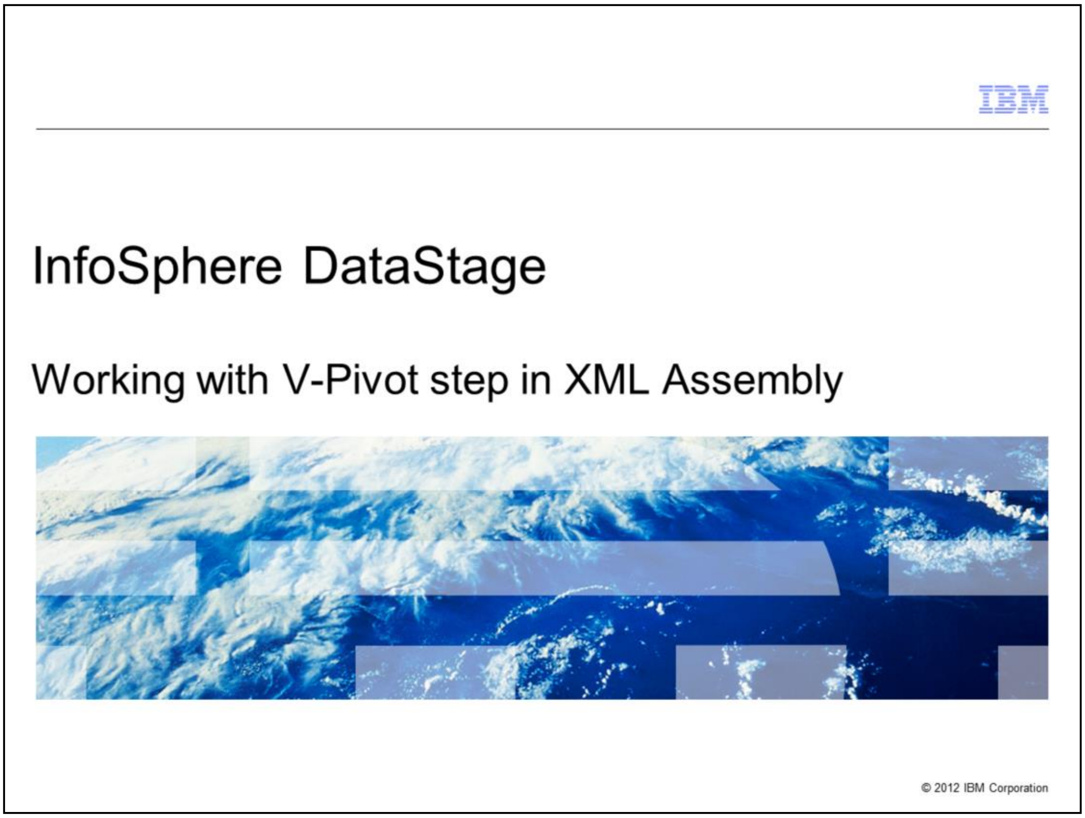


InfoSphere DataStage

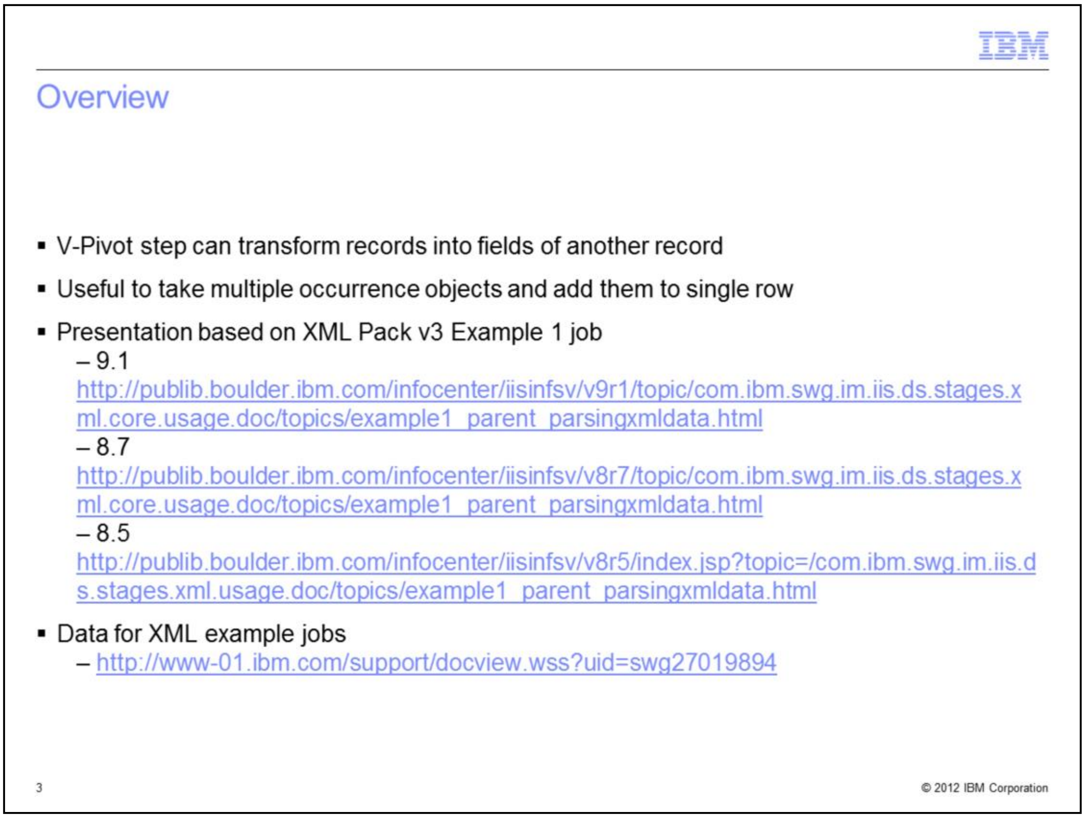
Working with V-Pivot step in XML Assembly

This presentation provides a working example of how to use the Vertical-Pivot step, referred to as V-Pivot step, in the XML Assembly in the XML Pack version 3. The XML Assembly in the XML Pack version 3 is shipped with Information Server DataStage® version 8.5 fix pack 1 and is included in version 8.7 and version 9.1 as well.

Table of contents

- Overview and prerequisites
- Understanding source XML data
- Adding V-Pivot step from palette
- Setting up V-Pivot step
- Mapping data in output step
- Reviewing results

This presentation will provide an overview and working example of how to use the V-Pivot step in the XML Assembly. The example demonstrates how to map XML data that contains multiple list objects into new columns for a single output row. The presentation will review the source XML data and, using an existing XML Example job, show how to add and configure the V-Pivot step. The data mapping in the output step and review of the data produced will also be covered.



Overview

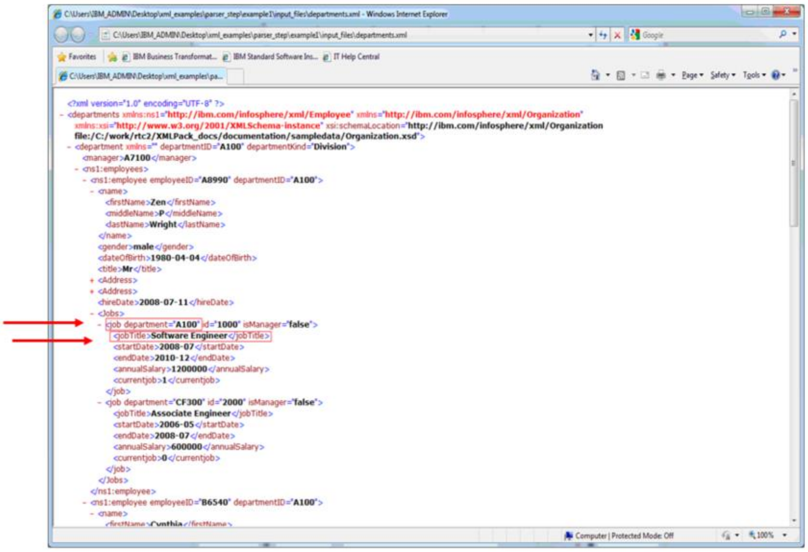
- V-Pivot step can transform records into fields of another record
- Useful to take multiple occurrence objects and add them to single row
- Presentation based on XML Pack v3 Example 1 job
  - 9.1 http://publib.boulder.ibm.com/infocenter/iisinfsv/v9r1/topic/com.ibm.swg.im.iis.ds.stages.xml.core.usage.doc/topics/example1_parent_parsingxmldata.html
  - 8.7 http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r7/topic/com.ibm.swg.im.iis.ds.stages.xml.core.usage.doc/topics/example1_parent_parsingxmldata.html
  - 8.5 http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/index.jsp?topic=/com.ibm.swg.im.iis.ds.stages.xml.usage.doc/topics/example1_parent_parsingxmldata.html
- Data for XML example jobs
  - http://www-01.ibm.com/support/docview.wss?uid=swg27019894

The V-Pivot step can help to manipulate XML data in a single stage so the required pivoting does not have to be done outside of the XML Stage. The V-Pivot step is capable of transforming multiple records into fields of another record. The V-Pivot Stage can be useful when XML data contains multiple child records that need to be added to a single row of output data.

This presentation is built from the XML Example 1 job that is shipped with Information Server and the XML pack. You should familiarize yourself with the example and have it working to get the most out of this presentation.

Source XML data

- Employee Input XML file
  - Multiple job positions
  - Multiple departments

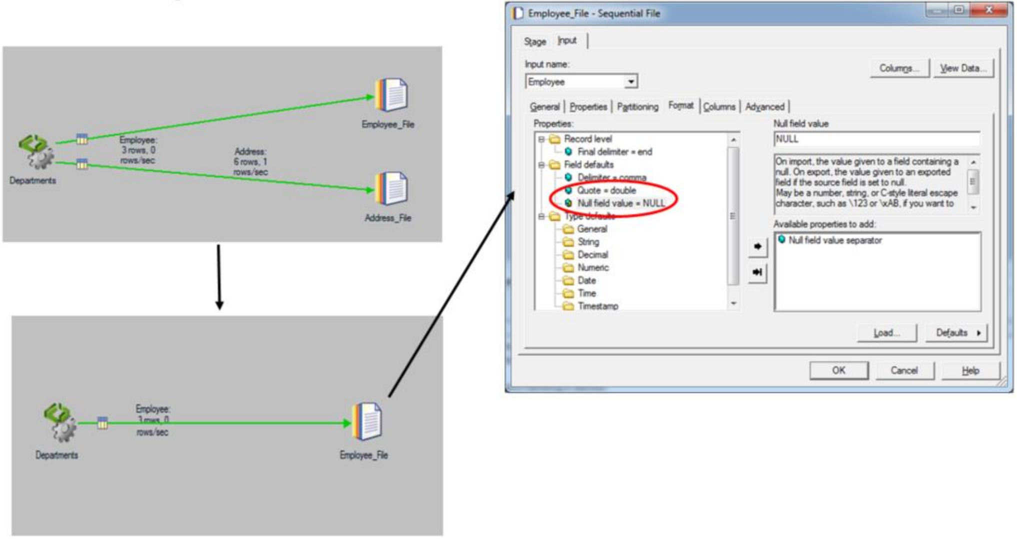

C:\Users\IBM_ADMIN\Desktop\xml_examples\parser_step\example1\input_files\departments.xml - Windows Internet Explorer
C:\Users\IBM_ADMIN\Desktop\xml_examples\parser_step\example1\input_files\departments.xml
Google
Favorites
IBM Business Transformat...
IBM Standard Software Ins...
IT Help Central
C:\Users\IBM_ADMIN\Desktop\xml_examples\pa...
Page
Safety
Tools
Computer | Protected Mode: Off
100%

```
  <?xml version="1.0" encoding="UTF-8" ?>
- <departments xmlns:ns1="http://ibm.com/infosphere/xml/Employee" xmlns="http://ibm.com/infosphere/xml/Organization"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://ibm.com/infosphere/xml/Organization
   file:/C:/work/rtc2/XMLPack_docs/documentation/sampledata/Organization.xsd">
  - <department xmlns="" departmentID="A100" departmentKind="Division">
      <manager>A7100</manager>
    - <ns1:employees>
      - <ns1:employee employeeID="A8990" departmentID="A100">
        - <name>
            <firstName>Zen</firstName>
            <middleName>P</middleName>
            <lastName>Wright</lastName>
          </name>
          <gender>male</gender>
          <dateOfBirth>1980-04-04</dateOfBirth>
          <title>Mr</title>
        + <Address>
        + <Address>
          <hireDate>2008-07-11</hireDate>
        - <Jobs>
          - <job department="A100" id="1000" isManager="false">
              <jobTitle>Software Engineer</jobTitle>
              <startDate>2008-07</startDate>
              <endDate>2010-12</endDate>
              <annualSalary>1200000</annualSalary>
              <currentjob>1</currentjob>
            </job>
          - <job department="CF300" id="2000" isManager="false">
              <jobTitle>Associate Engineer</jobTitle>
              <startDate>2006-05</startDate>
              <endDate>2008-07</endDate>
              <annualSalary>600000</annualSalary>
              <currentjob>0</currentjob>
            </job>
          </Jobs>
        </ns1:employee>
      - <ns1:employee employeeID="B6540" departmentID="A100">
        - <name>
            <firstName>Cynthia</firstName>
```

In the example displayed on this slide, the Input XML data contains multiple employee records and each employee can have multiple past and current job positions within different departments. In this example, the V-Pivot Stage is used to gather all current and past job titles from the child elements and puts them into a single output row for each employee. Displayed on this slide are arrows pointing at the two XML elements that are used in this example to pivot the data. The use of these two elements are explained in greater detail in later slides.

Job design (1 of 2)

- Edit XML Example 1 job
- Remove Address link and file
- Set null handling



Employee_File - Sequential File

Stage | Input

Input name: Employee

Columns... | View Data...

General | Properties | Partitioning | Format | Columns | Advanced

Properties:

- Record level
  - Final delimiter = end
- Field defaults
  - Delimiter = comma
  - Quote = double
  - Null field value = NULL
- Type defaults
  - General
  - String
  - Decimal
  - Numeric
  - Date
  - Time
  - Timestamp

Null field value: NULL

On import, the value given to a field containing a null. On export, the value given to an exported field if the source field is set to null. May be a number, string, or C-style literal escape character, such as \123 or \xAB, if you want to

Available properties to add:

- Null field value separator

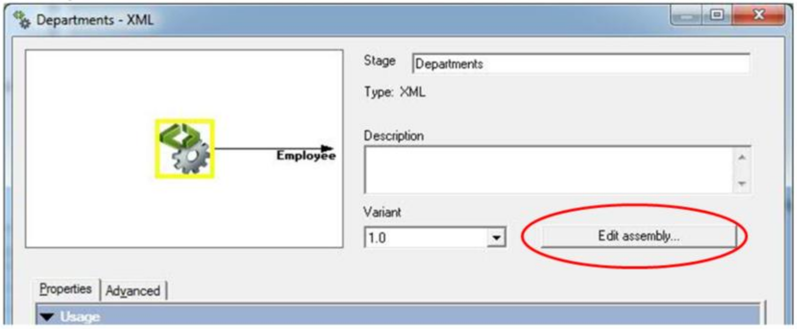
Load... | Defaults ▸

OK | Cancel | Help

This V-Pivot example uses the XML Example 1 job as a starting point. The first step is to remove the Address link and Address_File stages from the canvas, since the focus is on the Employee_File output. The second modification is to add null handling in the Employee_File stage, since the XML Stage in this example will produce null values for some of the pivoted columns.

Job design (2 of 2)

- Open XML Stage
- Click Edit assembly



Departments - XML

Stage: Departments

Type: XML

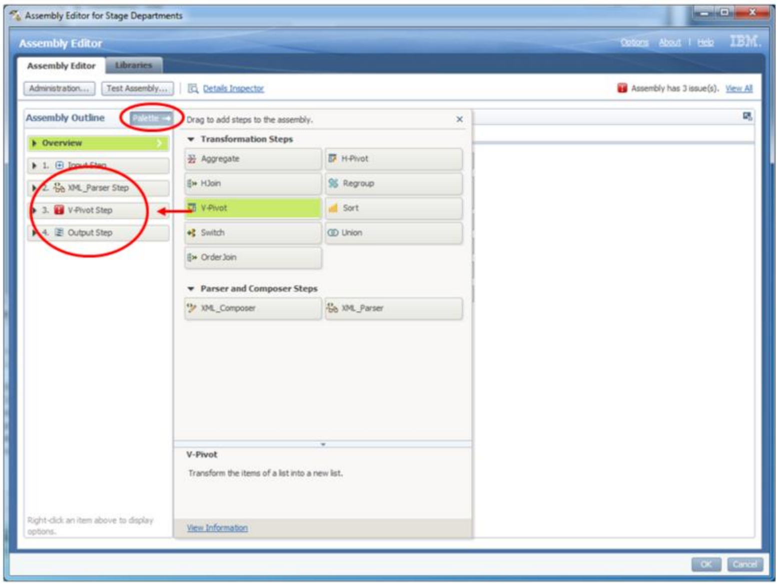
Description

Variant: 1.0

Once the unnecessary link and stage have been removed, it is time to work on the XML assembly and add the V-Pivot step. First, open the XML Stage Properties and click the Edit assembly button to launch the Assembly Editor.

Adding V-Pivot step to XML Assembly

- Add V-Pivot between XML Parser and Output



Assembly has 3 issue(s).

Drag to add steps to the assembly.

Transform the items of a list into a new list.

Right-click an item above to display options.

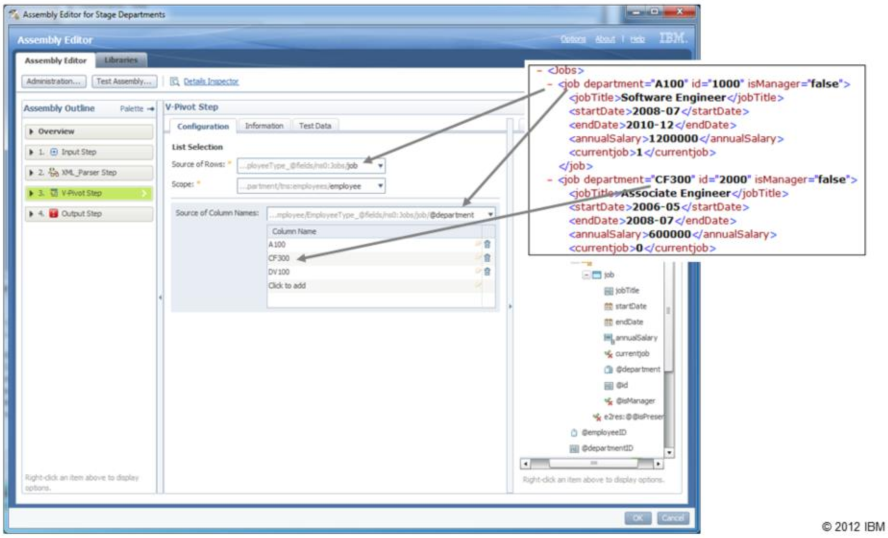
7

© 2012 IBM Corporation

Once in the Assembly Editor, the V-Pivot step can be found in the Palette. If your Palette does not display like it does here on the slide, click the Palette button. Grab the V-Pivot on the Palette and drag it to the left. Drop it between the XML Parser and Output Step so it appears as it does in the screen capture displayed on this slide. Click the V-Pivot step to begin the required setup.

Setting up V-Pivot step

- Set Source of Rows
- Set Scope
- Set Source of Column Names and add columns



Assembly Editor for Stage Departments

Assembly Editor

V-Pivot Step

List Selection

Source of Rows: * ...ployeeType_@fields/ns0:Jobs/job

Scope: * ...partment/tns:employees/employee

Source of Column Names: ...mployee/EmployeeType_@fields/ns0:Jobs/job/@department

Column Name
A100
CF300
DV100
Click to add

```
- <Jobs>
  - <job department="A100" id="1000" isManager="false">
      <jobTitle>Software Engineer</jobTitle>
      <startDate>2008-07</startDate>
      <endDate>2010-12</endDate>
      <annualSalary>1200000</annualSalary>
      <currentjob>1</currentjob>
    </job>
  - <job department="CF300" id="2000" isManager="false">
      <jobTitle>Associate Engineer</jobTitle>
      <startDate>2006-05</startDate>
      <endDate>2008-07</endDate>
      <annualSalary>600000</annualSalary>
      <currentjob>0</currentjob>
```

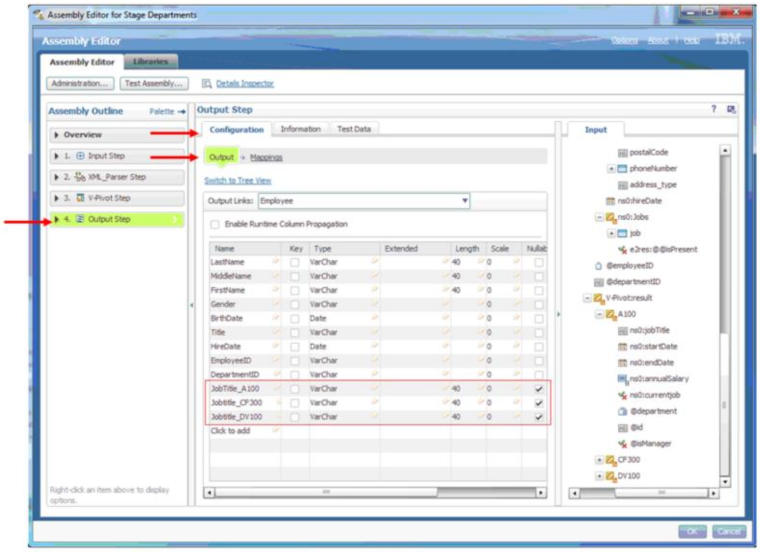
In the V-Pivot step there are several things that have to be setup correctly to obtain the desired output. The first step is to click the drop down for “Source of Rows”. This needs to point to “job” since that record contains the jobTitle element that is going to be appended to the output records.

Next, click the drop down for Scope and set it to “employee”, so the pivoting is limited to each employee's child records and not all job records within the XML data.

Next, the Source Of Column Names needs to be set. This needs to point to the @department element. Next, you need to create column names to represent each possible value for job department. You can call these columns anything you like but for simplicity, this example uses the same names as the job department. Basically, this sets the V-Pivot step to collect all the job records for an employee that has a matching @department value defined in the Column Name list. The matched records can then be mapped in the output step.

Output Step and Mapping data (1 of 2)

- Click Output Setup => Configuration => Output
- Add job title columns



Output Links: Employee

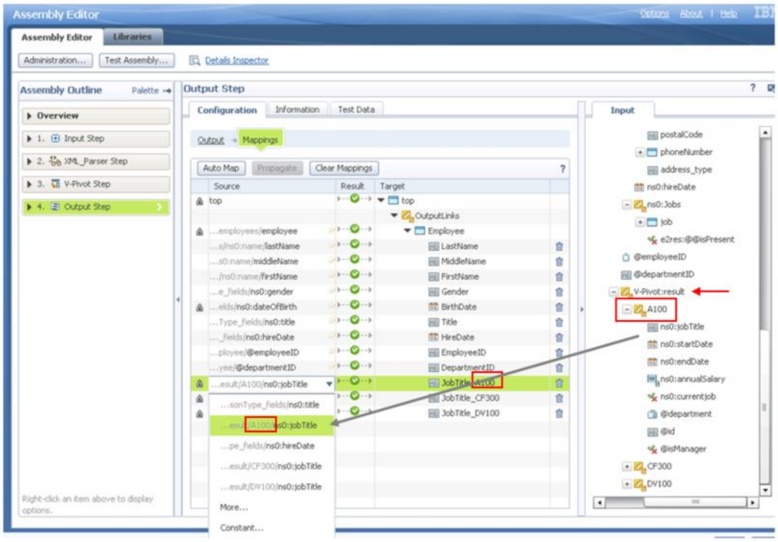
□ Enable Runtime Column Propagation

Name	Key	Type	Extended	Length	Scale	Nullab
LastName	□	VarChar		40	0	□
MiddleName	□	VarChar		40	0	□
FirstName	□	VarChar		40	0	□
Gender	□	VarChar			0	□
BirthDate	□	Date			0	□
Title	□	VarChar			0	□
HireDate	□	Date			0	□
EmployeeID	□	VarChar			0	□
DepartmentID	□	VarChar			0	□
JobTitle_A100	□	VarChar		40	0	☑
Jobtitle_CF300	□	VarChar		40	0	☑
Jobtitle_DV100	□	VarChar		40	0	☑
Click to add						

The next step is to set up the output step. Click “Output Step” on the left side of the dialog box and then the Output link under Configurations. Before the data from the V-Pivot step can be mapped, three new columns need to be added to the output step as displayed in the red box on this slide. These three columns will hold the job titles that the employees hold or held at these departments. Again, these column names can be what ever you choose.

Output Step and Mapping data (2 of 2)

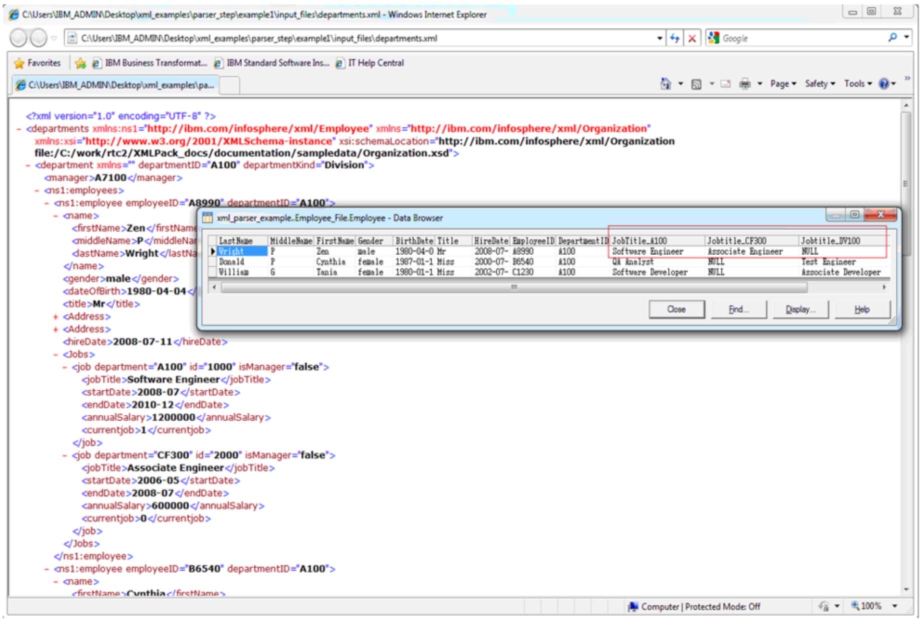
- Click Mappings link
- Match jobTile column to V-Pivot result set



Source	Result	Target
top	✓	top
		OutputLinks
...employees/employee	✓	Employee
...s/ns0:name/lastName	✓	LastName
...s0:name/middleName	✓	MiddleName
.../ns0:name/firstName	✓	FirstName
...e_fields/ns0:gender	✓	Gender
...elds/ns0:dateOfBirth	✓	BirthDate
...Type_fields/ns0:title	✓	Title
..._fields/ns0:hireDate	✓	HireDate
...ployee/@employeeID	✓	EmployeeID
...yee/@departmentID	✓	DepartmentID
...esult/A100/ns0:jobTitle	✓	JobTitle_A100
	✓	JobTitle_CF300
	✓	JobTitle_DV100

Next, click the Mappings link under the Configuration tab. Each jobTitle element from the matching V-Pivot results set, identified by the Column Names defined earlier in the V-Pivot step, must be mapped to target columns as seen in the screen capture displayed on this slide. Click the Source drop down for each of the new columns created on the previous slide and pick the matching jobTitle from the V-Pivot result set. Once this is completed, save, compile and run the job.

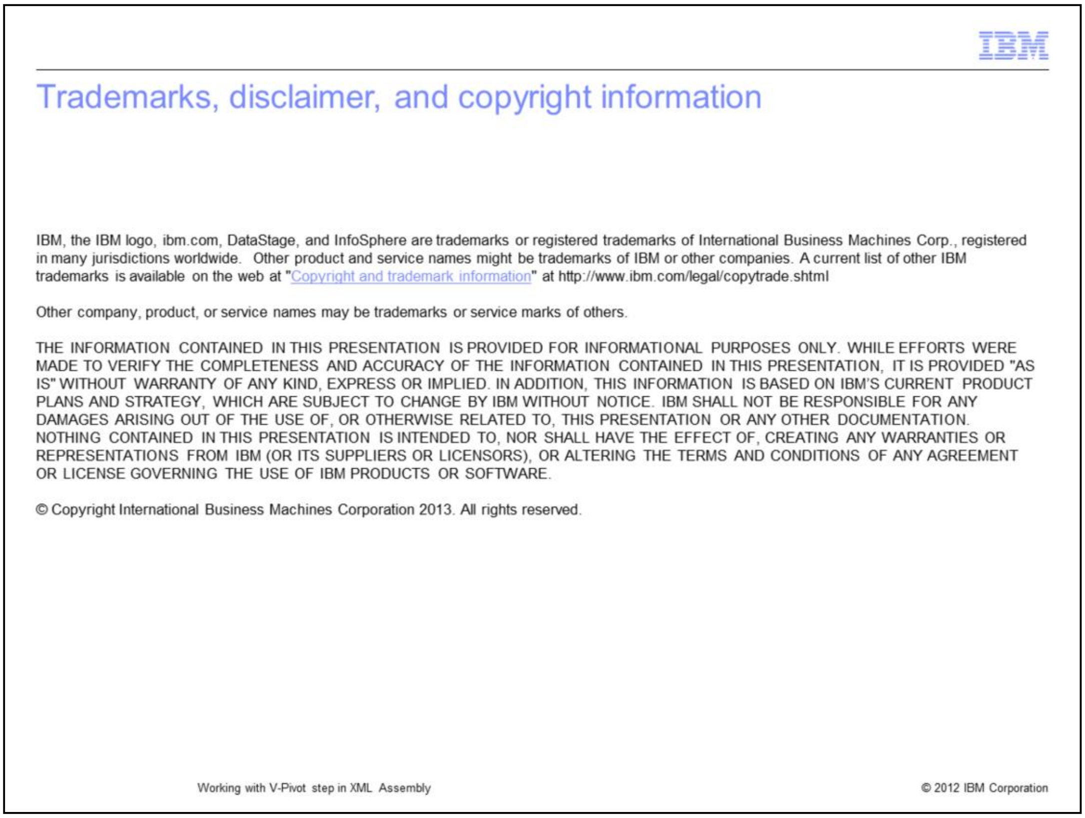
Output data produced


C:\Users\IBM_ADMIN\Desktop\xml_examples\parser_step\example1\input_files\departments.xml - Windows Internet Explorer
C:\Users\IBM_ADMIN\Desktop\xml_examples\parser_step\example1\input_files\departments.xml
Google
Favorites
IBM Business Transformat...
IBM Standard Software Ins...
IT Help Central
C:\Users\IBM_ADMIN\Desktop\xml_examples\pa...
Page
Safety
Tools
xml_parser_example..Employee_File.Employee - Data Browser
Close
Find...
Display...
Help
Computer | Protected Mode: Off
100%

```
  <?xml version="1.0" encoding="UTF-8" ?>
- <departments xmlns:ns1="http://ibm.com/infosphere/xml/Employee" xmlns="http://ibm.com/infosphere/xml/Organization"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://ibm.com/infosphere/xml/Organization
   file:/C:/work/rtc2/XMLPack_docs/documentation/sampledata/Organization.xsd">
  - <department xmlns="" departmentID="A100" departmentKind="Division">
      <manager>A7100</manager>
    - <ns1:employees>
      - <ns1:employee employeeID="A8990" departmentID="A100">
        - <name>
            <firstName>Zen</firstName
            <middleName>P</middleNa
            <lastName>Wright</lastNa
          </name>
          <gender>male</gender>
          <dateOfBirth>1980-04-04</
          <title>Mr</title>
        + <Address>
        + <Address>
          <hireDate>2008-07-11</hireDate>
        - <Jobs>
          - <job department="A100" id="1000" isManager="false">
              <jobTitle>Software Engineer</jobTitle>
              <startDate>2008-07</startDate>
              <endDate>2010-12</endDate>
              <annualSalary>1200000</annualSalary>
              <currentjob>1</currentjob>
            </job>
          - <job department="CF300" id="2000" isManager="false">
              <jobTitle>Associate Engineer</jobTitle>
              <startDate>2006-05</startDate>
              <endDate>2008-07</endDate>
              <annualSalary>600000</annualSalary>
              <currentjob>0</currentjob>
            </job>
          </Jobs>
        </ns1:employee>
      - <ns1:employee employeeID="B6540" departmentID="A100">
        - <name>
            <firstName>Cynthia</firstName>
```

LastName	MiddleName	FirstName	Gender	BirthDate	Title	HireDate	EmployeeID	DepartmentID	JobTitle_A100	Jobtitle_CF300	Jobtitle_DV100
Wright	P	Zen	male	1980-04-0	Mr	2008-07-	A8990	A100	Software Engineer	Associate Engineer	NULL
Donald	P	Cynthia	female	1987-01-1	Miss	2000-07-	B6540	A100	QA Analyst	NULL	Test Engineer
William	G	Tania	female	1980-01-1	Miss	2002-07-	C1230	A100	Software Developer	NULL	Associate Developer

Once the job is run, right click the sequential file stage and choose View Employee Data. The first output record clearly shows how the data from the multiple child records was pivoted to a single output row. Mr. Wright has not worked for department “DV100” thus, NULL is shown in the Jobtitle_DV100 column.



Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, DataStage, and InfoSphere are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

Other company, product, or service names may be trademarks or service marks of others.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.