Special Design Scenarios

Handling Text and XML Files

Stacia Misner
blog.datainspirations.com
@StaciaMisner

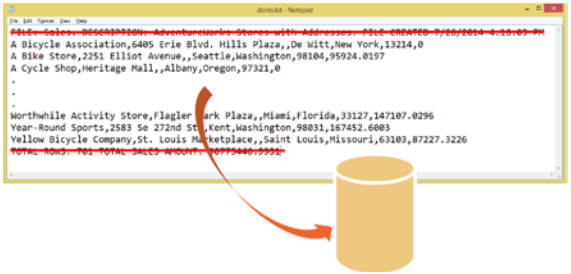
pluralsight
hardcore dev and IT training

Overview

- Working with Headers and Footers in Text Files
- XML Processing

Working with Headers and Footers in Text Files

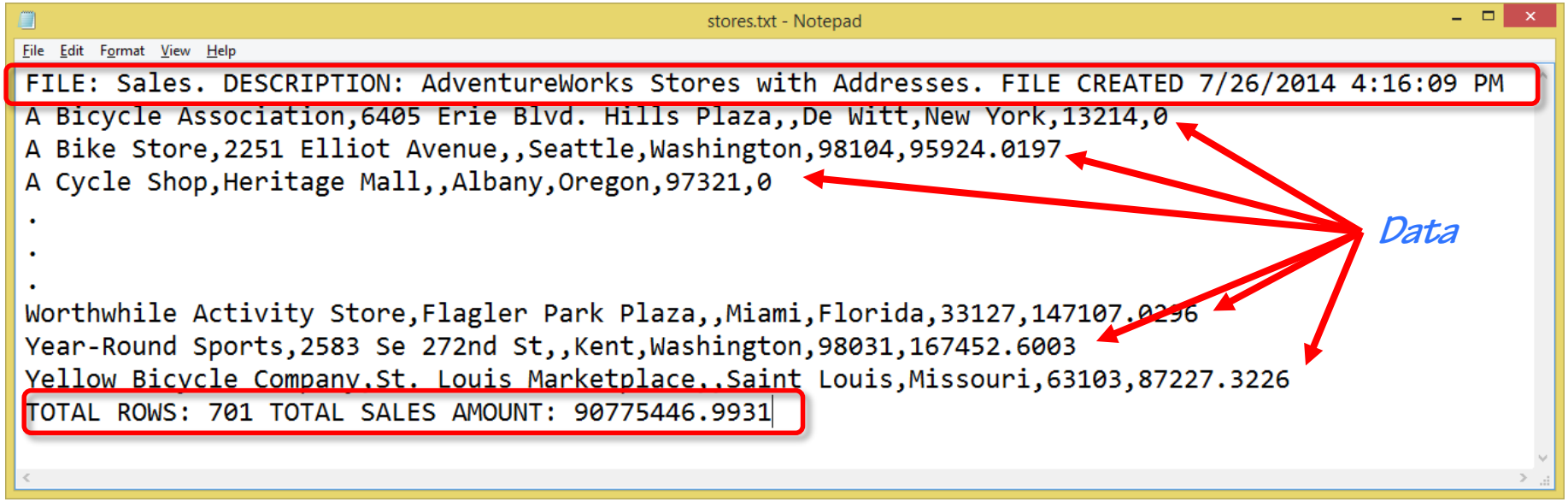
Adding Header and Footer Rows



Eliminating Header and Footer Rows

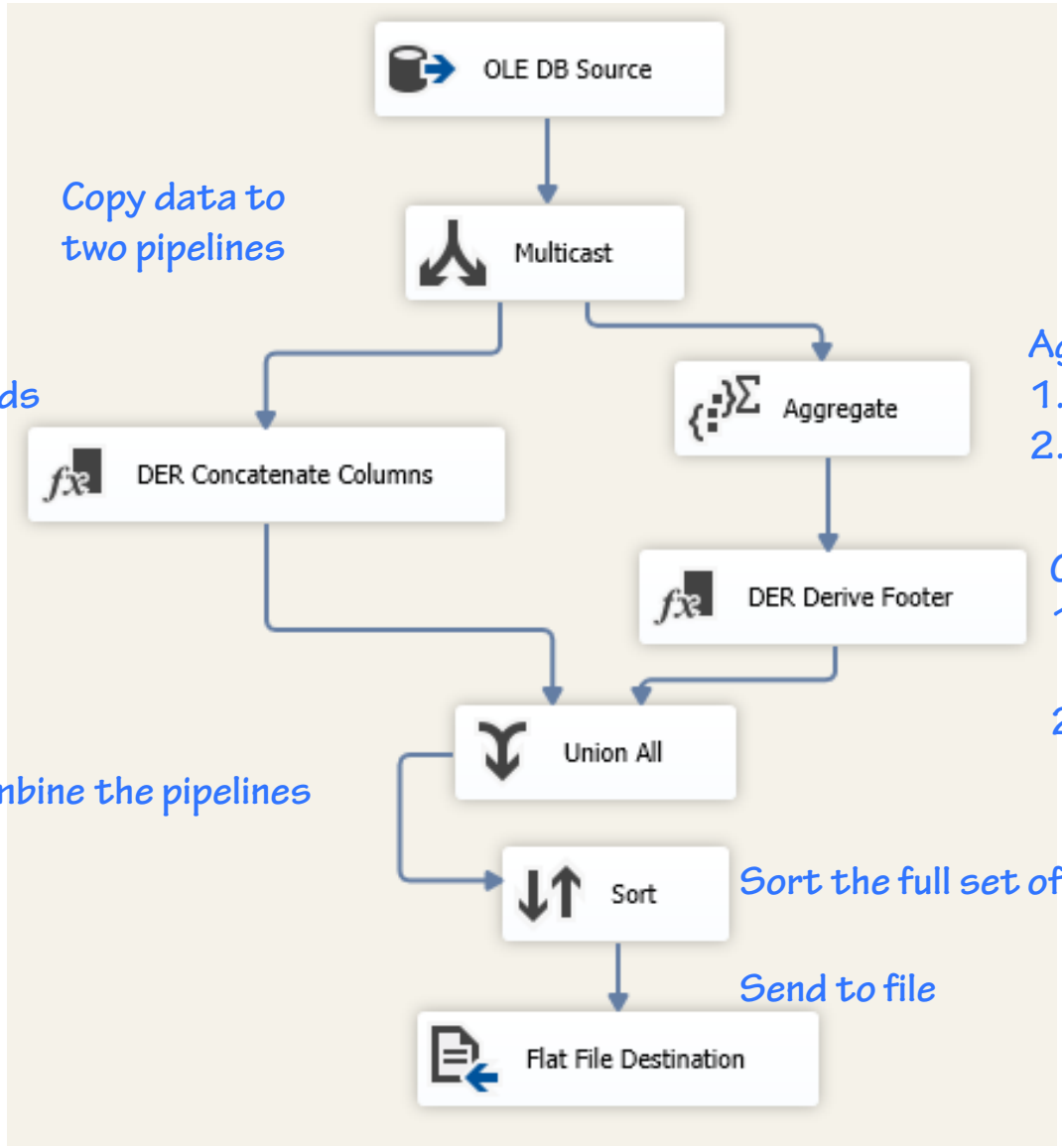
Adding Header and Footer Rows: No Data in Header

Header Row

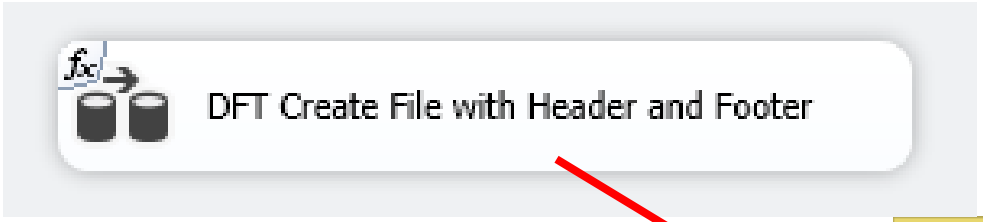


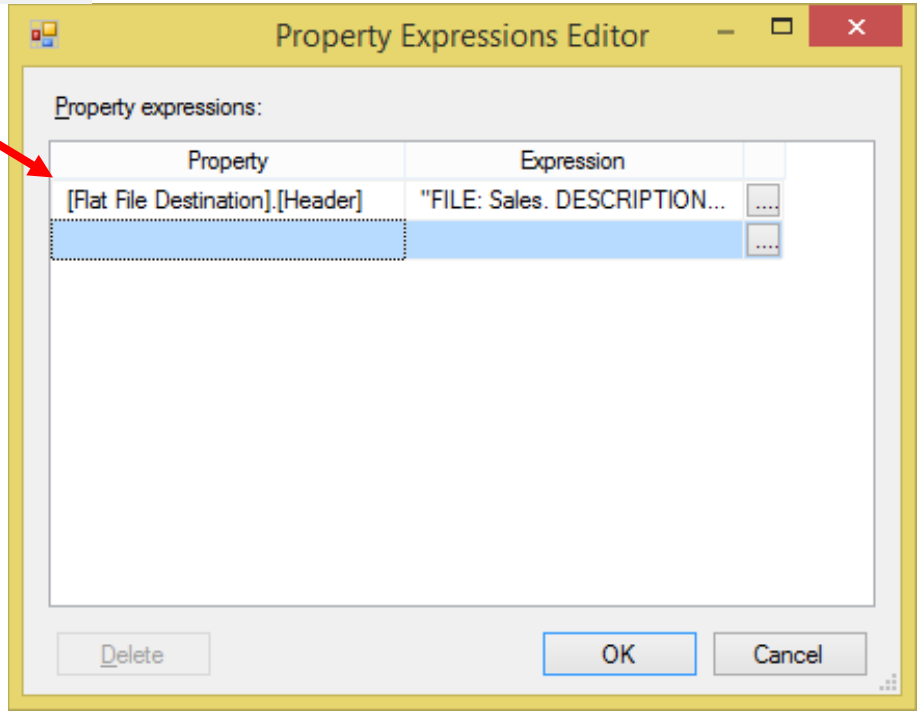
Footer Row

Adding Headers and Footer Rows: No Data in Header



Adding Headers and Footer Rows: No Data in Header

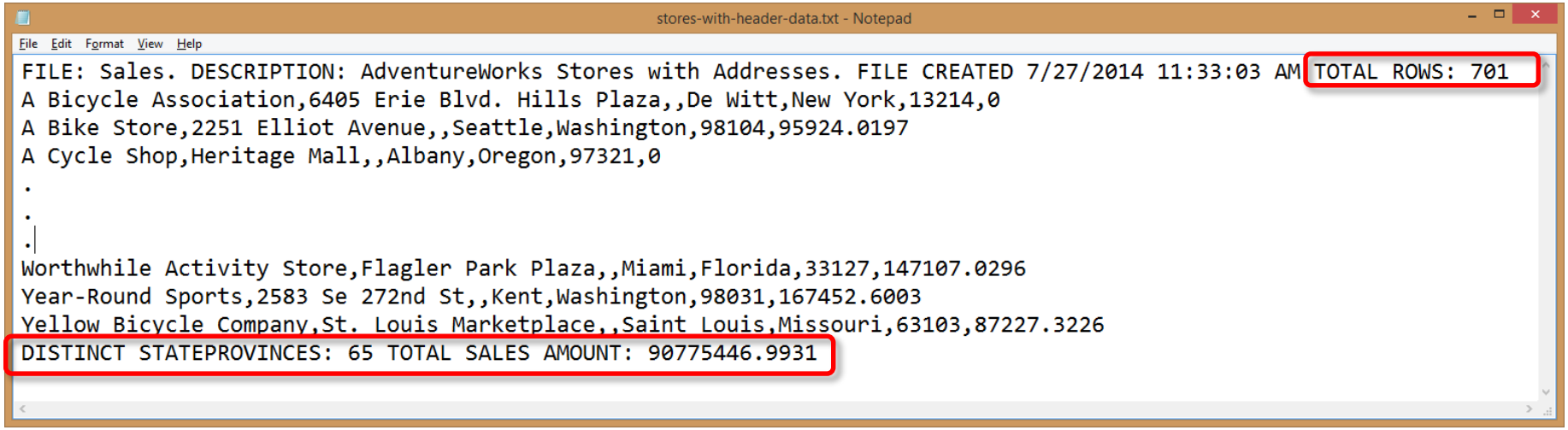





```
"FILE: Sales. DESCRIPTION:
AdventureWorks Stores with Addresses.
FILE CREATED " +
(DT_STR, 25, 1252)@[System::StartTime]
```

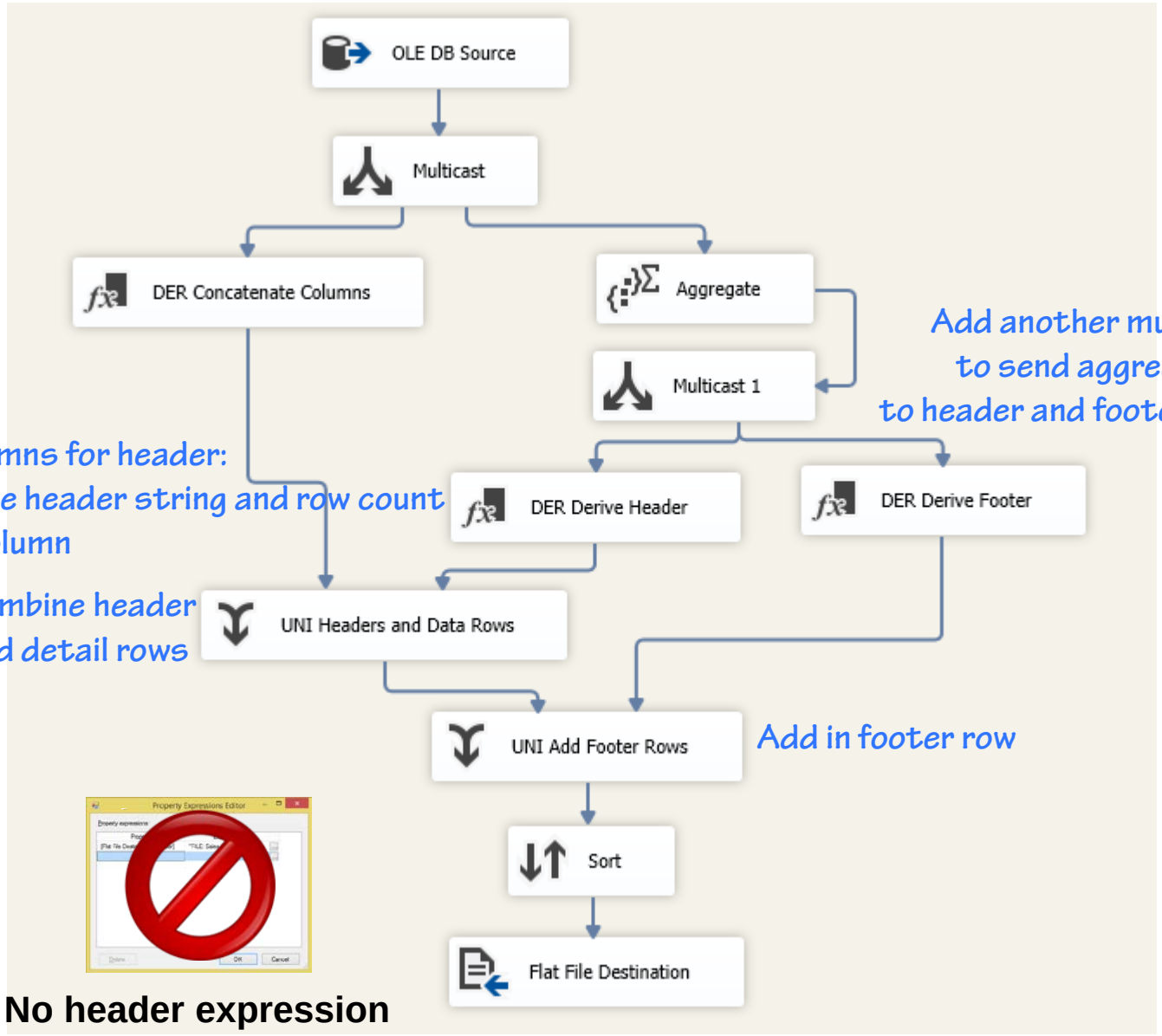
Adding Header and Footer Rows: Data in Header

Header Row includes Computed Data



Footer Row

Adding Headers and Footer Rows: Data in Header



Add another multicast to send aggregates to header and footer pipelines

Create two columns for header:

1. Concatenate header string and row count
2. Add sort column

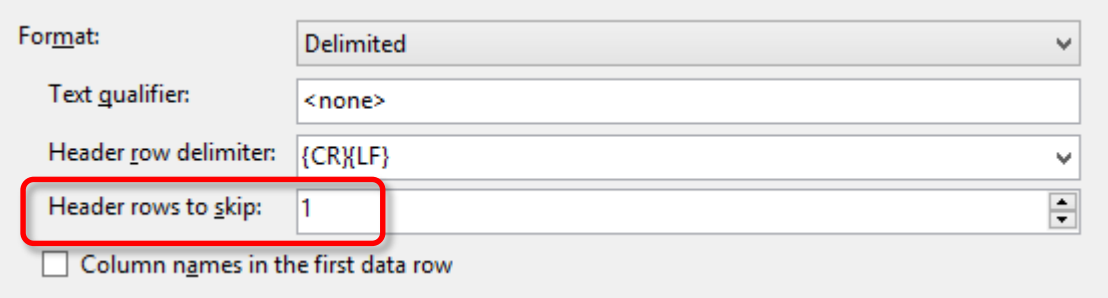
Combine header and detail rows

Add in footer row

No header expression

Eliminating Header Rows

Flat File Connection Manager



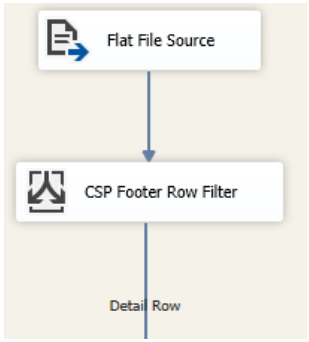
Eliminating Footer Rows

stores-with-header-data.txt - Notepad

File Edit Format View Help

```
FILE: Sales. DESCRIPTION: AdventureWorks Stores with Addresses. FILE CREATED 7/27/2014 11:33:03 AM TOTAL ROWS: 701
A Bicycle Association,6405 Erie Blvd. Hills Plaza,,De Witt,New York,13214,0
A Bike Store,2251 Elliot Avenue,,Seattle,Washington,98104,95924.0197
A Cycle Shop,Heritage Mall,,Albany,Oregon,97321,0
.
.
.
Worthwhile Activity Store,Flagler Park Plaza,,Miami,Florida,33127,147107.0296
Year-Round Sports,2583 Se 272nd St,,Kent,Washington,98031,167452.6003
Yellow Bicycle Company,St. Louis Marketplace,,Saint Louis,Missouri,63103,87227.3226
DISTINCT STATEPROVINCES: 65 TOTAL SALES AMOUNT: 90775446.9931
```

Identifiable Footer



Order	Output Name	Condition
1	Footer Row	LEFT(StoreName,8) == "DISTINCT"

Eliminating Footer Rows

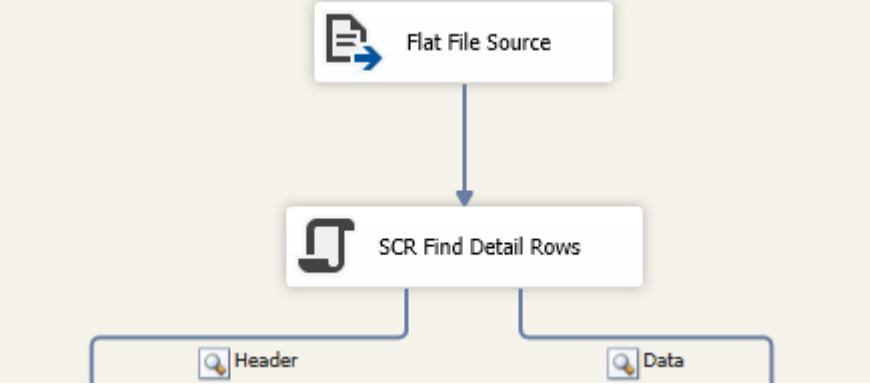
stores-with-header-data.txt - Notepad

File Edit Format View Help

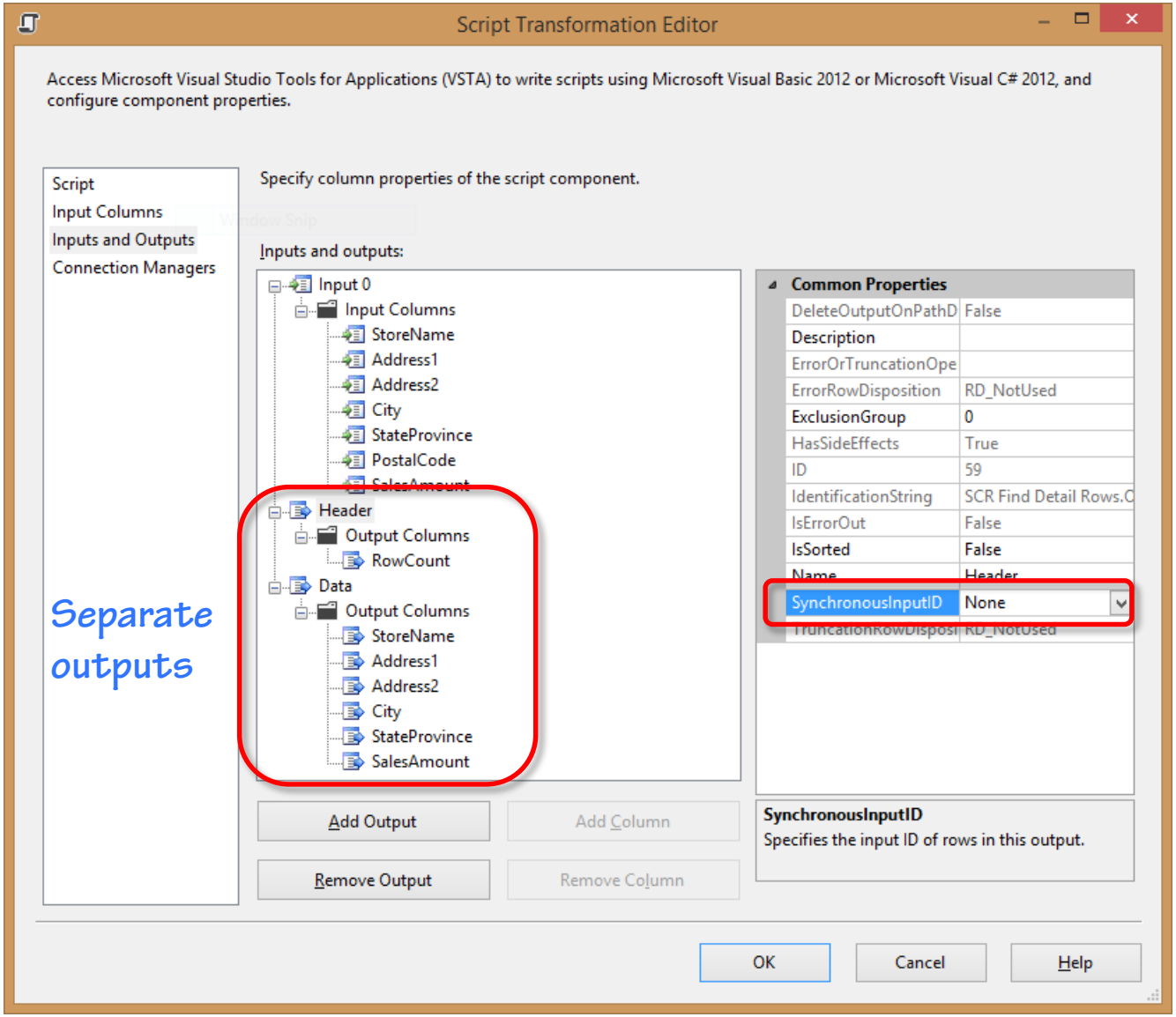
```
FILE: Sales. DESCRIPTION: AdventureWorks Stores with Addresses. FILE CREATED 7/27/2014 11:33:03 AM TOTAL ROWS: 701
A Bicycle Association,6405 Erie Blvd. Hills Plaza,,De Witt,New York,13214,0
A Bike Store,2251 Elliot Avenue,,Seattle,Washington,98104,95924.0197
A Cycle Shop,Heritage Mall,,Albany,Oregon,97321,0
.
.
.
Worthwhile Activity Store,Flagler Park Plaza,,Miami,Florida,33127,147107.0296
Year-Round Sports,2583 Se 272nd St,,Kent,Washington,98031,167452.6003
Yellow Bicycle Company,St. Louis Marketplace,,Saint Louis,Missouri,63103,87227.3226
DISTINCT STATEPROVINCES: 65 TOTAL SALES AMOUNT: 90775446.9931
```

Metadata

Script component parses header for metadata and separates data from header and footer rows



Eliminating Footer Rows

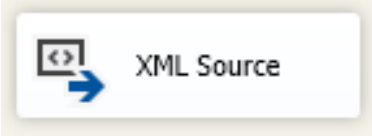


Eliminating Footer Rows

```
int iRowCount = 0;
int iTotalRows = 0;
 public override void Input0_ProcessInputRow(Input0Buffer Row)
    {
        iRowCount++;
        switch (iRowCount)
        {
            case 1:
                // header row
                string sTotalRows;
                sTotalRows = Row.StoreName.Substring(Row.StoreName.IndexOf("TOTAL ROWS:") + 12);
                iTotalRows = Convert.ToInt32(sTotalRows);
                HeaderBuffer.AddRow();
                HeaderBuffer.RowCount = iTotalRows;
                break;
            default:
                //data or footer row
                if (iRowCount <= iTotalRows + 1)
                {
                    DataBuffer.AddRow();
                    DataBuffer.StoreName = Row.StoreName;
                    DataBuffer.Address1 = Row.Address1;
                    DataBuffer.Address2 = Row.Address2;
                    DataBuffer.City = Row.City;
                    DataBuffer.StateProvince = Row.StateProvince;
                    DataBuffer.SalesAmount = Row.SalesAmount;
                }
                break;
        }
    }
```

Counter

XML Processing



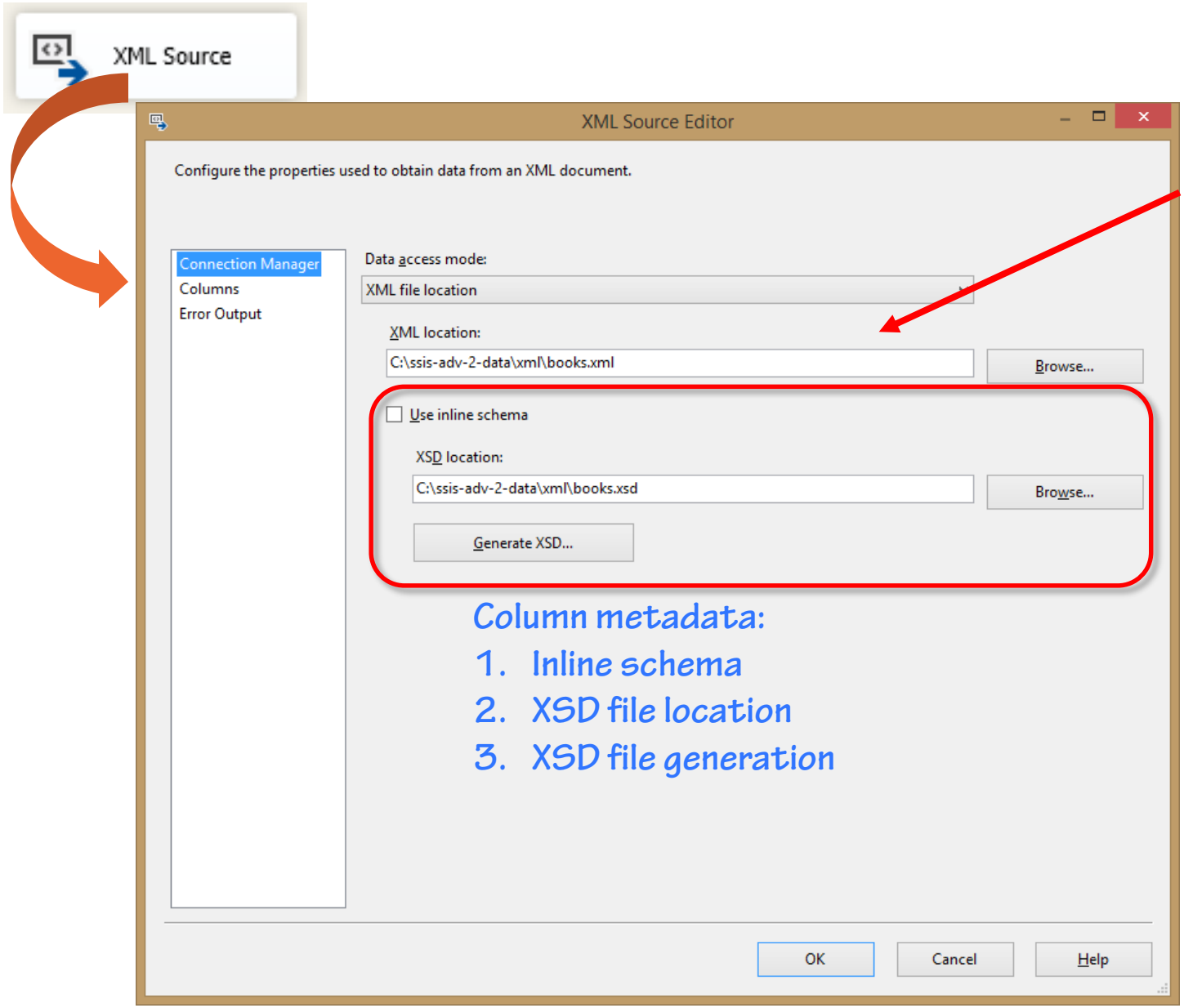
Importing
XML Data

Validating
XML Data

```
<?xml version="1.0" encoding="utf-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <xsl:output method="xml" indent="yes"/>
    <xsl:template match="catalog">
      <xsl:for-each select ="book">
        <book>
          <id>
            <xsl:value-of select ="@id"/>
          </id>
          <title>
            <xsl:value-of select ="title"/>
          </title>
          <genre>
            <xsl:value-of select ="genre"/>
          </genre>
          <price>
            <xsl:value-of select ="price"/>
          </price>
          <publish_date>
            <xsl:value-of select ="publish_info/publish_date"/>
          </publish_date>
          <publisher>
            <xsl:value-of select ="publish_info/publisher"/>
          </publisher>
          <description>
            <xsl:value-of select ="description"/>
          </description>
        </book>
      </xsl:for-each>
    </xsl:template>
</xsl:stylesheet>
```

Transforming
XML Data

Importing XML Data



XML data source:

1. File by location
2. File by variable location
3. Variable with XML data

Column metadata:

1. Inline schema
2. XSD file location
3. XSD file generation

Single Output from XML File

Elements and Sub-elements

```
<catalog>
   <book id="bk101">
      <author>Gambardella, Matthew</author>
      <title>XML Developer's Guide</title>
   </book>
</catalog>
```

Element = row

Attribute = column

Sub-element = column

Values as Attributes

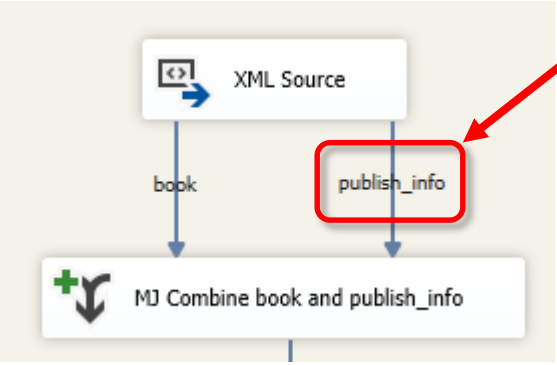
```
<catalog>
  <book id="bk101" author="Gambardella, Matthew" title="XML Developer's Guide" />
</catalog>
```

Element = row

Attribute = column

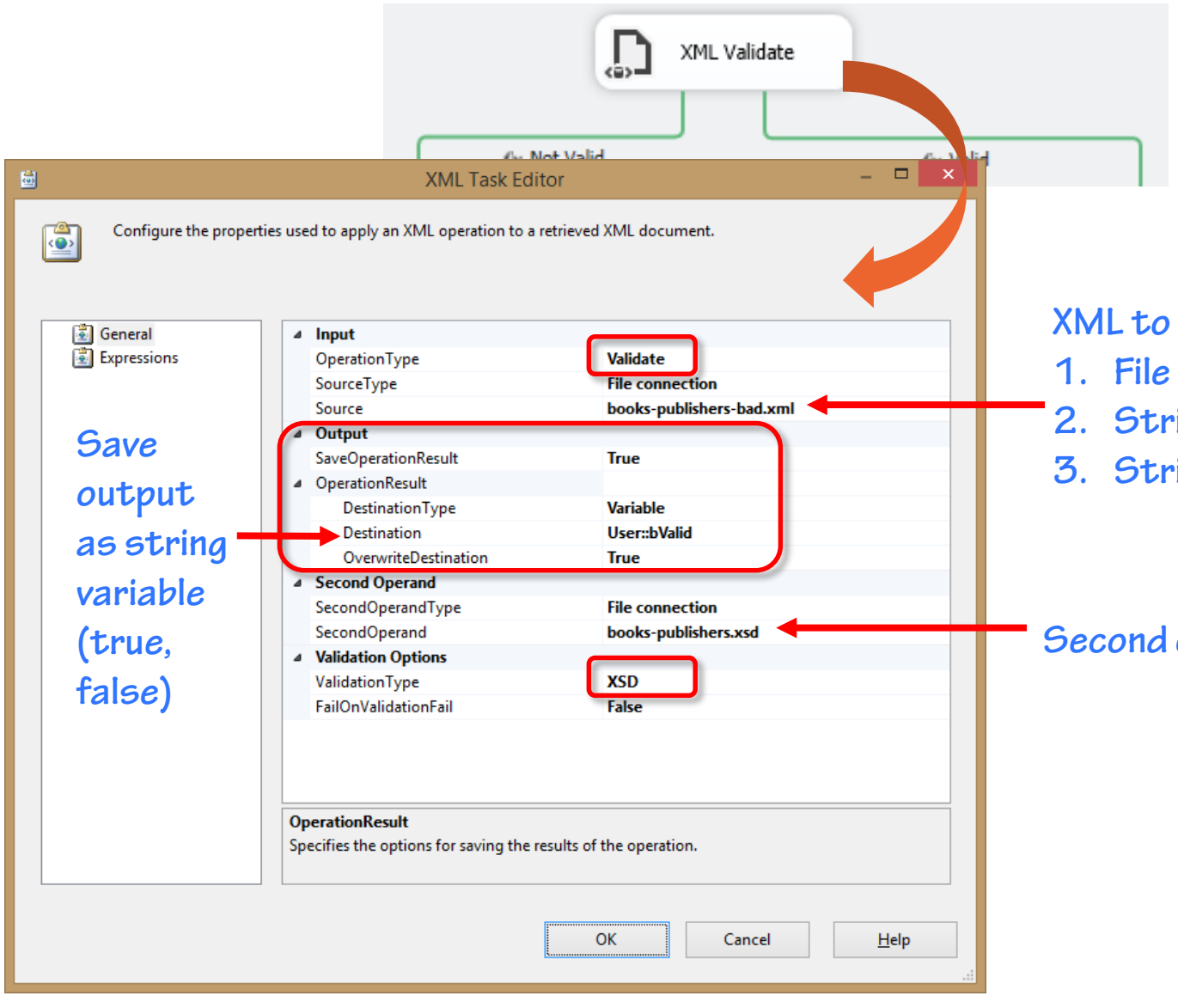
Multiple Outputs from XML File

```
<catalog>
   <book id="bk101">
      <author>Gambardella, Matthew</author>
      <title>XML Developer's Guide</title>
      <genre>Computer</genre>
      <price>44.95</price>
      <publish_info>
         <publish_date>2000-10-01</publish_date>
               <publisher>XML Publications Ltd.</publisher>
      </publish_info>
      <description>An in-depth look at creating applications
      with XML.</description>
   </book>
</catalog>
```



Nested elements produce additional output

Validating XML Data



Save output as string variable (true, false)

XML to validate:

1. File
2. String constant
3. String variable

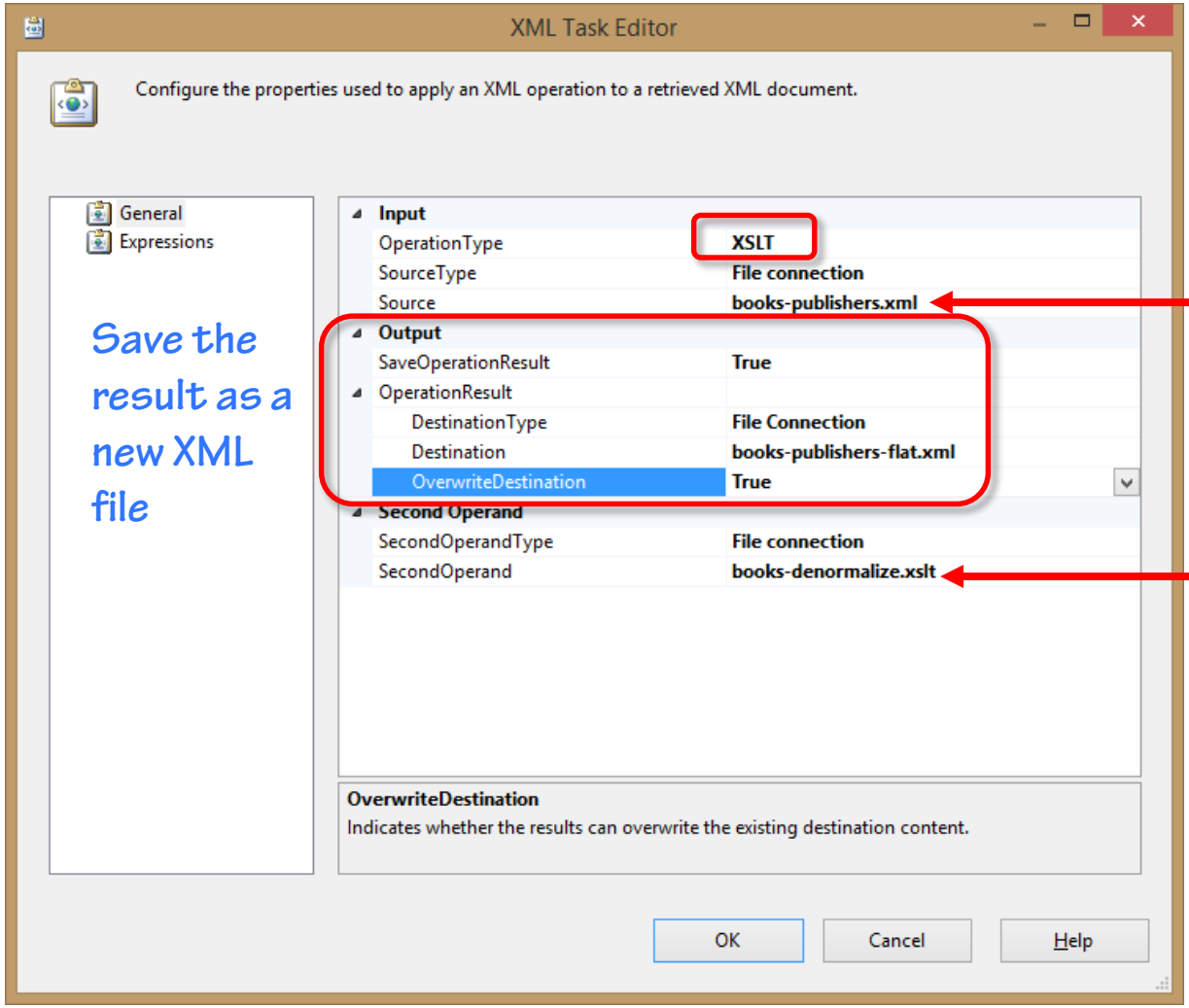
Second operand = XSD file

Transforming XML Data

```
<?xml version="1.0" encoding="utf-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
    <xsl:output method="xml" indent="yes"/>
    <xsl:template match="catalog">
        <xsl:for-each select ="book">
          <book>
            <id>
              <xsl:value-of select ="@id"/>
            </id>
            <title>
              <xsl:value-of select ="title"/>
            </title>
            <genre>
              <xsl:value-of select ="genre"/>
            </genre>
            <price>
              <xsl:value-of select ="price"/>
            </price>
            <publish_date>
              <xsl:value-of select ="publish_info/publish_date"/>
            </publish_date>
            <publisher>
              <xsl:value-of select ="publish_info/publisher"/>
            </publisher>
            <description>
              <xsl:value-of select ="description"/>
            </description>
          </book>
        </xsl:for-each>
    </xsl:template>
</xsl:stylesheet>
```

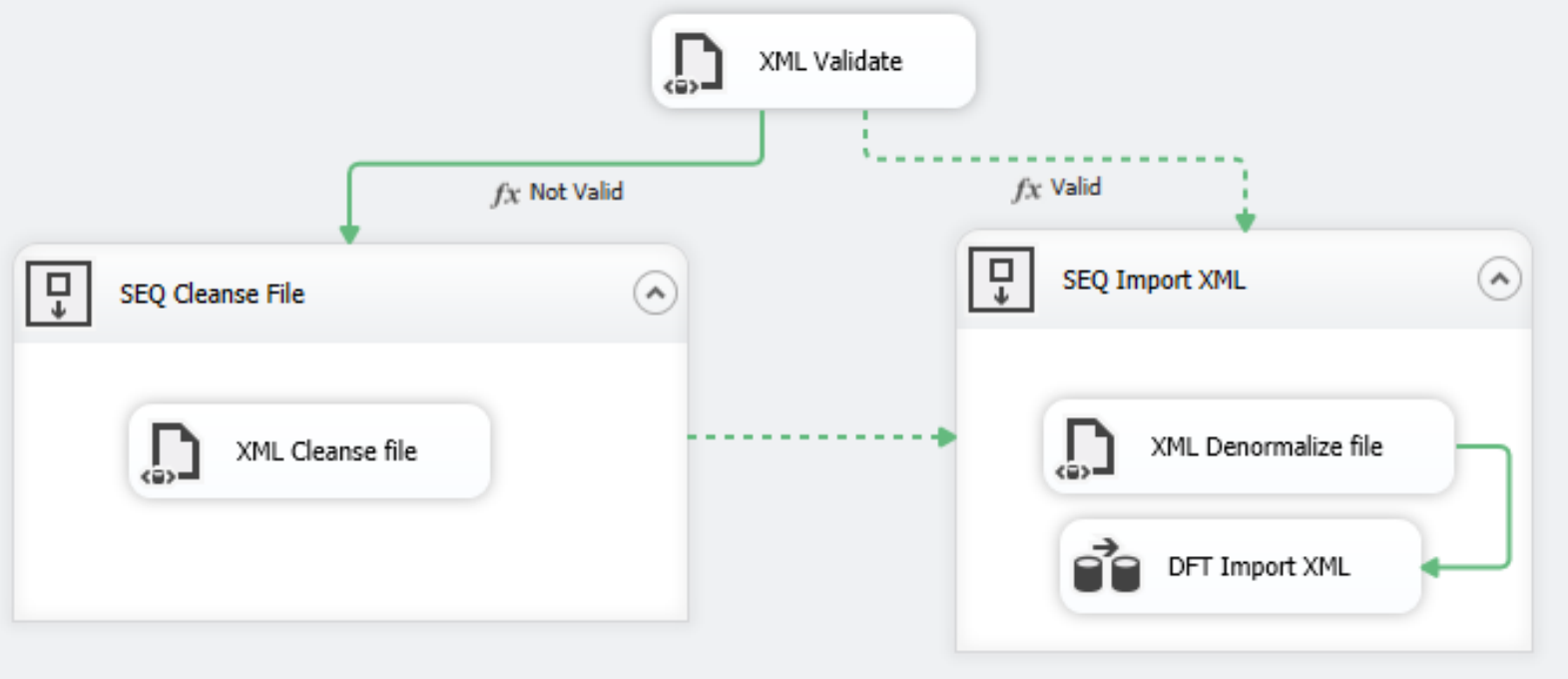
Promote nested sub-elements

Transforming XML Data



Transforming XML Data

Validate XML with XSD file



Apply business rules to cleanse XML using XSLT script

Simplify XML using XSLT script and run Data Flow Task

Summary

- **Working with Headers and Footers in Text Files**
  - Adding headers and footers
    - No Header Data: Header property of Data Flow Task
    - Header Data and/or Footers: Multicast plus Aggregate plus Derived Column transforms for both headers and footers then Union and Sort transforms
  - Eliminating headers and footers
    - Flat File connection: set property to skip header rows
    - Identifiable footer: use Conditional Split to filter
    - Non-identifiable: use Script Component to determine row count and filter
- **XML Processing**
  - XML Source component requires clean, simply structured XML
  - XML Task provides option to validate XML by using XSD
  - XML Task supports XSLT operations to cleanse or restructure XML

Resources

- **Data Flow Properties that can be Set by Using Expressions**
  - http://tinyurl.com/cyqkxx
- **Configuring the Script Component in the Script Component Editor**
  - http://tinyurl.com/qd3uqfp
- **XSL Transformations**
  - http://www.w3.org/TR/xslt11/
- **Pluralsight course: XSLT 2.0 and 1.0 Foundations**
  - http://tinyurl.com/l74g9kt
- **Create an XML Destination for an SSIS Package**
  - http://tinyurl.com/k4mnzpq