

SQL Server Integrated Service ETL

SQL Server Integration Services (SSIS) · 這是 Microsoft SQL Server 提供的一個企業級數據整合和轉換平台。SSIS 是一個 ETL (Extract, Transform, Load · 即擷取、轉換、載入) 工具，用於：

1. 從各種不同的數據源中提取數據
2. 對數據進行清理、轉換和整合
3. 將處理後的數據載入目標系統 (通常是數據倉庫)

有時候人們會非正式地將其稱為 "SQL Server Integrated Service ETL"，強調其 ETL 功能，但正確的官方名稱是 SQL Server Integration Services (SSIS)。

SSIS 是 SQL Server 商業智慧(BI)套件的一部分，與 SQL Server Analysis Services (SSAS) 和 SQL Server Reporting Services (SSRS) 共同組成 SQL Server 的商業智慧解決方案。

- 資料整合服務概覽
- SSIS Data Tools環境介紹
- SSIS封裝部署與執行
- SSIS中的簡易輔助工具
- SSIS的運算式
- 傳送和維護資料庫物件的相關工作
- 資料流程
- 容器交易與管理
- 分析服務整合
- SSIS的安全架構
- SSIS的進階設計

建立環境

- [Visual Studio](#)
- [SQL Server Data Tools \(SSDT\)](#)
- [SQL Server Integration Services Projects 2022](#)
- Open SSMS
 1. Create Database SSIS
 2. open download/ssis/SettingUp.sql
 3. use SSIS
 4. exec SettingUp.sql
- Setup Tools -> option
- Setup Window
- Create New Project
- 卸載\開啟專案
- Open SSIS ToolBar extensions->SSIS

- 5 tabs check
Data Flow Tab
每個 tab 都有專用的工具<如果找不到就是在錯誤的 tab

Data Flow

- drag toolbox data flow task
- scroll toolbox down to other tasks about DBA tools
- double click component
- component name is unique
- 2nd same name component will have error msg
- check data flow tab 下拉選單 have 2 component can choose
- delete 2nd component
- right click component then click edit
- SSIS is a ETL (Extract, Transform, Load · 即擷取、轉換、載入)
load database, excel or document

connection manager

- connect old version 2012, 2014 or 2016 use OLE DB connection
OLE DB Driver for sql server
rename
- switch to Data Flow Tab
toolbox -> double click Destination Assistant
right click connection manager area to add new file
- view toolbox other source
- source assistant
 1. add new source SQL Server
 2. control flow icon data go out
 3. Data Flow X A destination table name has not been provided.
 4. right click component -> edit to add table
 5. start
 6. 點中間的綠勾勾 package execution completed with ...
- connect a new table
 1. drag destination Assistant
 2. double click new item(table)
 3. select tbl dbo.Result01
 4. click right side Mappings
 5. start
 6. query from SSMS Result01
- CHECK RESULT

1. RIGHT CLICK ARROW
 2. checked enable data viewer
 3. start
 4. copy data past to excel
 5. attach/ detach
 6. checked from SSMS 影響不只一千行
- Open SSMS
 1. Create Database TAIWANJOBS
 2. import CSV file
工作 -> 匯入一般檔案
 - 匯入練習
 - 104.csv 匯入 tbl JOBS
 - 練習一
 1. 建立一個新 package/project
 2. source tbl JOB_LISTINGS
 3. destination tbl NEW_JOB_LISTINGS

```
CREATE TABLE [dbo].[NEW_JOB_LISTINGS](
    [OCCU_DESC] [nvarchar](100) NULL,
    [JOB_PERSON] [tinyint] NOT NULL,
    [JOB_DETAIL] [nvarchar](1400) NOT NULL,
    [CITYNAME] [nvarchar](50) NOT NULL,
    [WKTIME] [nvarchar](50) NULL,
    [SALARYCD] [nvarchar](50) NOT NULL,
    [EDGRDESC] [nvarchar](50) NOT NULL,
    [COMPNAME] [nvarchar](50) NOT NULL,
)
```

Transforming

- 轉換驗證
 1. open create_table.sql
 2. view Customer.CustomerID.type vs Output02.CustomerID.type is diffence
 3. RIGHT CLICK ARROW
 4. checked enable data viewer
 5. start
 6. goto data flow and right click to show advanced editor
 7. view input and output per... tab
- data conversion
 1. delete arrow
 2. add right side common : data Conversion

3. connect 3 item

4. double click data conversion item or mouse right button edit

配置用於將輸入列的資料類型轉換為不同資料類型的屬性。根據列轉換為的資料類型，設定列的長度、精確度、小數位數和代碼頁。

5. checked CustomerID (if the same name OLE CustomerID and Conversion.CustomerID)

- derived column (Customed -> output02)

1. double click derived column item

2. drag down CustomedID to down area

3. derived column name : CustomedID

4. derived column : replace CustomedID

5. (DT_STR,5,950)[CustomerID] or 1252

- 練習二

1. 清空 NEW_JOB_LISTINGS

2. 使用 data conversion 導入資料

- 練習三

1. 清空 NEW_JOB_LISTINGS

2. 將 [NEW_JOB_LISTINGS] [JOB_PERSON] 資料型態改為 varchar NOT NULL

3. 使用 derived column 導入資料

4. 建立 NUMBER_OF_PEOPLE 對應 JOB_PERSON

The screenshot shows the SQL Server Data Tools interface. On the left, a Data Flow Task is visible with the following components: TaiwanJob (OLE DB Source), Derived Column (Derived Column Transformation), and OLE DB Destination. The Derived Column transformation is selected, and the 'Derived Column Transformation Editor' dialog box is open.

The 'Derived Column Transformation Editor' dialog box has a tree view on the left with 'Variables and Parameters' and 'Columns'. The 'Columns' folder is expanded, showing a list of columns. The 'Description' field is empty.

Derived Column Name	Derived Column	Expression	Data Type	Length	Precision
NUMBER_OF_PEOPLE	<add as new column>	(DT_STR,5,950)JOB_PERSON	string [DT_STR]	5	

At the bottom of the dialog box, there are buttons for 'Configure Error Output...', 'OK', 'Cancel', and 'Help'.

- BOSS

將 JOB_LISTINGS 匯入 JOBS

聚合 **Aggregate (Customer -> output3)**

1. delete conversion and derived item
2. add right side Aggregate
3. double click Aggregate

Aggregate Transformation Editor

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Available Input Columns

- ☒ Name
- ☒ (*)
- ☐ CustomerID
- ☐ NameStyle
- ☐ Title
- ☐ FirstName
- ☐ MiddleName
- ☐ LastName
- ☐ Suffix
- ☐ CompanyN...
- ☒ SalesPerson
- ☐ EmailAddr...

Input Column	Output Alias	Operation	Com
SalesPerson	SalesPerson	Group by	
(*)	Count all	Count all	

OK Cancel Help

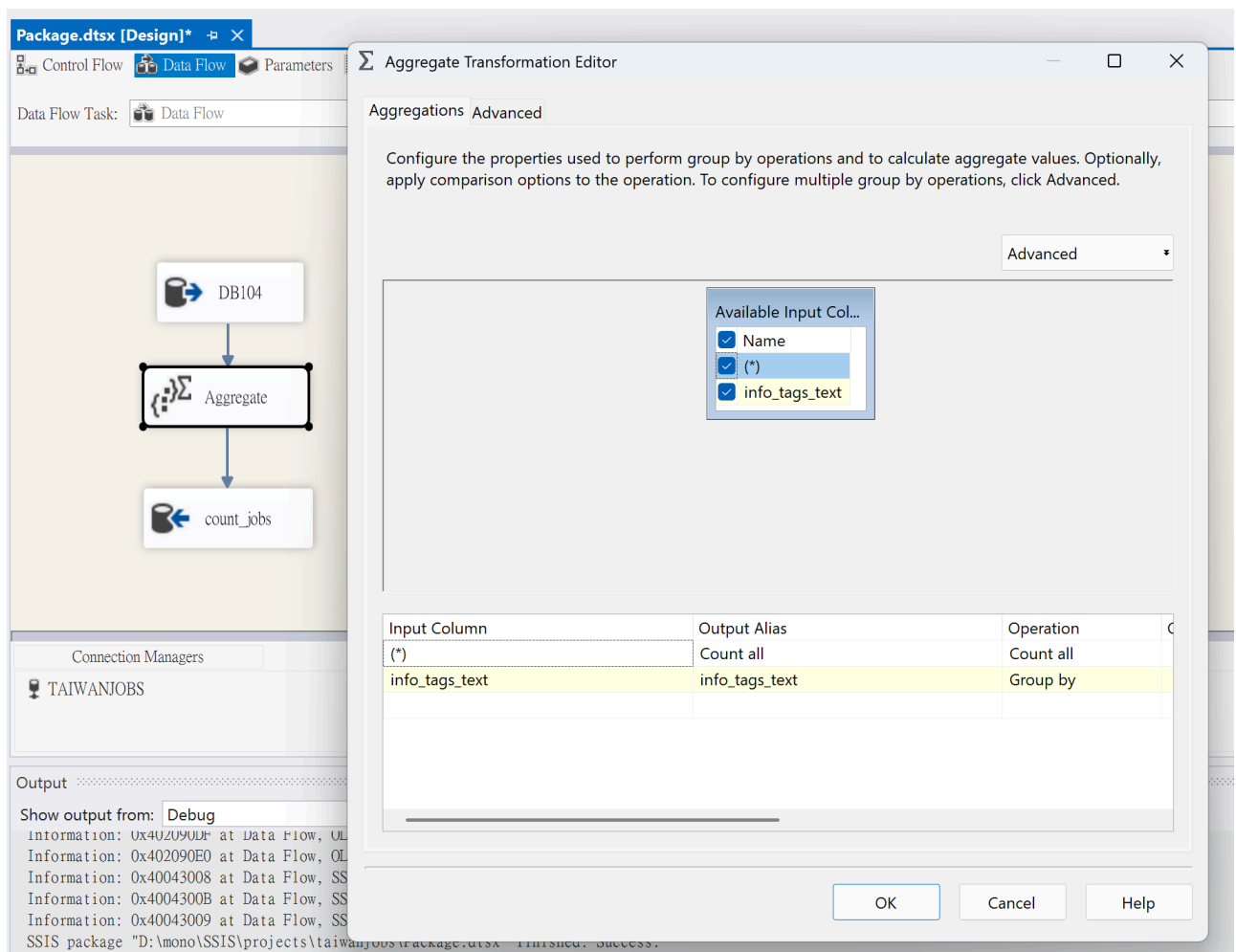
- 練習四

1. 建立新表單

```
create table [dbo].[count_jobs] (
    [region] varchar(50) NOT NULL,
    [jobs] [tinyint] NOT NULL)
```

2. 計算 104 表單各地區的總工作數

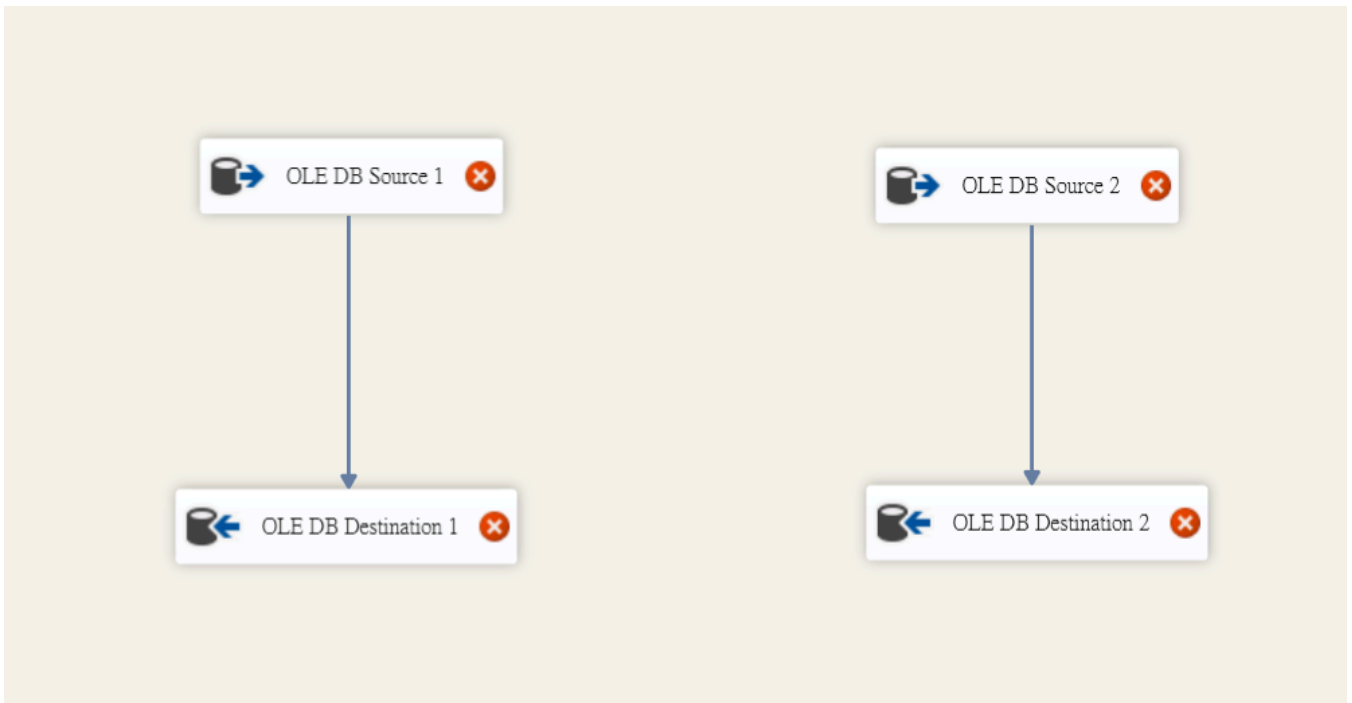
```
select info_tags_text, count(*)
from db104
group by info_tags_text
```



多播 multicast

1. source assistant Member
2. 新增 multicast tool
3. 新增兩個目的助手 output01, 02
4. 完成連接
5. 清空 output01, 02
6. start

雖然也可以這樣做，但會讀取兩次原始資料，如果只有幾百幾千筆還 ok，但百萬千萬筆就有問題了。



最重要是能節省資料讀取時間

練習五

將 104 資料表分成兩分，一分是完整複製，另一份是只保留公司名稱，所在地，職稱跟待遇

```
CREATE TABLE [dbo].[db104](
    [info_job_text] [nvarchar](100) NULL,
    [info_company_text] [nvarchar](100) NULL,
    [info_company_addon_type] [nvarchar](50) NULL,
    [info_tags_text] [nvarchar](50) NULL,
    [info_tags_text_2] [nvarchar](50) NULL,
    [info_tags_text_3] [nvarchar](50) NULL,
    [info_tags_text_4] [nvarchar](50) NULL,
    [info_description] [varchar](max) NULL,
    [text_highlight] [nvarchar](50) NULL,
    [info_othertags_text] [nvarchar](50) NULL,
    [info_othertags_text_2] [nvarchar](50) NULL,
    [info_othertags_text_4] [nvarchar](50) NULL,
    [info_othertags_text_5] [nvarchar](50) NULL,
    [last_processed_resume_at_text] [nvarchar](50) NULL,
    [date_container] [time](7) NULL,
    [actively_hiring_tag] [nvarchar](50) NULL,
    [text_highlight_6] [nvarchar](50) NULL,
    [last_cust_reply_text] [nvarchar](50) NULL
)
```

條件分割 Conditional split

1. source assistant Member
2. 新增 Conditional split
3. 新增兩個目的助手 output01, 02

4. 完成連接
5. 清空 output01, 02
6. Configure Error Output -> Ignore failure
7. 添加 data viewer 觀察執行結果

練習六

將 104 報表 **學歷不拘** 的工作獨立出一張表單

8. Configure Error Output -> Redirect Row (v.33)
9. 新增衍生欄位 derived column
10. 連接紅線
11. 執行
12. 添加 data viewer 觀察執行結果
13. 編輯 Conditional split
14. 添加 REPLACENULL(Title , "不確定") == "Mr."
15. 清空 output01, 02
16. 執行
17. ssms 觀看結果，原來的欄位 null 排除掉了

練習七

將台灣就業通 **待遇面議** 的工作獨立出一張表單，並處理錯誤

Row Sampling

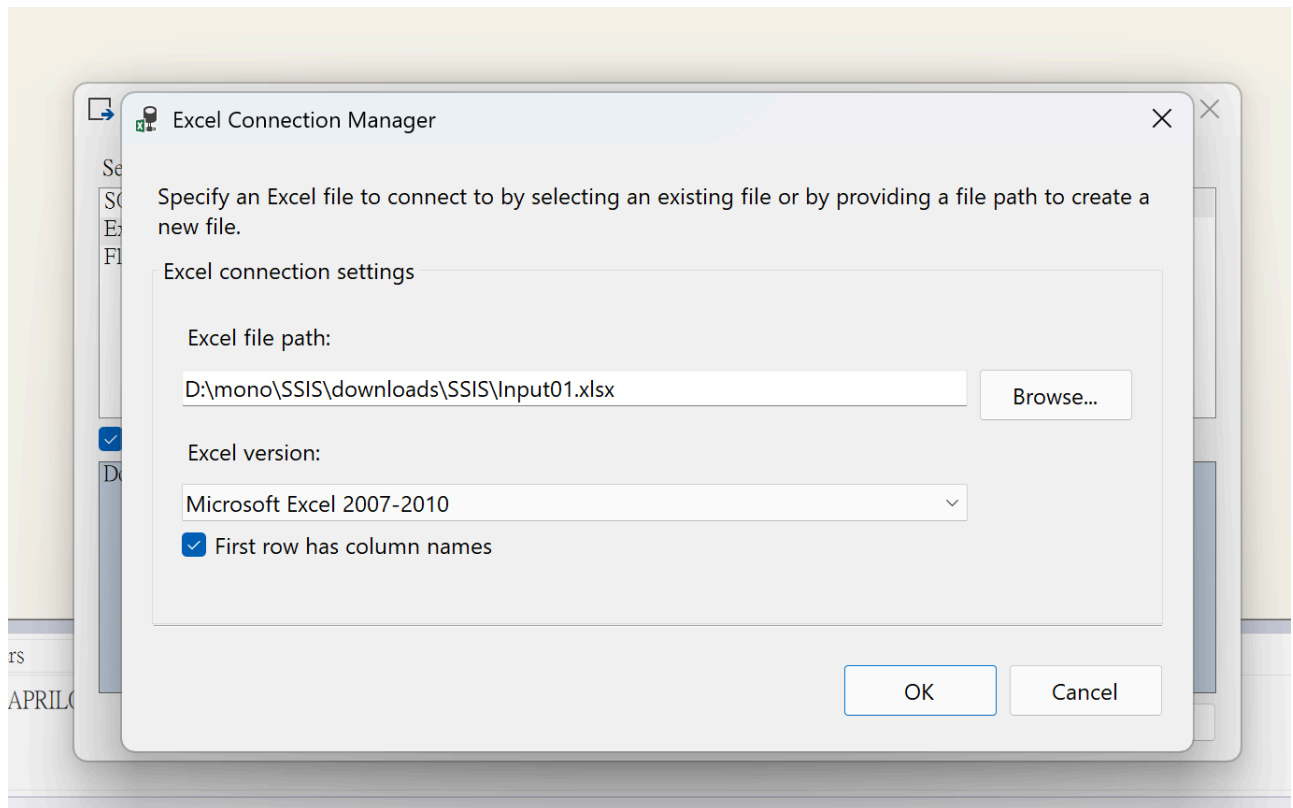
1. 刪除全部 item 只留下 customer
2. 新增 Row Sampling (other transforms) 跟兩個衍生欄位 derived column
3. 設定 Row Sampling
4. number of rows : sample 數
5. 完成連接並添加 viewr

練習八

從 104 報表隨機取出 20 份工作

連接 excel, 二維文件

1. 新增來源助手 Excel -> New



2. 設定 Excel 物件，選擇 \$sheet1 -> preview

3. 檢視 column, error output -> 確定

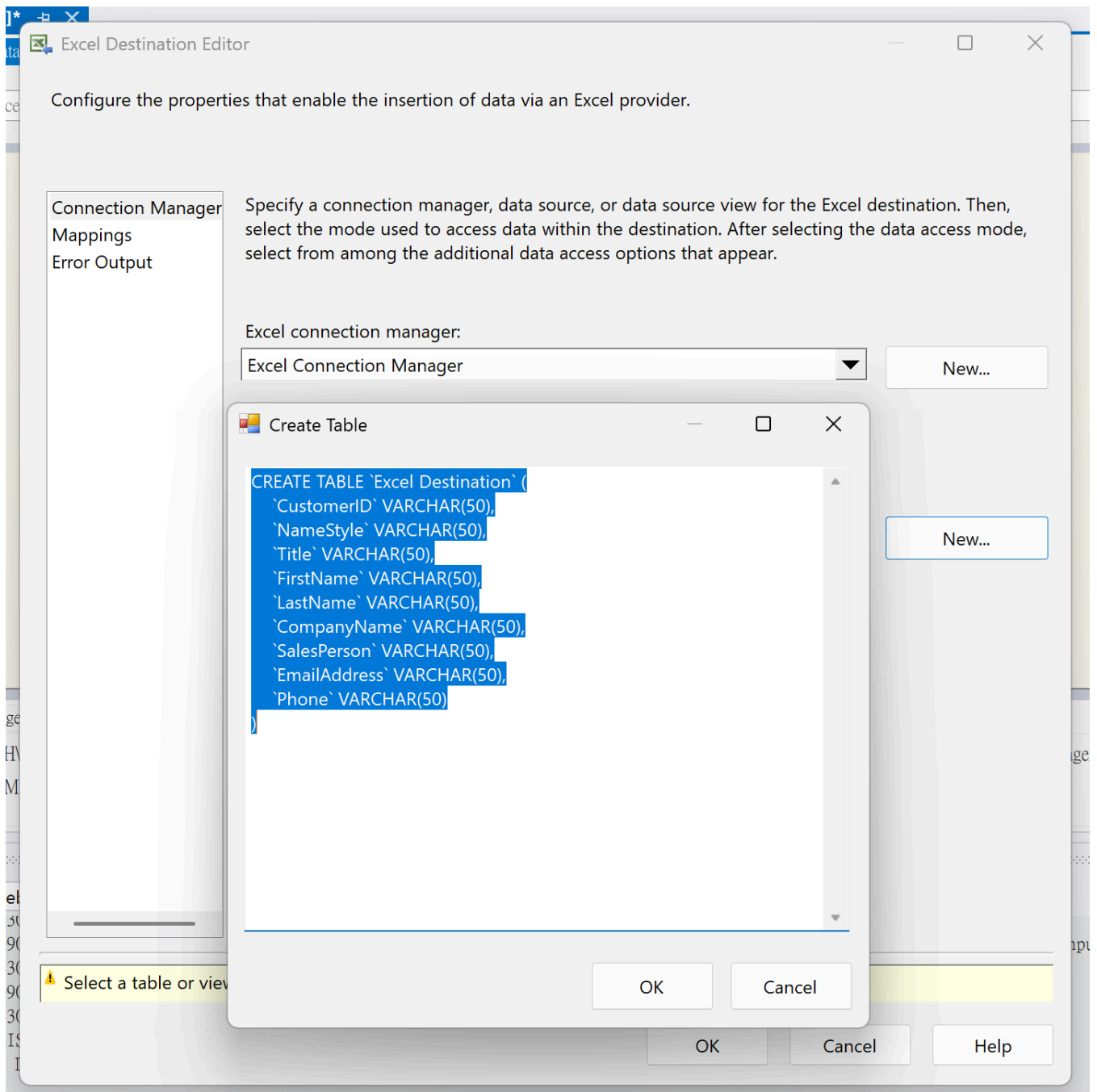
4. 新增衍生欄位 derived column 並連接

5. 加上 viewer 後執行

Derived Column

Derived Column Transformation 是 SSIS (SQL Server Integration Services) 資料流任務中的一個核心轉換元件，用於在資料流中動態建立、修改或轉換欄位值。它的主要用途是根據現有欄位、運算式或常數產生新的欄位，或修改現有欄位的值，以滿足資料處理、清洗或目標表結構的需求。

CSV 輸出到 excel



1. 新增來源 Input02.txt
2. 建立新 excel 目的
 - 附檔名使用 xlsx 才能建立成功
 - 連接，無法正確執行，資料型別錯誤

```
CREATE TABLE `JOBS2025` (
  `CustomerID` VARCHAR(50),
  `NameStyle` VARCHAR(50),
  `Title` VARCHAR(50),
  `FirstName` VARCHAR(50),
  `LastName` VARCHAR(50),
  `CompanyName` VARCHAR(50),
  `SalesPerson` VARCHAR(50),
  `EmailAddress` VARCHAR(50),
  `Phone` VARCHAR(50)
)
```

- 檢視進階設定
將外部欄位的資料型別修改為 DT-STR
- 刪除連接
- 加入轉換器並建立連接，將所有型別改為 DT-WSTR
- 編輯 excel 從新對應到轉換後的副本。CustomerID -> copy of CustomerID

百分比 Percentage Sampling

練習十

取得表 Address 中的 10% 的行。

將它們分成兩個輸出，一個用於美國地址，一個用於其他地址。

使用檢視器檢查結果。

排序 SORT

搜尋字串函式

FINDSTRING(«character_expression», «string», «occurrence»)

- 第一個參數：要被搜尋的欄位或字串
- 第二個參數：想要尋找的子字串
- 第三個參數：開始搜尋的位置（索引位置，從 1 開始計數，而非從 0 開始）

練習十

取得表 Taiwanjobs 中 **台北信義區** 薪資最高的前 10% 的工作。

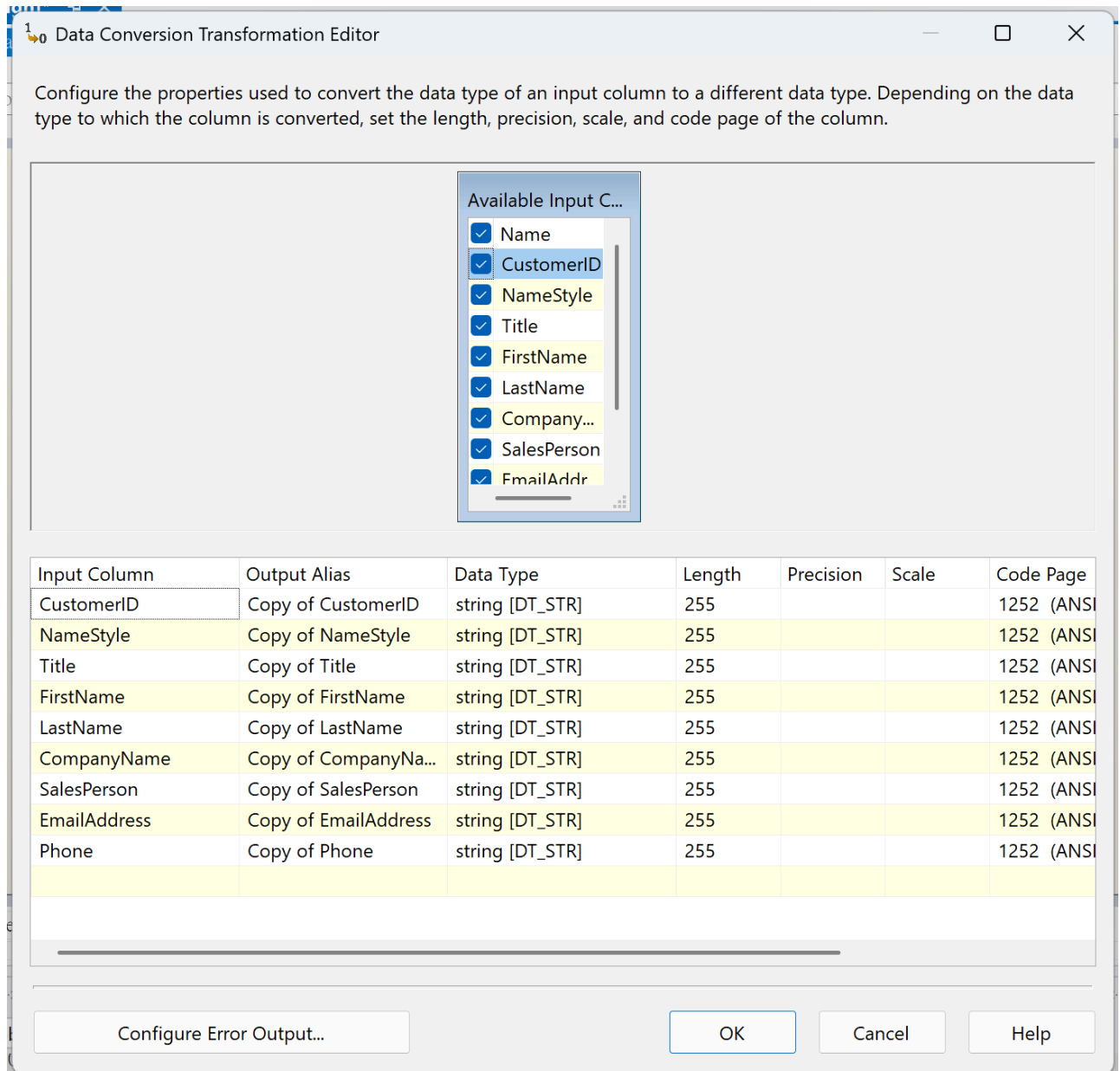
將它們分成兩個輸出，一個是台北信義區，另一個用於其他工作。

使用檢視器檢查結果。

Union All()

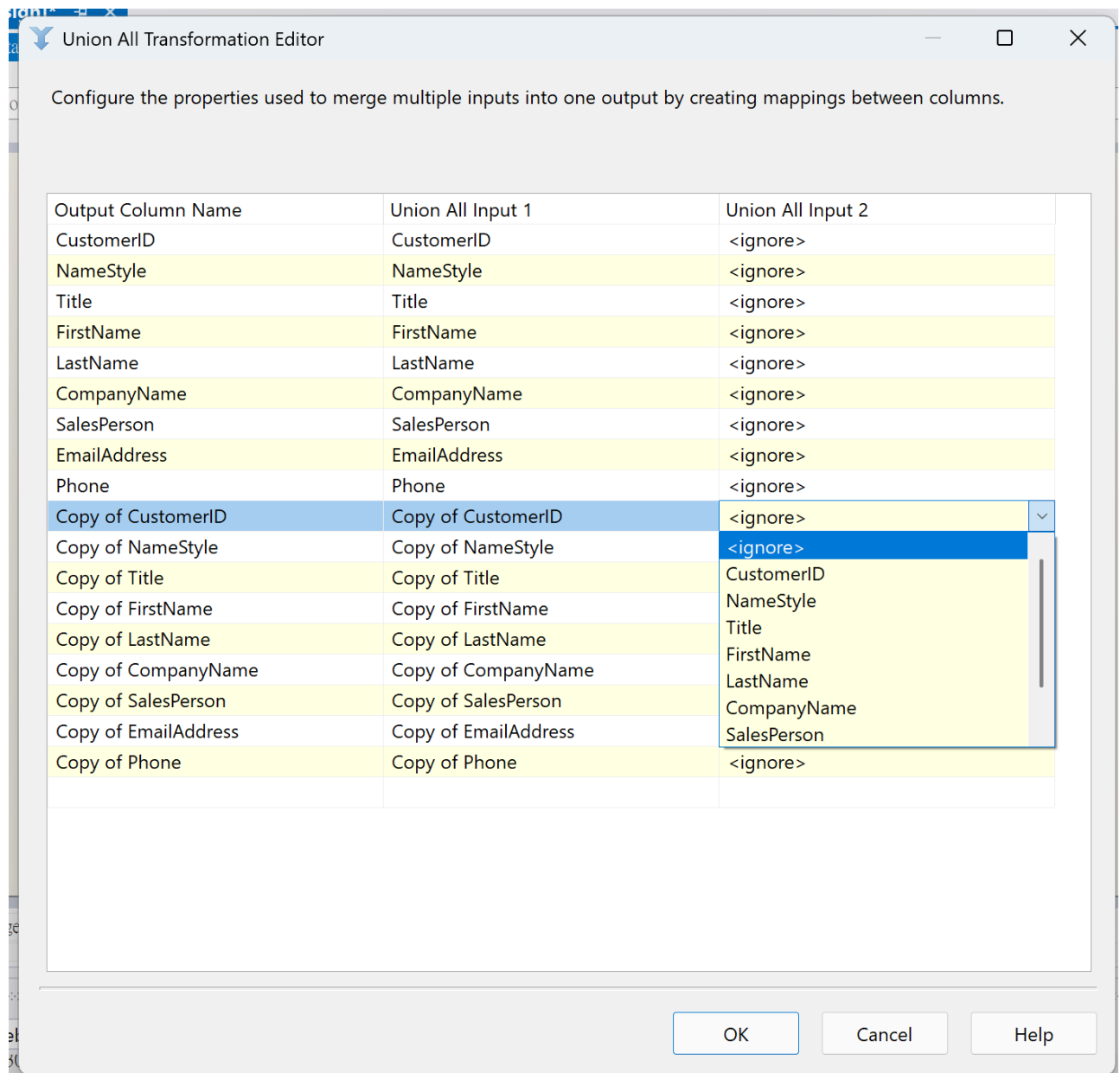
1. 建立兩個新來源 input01.xlsx(選擇\$sheet1), input02.txt
2. 建立兩個衍生欄位確認輸出結果
3. 刪除衍生欄位
4. 加入 union all 並完成連接
5. 加入衍生欄位並完成連接
6. start

7. 結果不是我要的；編輯 union all
8. union input 2 無法進行對應
9. 開啟 excel, txt 進階設定 · 檢查資料型別
10. 刪除連結
11. excel 加入資料轉換 data conversion



12. 編輯 union all 進行對應
13. 無法對應 · 因為 union all 未針對新設定完成對應

14. delete union all後重新加入並進行對應



15. start

16. 得到一個混亂結果

17. 加入排序並進行編輯

- 消除 pass through 勾選

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

Available Input Columns	Pass Through
<input checked="" type="checkbox"/> Name	<input type="checkbox"/>
<input type="checkbox"/> SalesPerson	<input type="checkbox"/>
<input type="checkbox"/> EmailAddr...	<input type="checkbox"/>
<input type="checkbox"/> Phone	<input type="checkbox"/>
<input checked="" type="checkbox"/> Copy of C...	<input checked="" type="checkbox"/>
<input type="checkbox"/> Copy of N...	<input checked="" type="checkbox"/>
<input type="checkbox"/> Copy of Ti...	<input checked="" type="checkbox"/>
<input type="checkbox"/> Copy of Fi...	<input checked="" type="checkbox"/>
<input type="checkbox"/> Copy of L...	<input checked="" type="checkbox"/>
<input type="checkbox"/> Copy of C...	<input checked="" type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Com
Copy of CustomerID	Copy of CustomerID	ascending	1	

☐ Remove rows with duplicate sort values

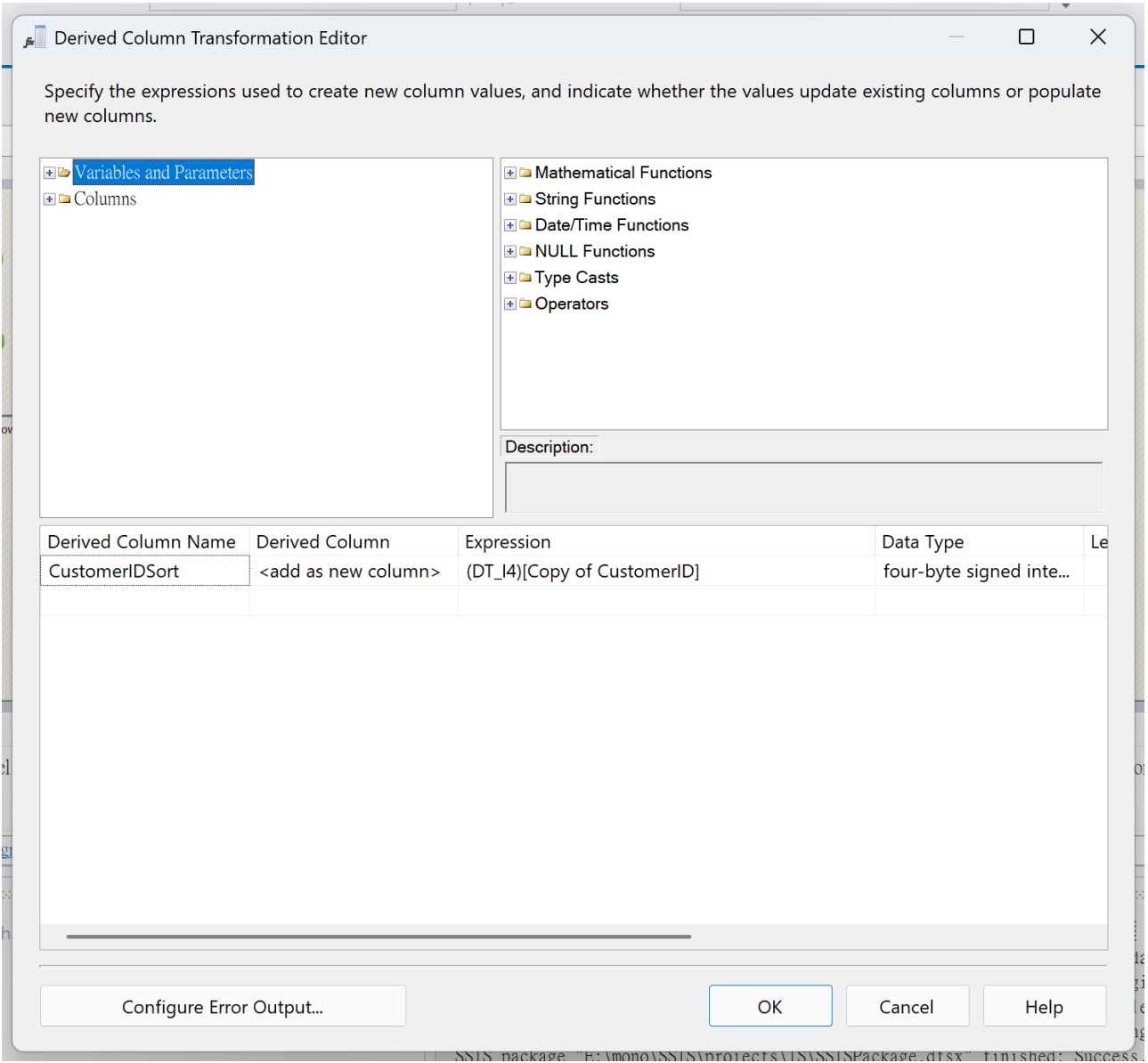
OK Cancel Help

18. start

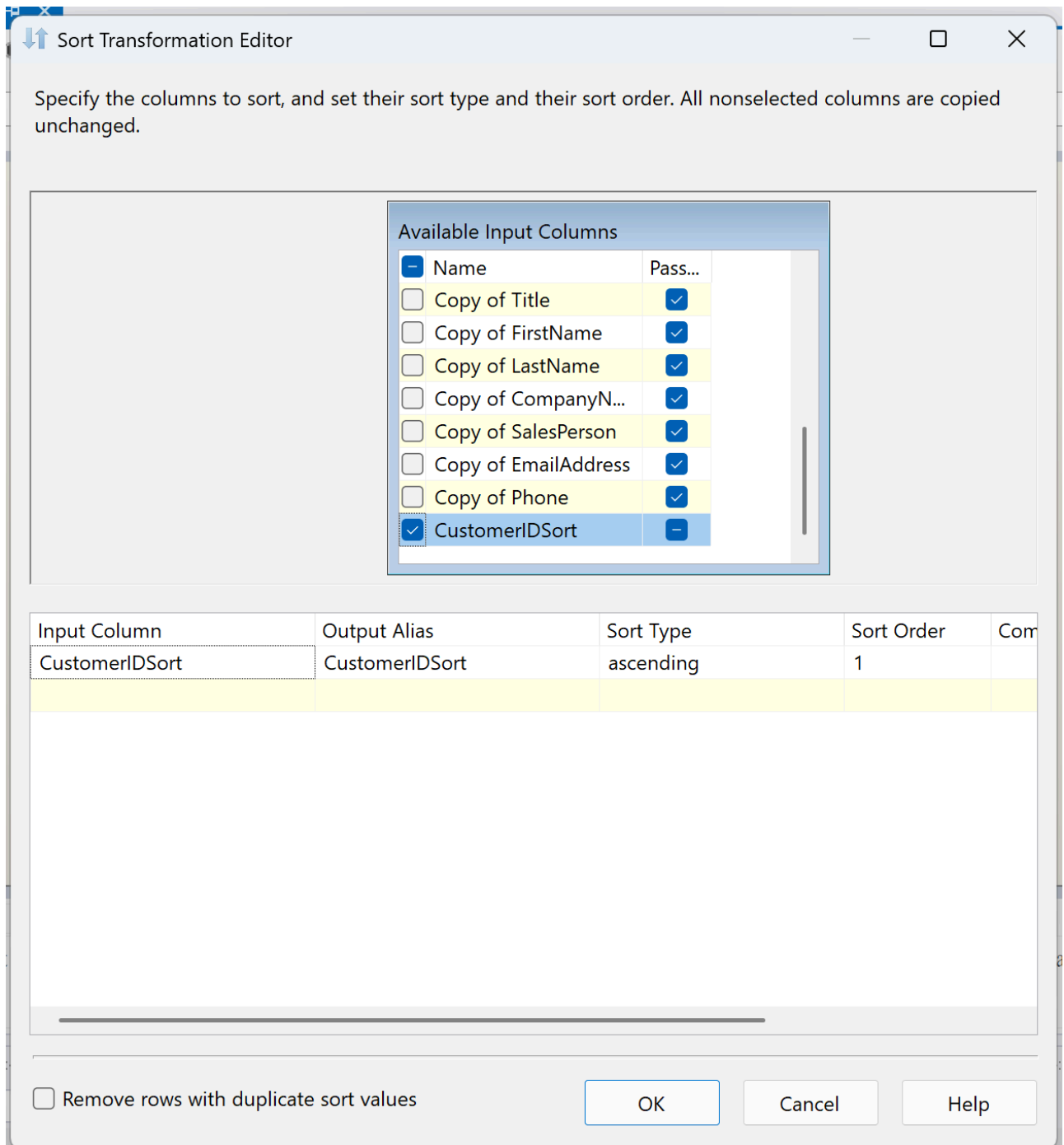
19. 得到一個排序錯誤的表單，因為 customerID 是字串型態

20. 在 union all 跟 sort 間插入一個衍生欄位進行轉換

21. 將 customerID 轉型為整數型態



22. 編輯 sort



23. Down

練習十一

將 source01.xls, source02.txt 合併

- 先將 xls 轉換成xlsx

合併 Merge (v.44)

source : Input01.xlsx, Input02.txt

1. 移除 union all 後加入 Merge 並完成連接
2. 觀察 Merge error msg 需先進行 sort
3. 加入兩個 sort 並完成連接

4. 編輯 sort2 by customerID

5. 編輯 sort1 取消 pass through 原欄位勾選

- 勾選 copy of customerID 修改別名 customerID

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...
<input type="checkbox"/>	SalesPerson	<input type="checkbox"/>
<input type="checkbox"/>	EmailAddr...	<input type="checkbox"/>
<input type="checkbox"/>	Phone	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Copy of C...	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Copy of N...	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Copy of Ti...	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Copy of Fi...	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Conv of I	<input checked="" type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Com
Copy of CustomerID	CustomerID	ascending	1	

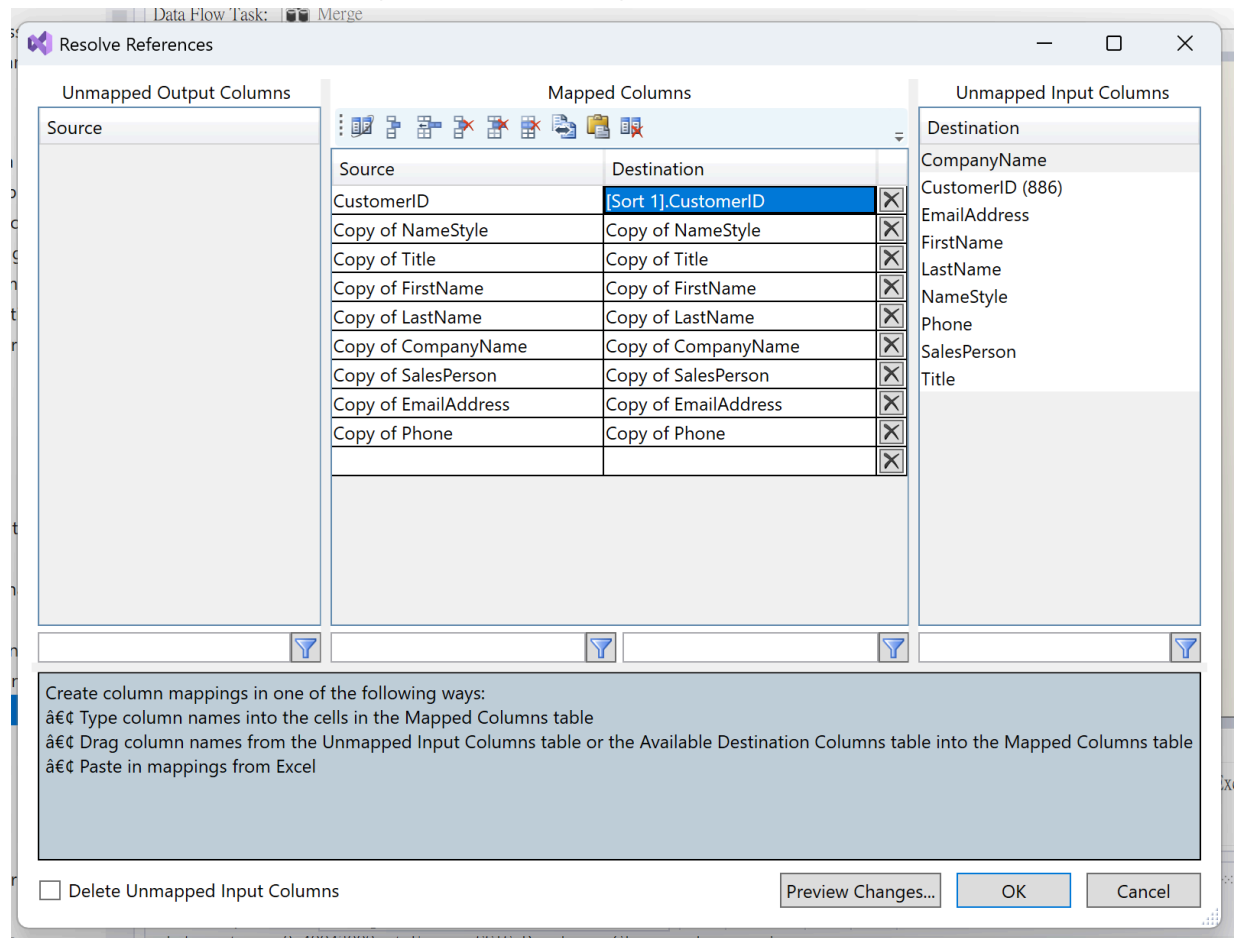
⚠ There is more than one output column with the name "CustomerID".

☐ Remove rows with duplicate sort values

OK Cancel Help

6. 處理箭頭的X

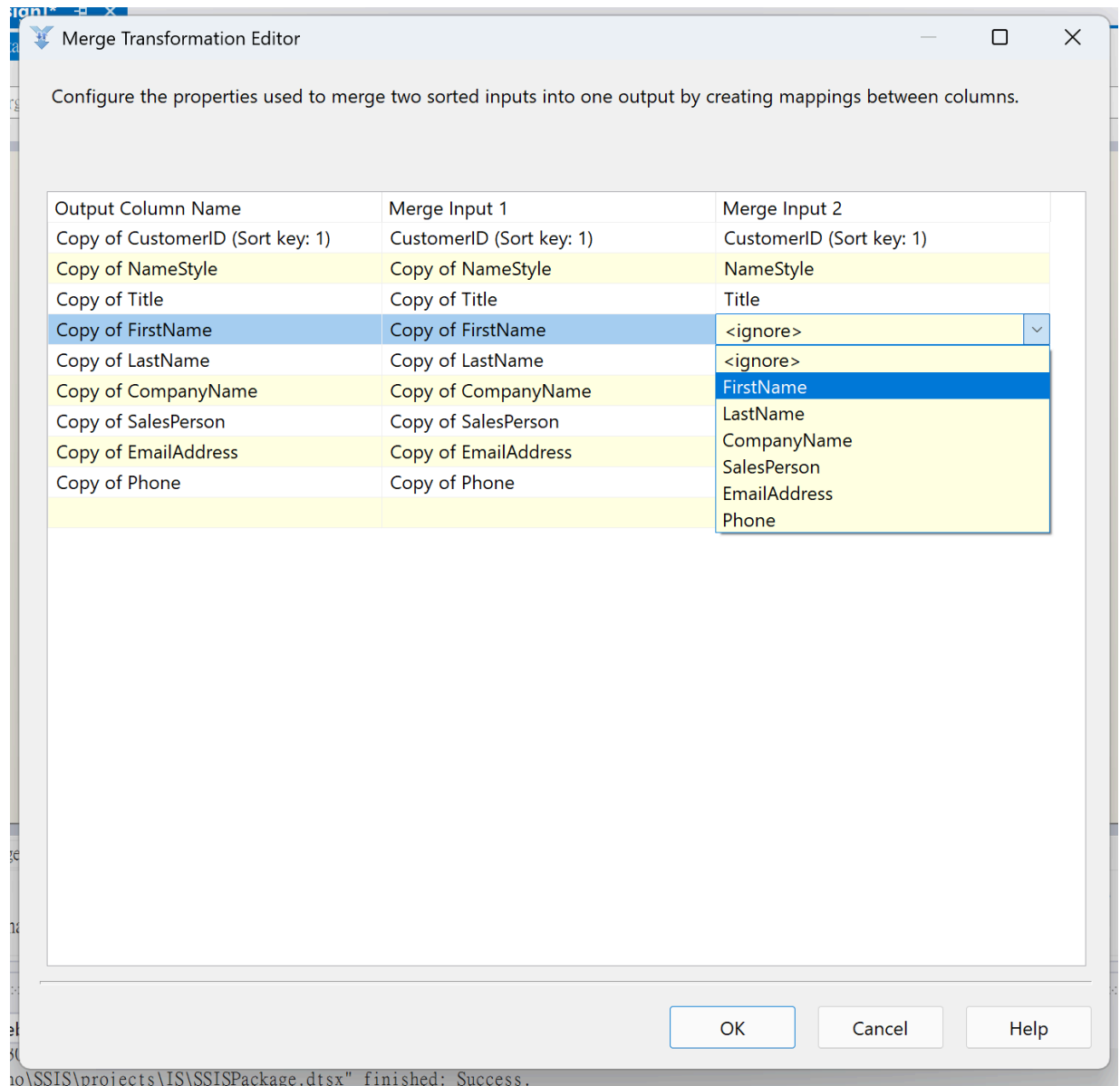
- 在箭頭上右鍵點擊解決參考文獻(resolve references)



- [checked] delete unmapped from Excel

7. 編輯 merge

- 刪上面沒用到的並完成對應



稽核 (Audit)

練習十二

1. 合併文字檔案 Source01 和 Source02 。
2. 添加審計，並且
3. 將合併結果匯出到 Excel 。

參考資料與下載連結:

- [SQL Server Integration Services 官方文件](#)
- [SQL Server Data Tools 下載](#)
- [Microsoft Learn: SSIS 入門課程](#)
- [ETL基礎概念白皮書](#)