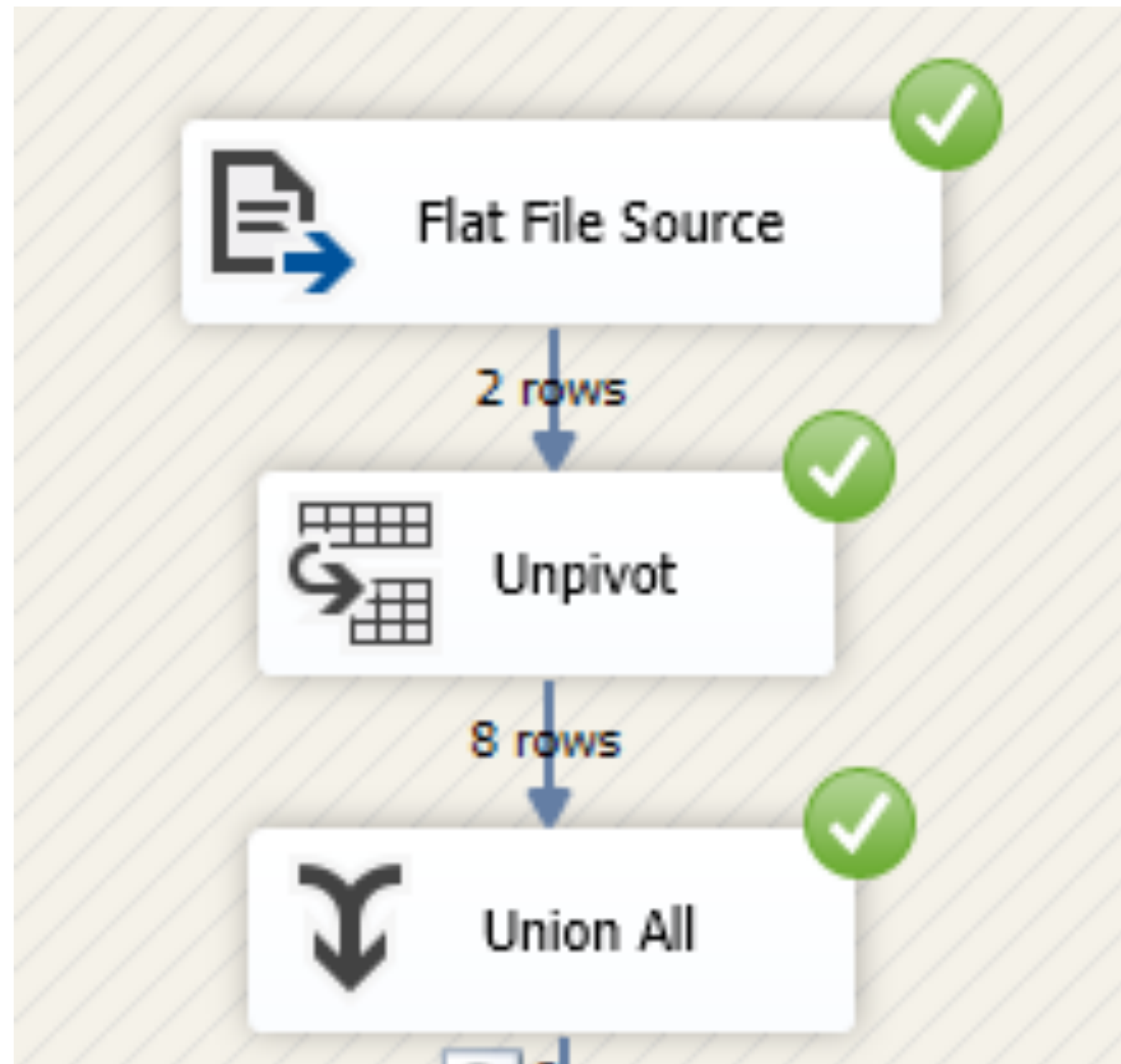
Transformation de pivot - exemple

Le tableau suivant montre un ensemble de données après le pivotement des données dans la colonne Année

| | 2002 | 2003 | 2004 |
|-----------------------|-------------|-------------|------------|
| HL Mountain Tire | 141164.10 | 446297.775 | 1504884.15 |
| Road Tire Tube | 3592.05 | 35920.50 | 89801.25 |
| Water Bottle – 30 oz. | <i>NULL</i> | <i>NULL</i> | 2805.00 |
| Touring Tire | 62364.225 | 375051.60 | 1041810.00 |

- **Clé de pivot** : année
- **Définir la clé** : nom du produit
- **Valeur pivot** : total
- **Générer des colonnes de sortie pivot à partir des valeurs**: on entre [2002], [2003], [2004]

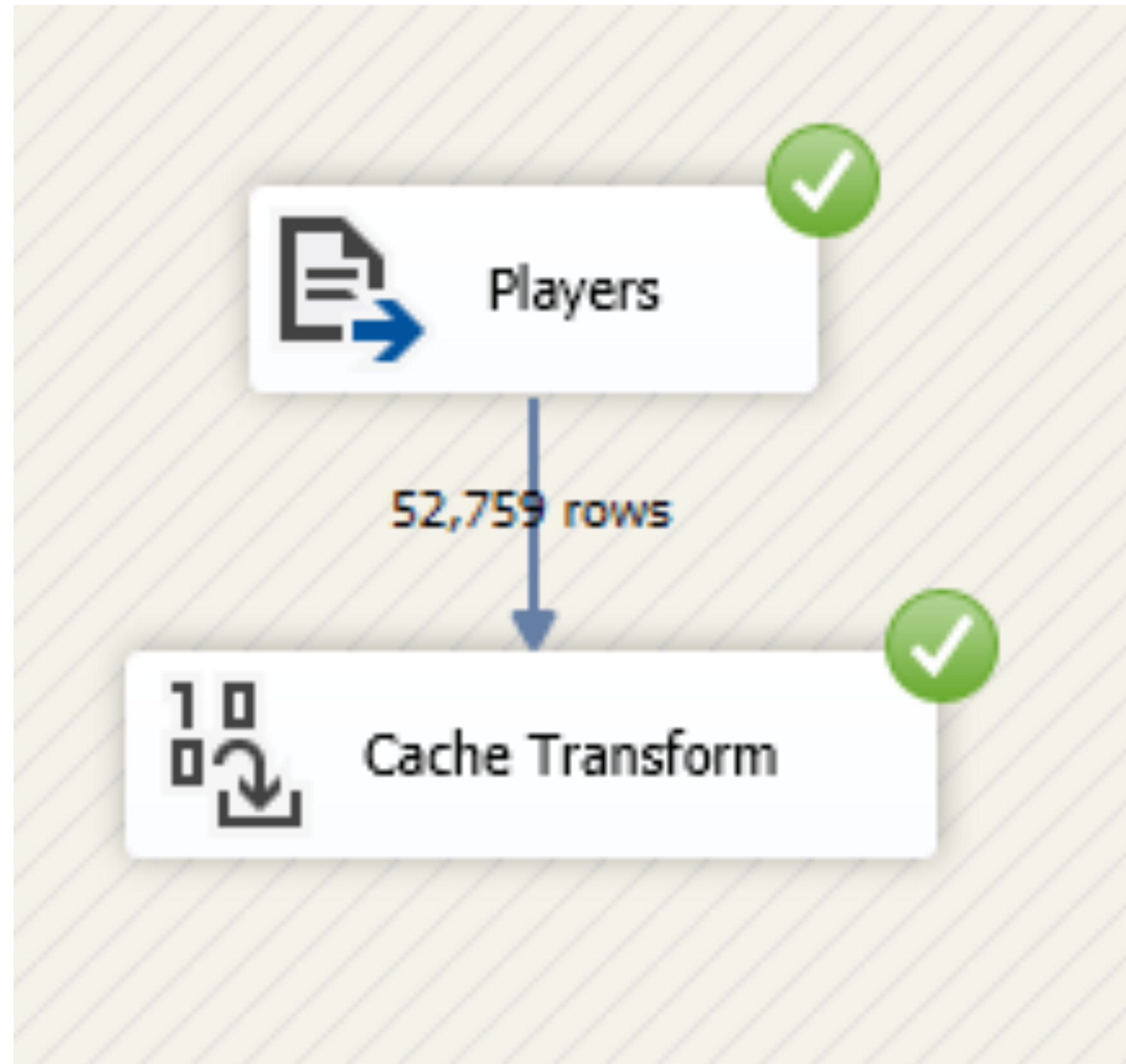
Transformation non pivotante



Elle transforme un ensemble de données non normalisé en une version plus normalisée.

- Par exemple, un ensemble de données qui répertorie les noms des clients a une ligne pour chaque client, avec les produits et la quantité achetée affichés dans les colonnes de la ligne
- Une fois que la transformation Unpivot a normalisé l'ensemble de données, l'ensemble de données contient une ligne différente pour chaque produit acheté par le client

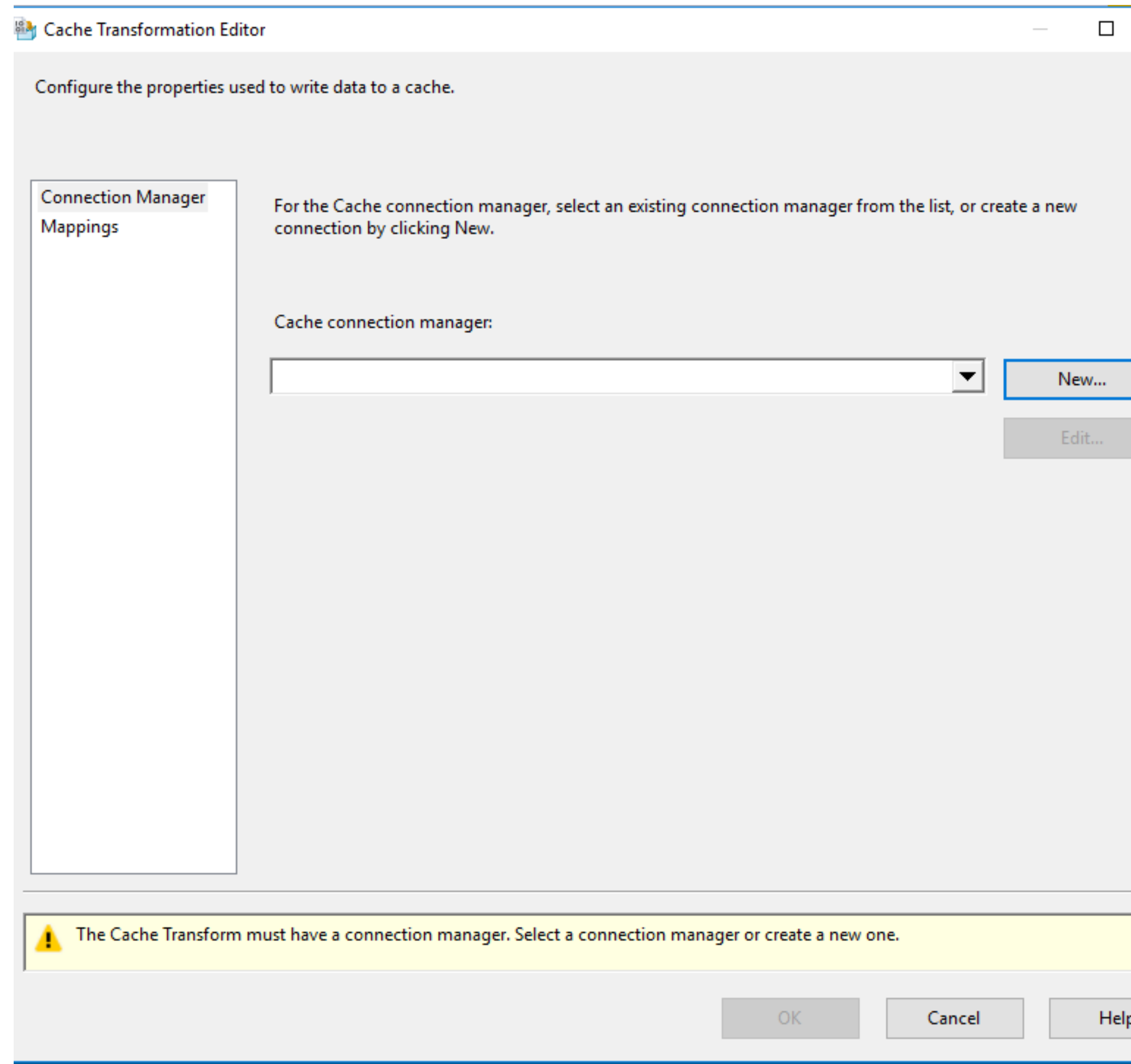
Transformation de cache



Elle génère un ensemble de données de référence pour la transformation de recherche en écrivant des données d'une source de données connectée dans le flux de données vers un gestionnaire de connexions de cache

- Un fichier avec l'extension de fichier CACHE contient des informations temporaires qu'un programme met de côté car il suppose que vous voudrez le réutiliser bientôt. Cela permet au logiciel de charger les informations plus rapidement qu'il ne le faudrait pour trouver les données d'origine
- Exemple : nous voulons obtenir la date de naissance du gagnant

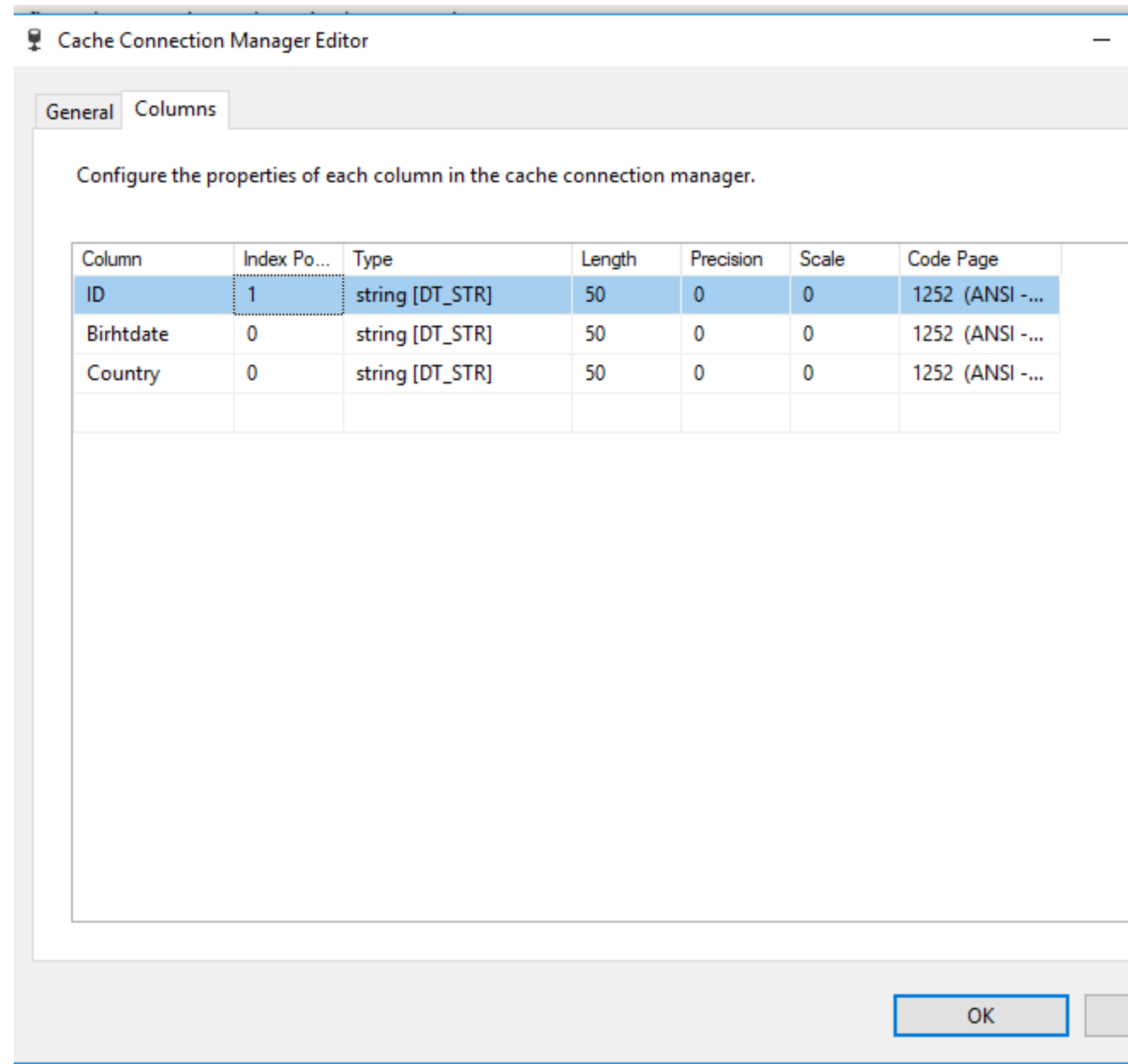
Transformation de cache



Elle génère un ensemble de données de référence pour la transformation de recherche en écrivant des données d'une source de données connectée dans le flux de données vers un gestionnaire de connexions de cache

- Un fichier avec l'extension de fichier CACHE contient des informations temporaires qu'un programme met de côté car il suppose que vous voudrez le réutiliser bientôt. Cela permet au logiciel de charger les informations plus rapidement qu'il ne le faudrait pour trouver les données d'origine
- Exemple : nous voulons obtenir la date de naissance du gagnant

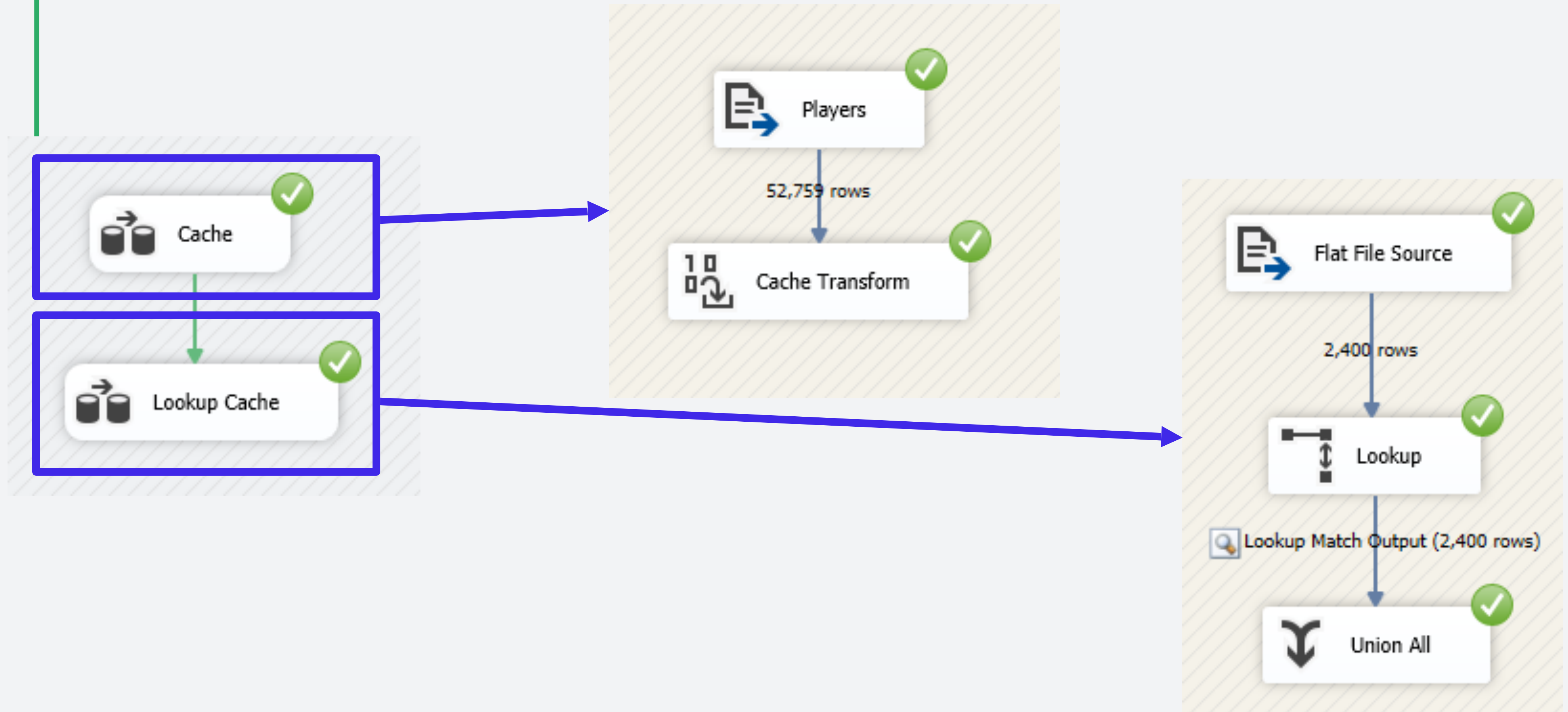
Transformation de cache



Elle génère un ensemble de données de référence pour la transformation de recherche en écrivant des données d'une source de données connectée dans le flux de données vers un gestionnaire de connexions de cache

- Un fichier avec l'extension de fichier CACHE contient des informations temporaires qu'un programme met de côté car il suppose que vous voudrez le réutiliser bientôt. Cela permet au logiciel de charger les informations plus rapidement qu'il ne le faudrait pour trouver les données d'origine
- Exemple : nous voulons obtenir la date de naissance du gagnant

Cache Transformation



Texte du titre

C. Burnay, Formation «Introduction à la Business Intelligence»

CogniTIC, Formation «SSIS»

K. Withee, (2010), «Microsoft Business Intelligence for Dummies», Wiley Publishing

<https://docs.microsoft.com/en-us/sql/integration-services/data-flow/transformations>