



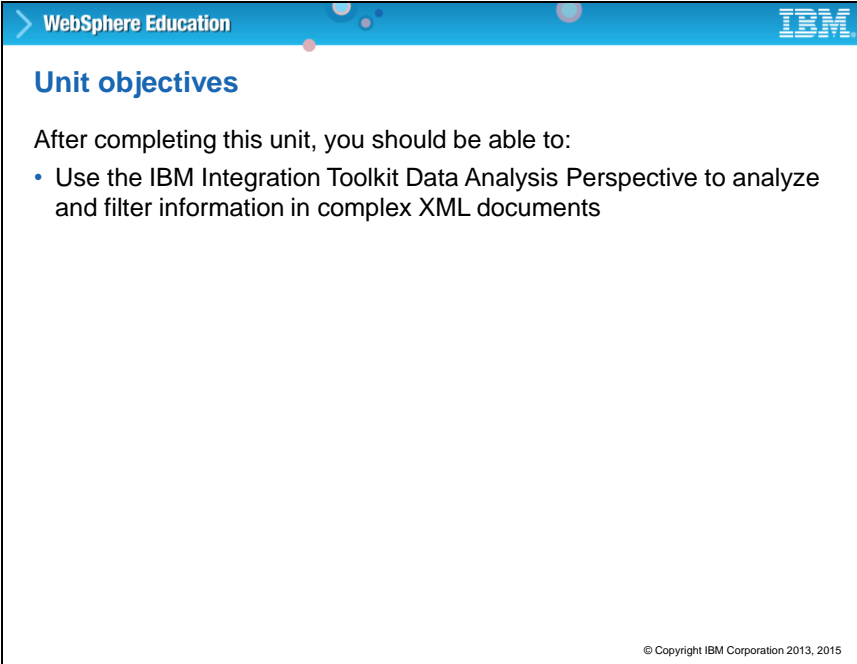
Slide 1


> WebSphere Education 

Analyzing XML documents



© Copyright IBM Corporation 2013, 2015
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

The slide features a blue header bar with the text 'WebSphere Education' on the left and the IBM logo on the right. Below the header, the title 'Unit objectives' is displayed in blue. The main content area contains a paragraph stating 'After completing this unit, you should be able to:' followed by a single bullet point. A copyright notice is located in the bottom right corner of the slide.

> WebSphere Education 

Unit objectives

After completing this unit, you should be able to:



- Use the IBM Integration Toolkit Data Analysis Perspective to analyze and filter information in complex XML documents

© Copyright IBM Corporation 2013, 2015

Unit objectives

The IBM Integration Bus Data Analysis function analyzes and filters information in complex XML documents. You can use the analysis to create a library that contains Data Analysis tools to quickly and easily transform the data in IBM Integration Bus. This unit describes the Data Analysis tools that are available in the IBM Integration Toolkit for analyzing XML data.

After completing this unit, you should be able to use the IBM Integration Toolkit Data Analysis Perspective to analyze and filter information in complex XML documents.

 WebSphere Education 

Using XML data

The problem:

- XML documents are getting bigger
- XML documents are getting more complex
- Non-descriptive XML schema definitions are poorly constructed either accidentally or deliberately in an attempt to construct more dynamic integration capabilities

The solution:

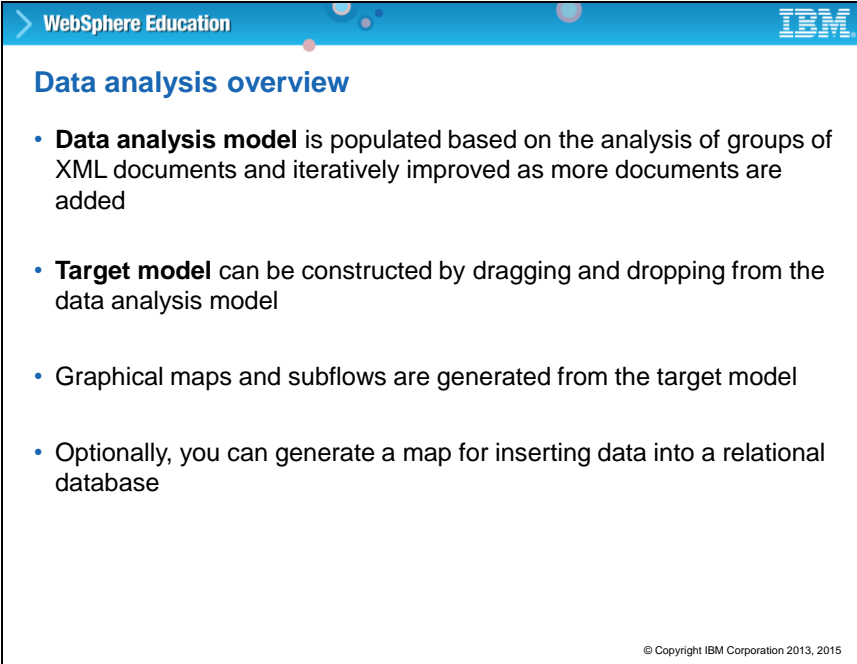
- Integration Toolkit Data Analysis Perspective provides a set of models, views, and tools that analyze and filter information in complex XML documents
- Use data analysis to model your data, and create libraries, maps, and subflows to easily transform it within the Integration Bus

© Copyright IBM Corporation 2013, 2015

Using XML data

XML files are flexible, but can be long and complex. For example, healthcare XML files can contain information for many patients, doctors, hospitals, and prescriptions. If you are working with complex XML documents, you might be interested in a small subset of data. However, it can be difficult to construct a model to filter and analyze this subset of data by traditional methods.

The Integration Toolkit Data Analysis Perspective analyzes and filters information in complex XML documents. You can use this analysis to create a library that contains data analysis tools to quickly and easily transform your complex XML data in Integration Bus.



The slide is titled "Data analysis overview" and is part of a "WebSphere Education" presentation, as indicated by the header. It features a blue header bar with the IBM logo on the right. The main content is a bulleted list of four points. The first point states that the "Data analysis model" is populated based on the analysis of groups of XML documents and is iteratively improved as more documents are added. The second point states that the "Target model" can be constructed by dragging and dropping from the data analysis model. The third point states that Graphical maps and subflows are generated from the target model. The fourth point states that optionally, you can generate a map for inserting data into a relational database. The slide is copyrighted by IBM Corporation in 2013 and 2015.

- **Data analysis model** is populated based on the analysis of groups of XML documents and iteratively improved as more documents are added
- **Target model** can be constructed by dragging and dropping from the data analysis model
- Graphical maps and subflows are generated from the target model
- Optionally, you can generate a map for inserting data into a relational database

© Copyright IBM Corporation 2013, 2015

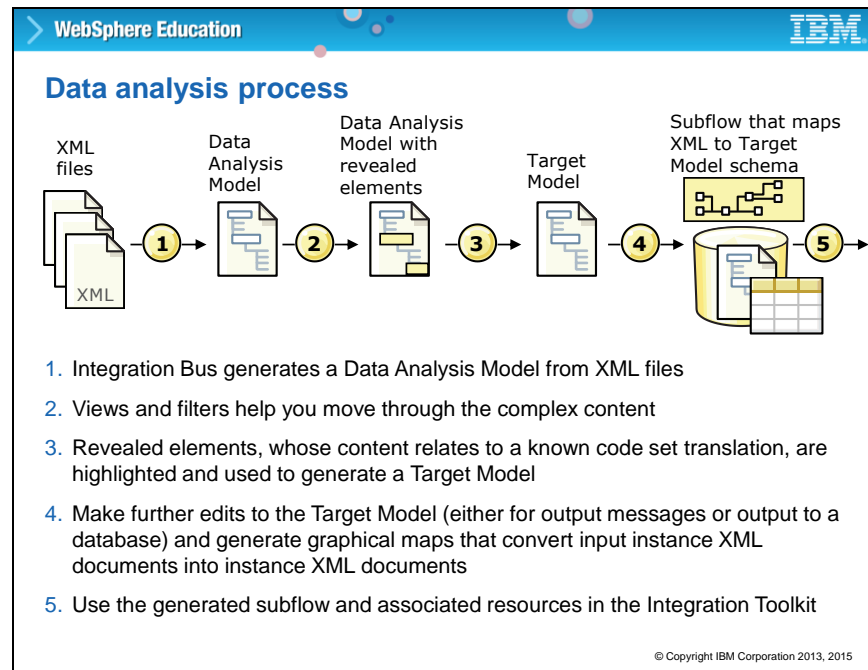
Data analysis overview

In a Data Analysis project, you analyze a set of sample XML documents according to the content of the data. This content is defined in a Data Analysis profile that you specify when you create the project.

When you analyze your sample XML documents, new data and new data structures within them are retained in the *Data analysis model*. Repeated data and data structures are not added.

After the analysis is complete, you can select relevant data structures in the Data analysis model, and then add them to a **target model**. You can use the target model to produce a library that contains your Data Analysis tools, which you can use at run time to transform incoming data.

The Data Analysis tools that are created include a map, schema file, validation stylesheet, and subflow. As an option, you can create a database representation of the target model.



Data analysis process

The Data Analysis tools are all grouped into a single Data Analysis Perspective, which consists of a set of coordinated views. Selecting an element in one view, moves you to the relevant part of another view.

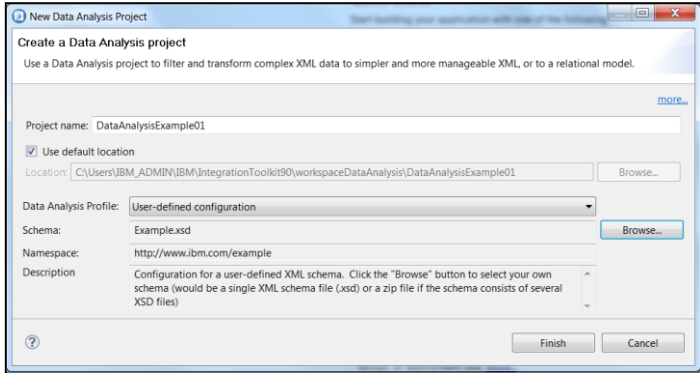
The key steps for analyzing XML documents are:

1. Create a Data Analysis project that generates a data analysis model from a representative sample of XML documents.
2. Use views and filters to analyze the sample XML documents.
3. Create a target model from revealed elements, whose content relates to a known code set conversion.
4. Populate and edit your target model and generate graphical maps that convert input instance XML documents into new XML documents.
5. Use the generated subflow and associated resources in the Integration Toolkit. You can optionally view and edit your target model in database format.

WebSphere Education

IBM

Creating a Data Analysis project



- Data Analysis projects must use a Data Analysis Profile
 - User-defined configuration Data Analysis Profile
 - IBM Integration Bus Healthcare Pack Data Analysis profiles
 - Book Series (sample configuration)

© Copyright IBM Corporation 2013, 2015

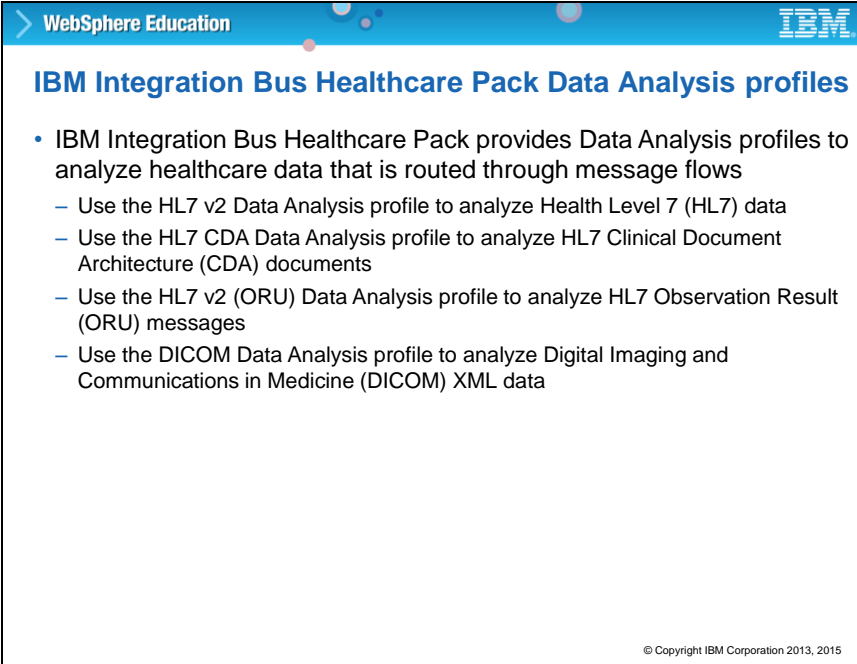
Creating a Data Analysis project

You must have a Data Analysis project to create, contain, and develop Data Analysis models, target models, and Data Analysis tools.


This slide shows the Data Analysis Project creation dialog. As with all projects, you specify a name and a location.

If you are analyzing a specific type of data when you create the project, you can specify a Data Analysis Profile. For example, the Data Analysis profiles that are provided with the Integration Bus Healthcare Pack are used to analyze specific types of healthcare data.

You can also specify a schema. The schema that you select provides a common definition against which all the instance documents successfully validate.



The slide features a blue header bar with the text 'WebSphere Education' on the left and the IBM logo on the right. Below the header, the title 'IBM Integration Bus Healthcare Pack Data Analysis profiles' is displayed in bold blue text. The main content area contains a bulleted list of four items, each preceded by a blue dash. The slide is enclosed in a thin black border, and a small copyright notice is located in the bottom right corner.

WebSphere Education 

IBM Integration Bus Healthcare Pack Data Analysis profiles

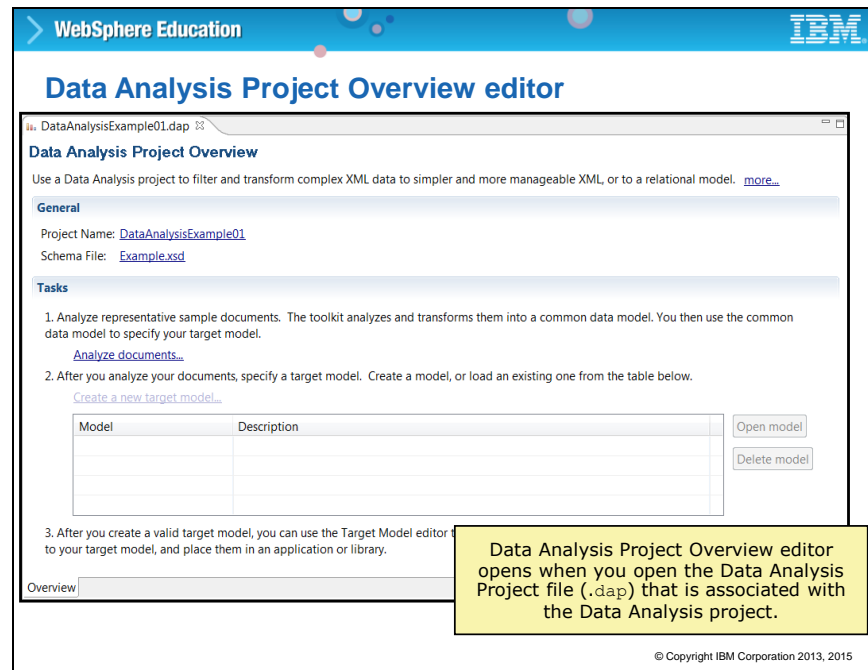
- IBM Integration Bus Healthcare Pack provides Data Analysis profiles to analyze healthcare data that is routed through message flows
 - Use the HL7 v2 Data Analysis profile to analyze Health Level 7 (HL7) data
 - Use the HL7 CDA Data Analysis profile to analyze HL7 Clinical Document Architecture (CDA) documents
 - Use the HL7 v2 (ORU) Data Analysis profile to analyze HL7 Observation Result (ORU) messages
 - Use the DICOM Data Analysis profile to analyze Digital Imaging and Communications in Medicine (DICOM) XML data

© Copyright IBM Corporation 2013, 2015

IBM Integration Bus Healthcare Pack Data Analysis profiles

You can create your own data analysis profiles or use predefined profiles.

Integration Bus Healthcare Pack provides Data Analysis profiles that are used with an Integration Bus Data Analysis project to analyze healthcare data that is routed through message flows. This slide summarizes the Integration Bus Healthcare Pack Data Analysis profiles. Each profile is used for a specific type of healthcare data.

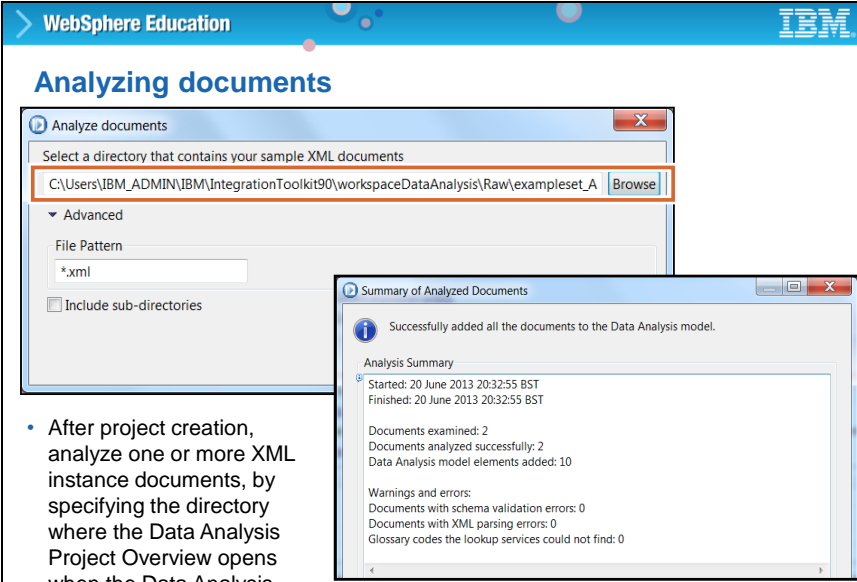


Data Analysis Projective Overview editor

Clicking **Finish** in the New Data Analysis Project dialog creates the project and opens the Data Analysis Project Overview file (.dap) in an editor view.

By default, the editor view in the Data Analysis perspective contains the Data Analysis Project Overview and Target Model editors. Use the **Data Analysis Project Overview** tab to analyze sample XML documents, and to add data structures that conform to your schema into a Data Analysis model.

To start analyzing XML documents, click **Analyze documents** in the **Data Analysis Project Overview** tab.



The screenshot displays the 'WebSphere Education' interface with the title 'Analyzing documents'. It features two overlapping windows. The 'Analyze documents' window is in the background, showing a text field with the path 'C:\Users\IBM_ADMIN\IBM\IntegrationToolkit90\workspaceDataAnalysis\Raw\exampleset_A' and a 'Browse' button. Below this, under an 'Advanced' section, is a 'File Pattern' field containing '*.xml' and an unchecked 'Include sub-directories' checkbox. The 'Summary of Analyzed Documents' window is in the foreground, displaying a message: 'Successfully added all the documents to the Data Analysis model.' It includes an 'Analysis Summary' section with the following details: Started: 20 June 2013 20:32:55 BST, Finished: 20 June 2013 20:32:55 BST, Documents examined: 2, Documents analyzed successfully: 2, and Data Analysis model elements added: 10. A 'Warnings and errors' section at the bottom shows zero errors for schema validation, XML parsing, and glossary codes.

- After project creation, analyze one or more XML instance documents, by specifying the directory where the Data Analysis Project Overview opens when the Data Analysis Project file is opened

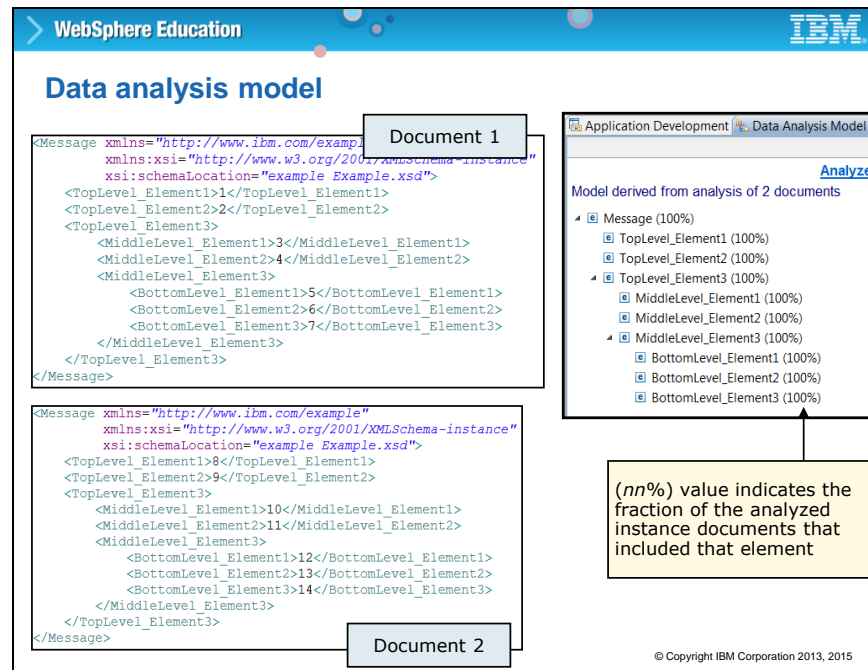
© Copyright IBM Corporation 2013, 2015

Analyzing documents

The next step is to select a directory that contains a representative sample of XML documents.

Ensure that your sample documents are in the XML format. The sample XML documents must be representative of your relevant data so that you can later produce a useful target model. If data is missing from your sample XML documents, subsequent input might not conform to the generated target model.

Also, ensure that the sample XML documents conform to a schema and that the schema is available. If the sample data does not conform to your schema, you cannot analyze your sample XML documents, or load the relevant data from within them.



Data Analysis model

The next step is to select elements and attributes from your Data Analysis model to form a target model, which you use to create a library that contains Data Analysis tools.

The data analysis example in the slide uses a single XML schema to which two separate XML instance documents conform.

The **Data Analysis Model** tab is in the upper left of the Data Analysis Perspective. You can switch between the **Data Analysis Model** tab and the **Application Development** tab.

The document instance percentage button configures the windows to show the percentages in brackets at each level in the hierarchy of the Data Analysis Model. The percentages indicates the percentage of the instance documents that were analyzed that include that element.

WebSphere Education **IBM**

Data Filter view

Filter (by Element Name or Element Data): 1

Element Name	Element Data	Element Type	Depth of Descendant [min, max]	Docun
TopLevel_Element1	TopLevel_Element1	string	[0,0]	3
MiddleLevel_Element1	MiddleLevel_Element1	string	[0,0]	2
BottomLevel_Element1	BottomLevel_Element1	string	[0,0]	2

Use the **Data Filter** view to search the Data Analysis model

Select a row in the **Data Filter** view to shows its **Properties** and **Data Path**

Properties Project Explorer

Property	Value
General	
Data Path	
Element Data	MiddleLevel_Element1
Element Name	MiddleLevel_Element1
Element Namespace URI	http://www.ibm.com/example
Statistics	
Document count	2
Total count	2
Type	
XML Schema Type	string
XML Schema Type Namespace	http://www.w3.org/2001/XMLSchema

Data Paths Problems

MiddleLevel_Element1 (Message/TopLevel_Element3/)


© Copyright IBM Corporation 2013, 2015

Data Filter view

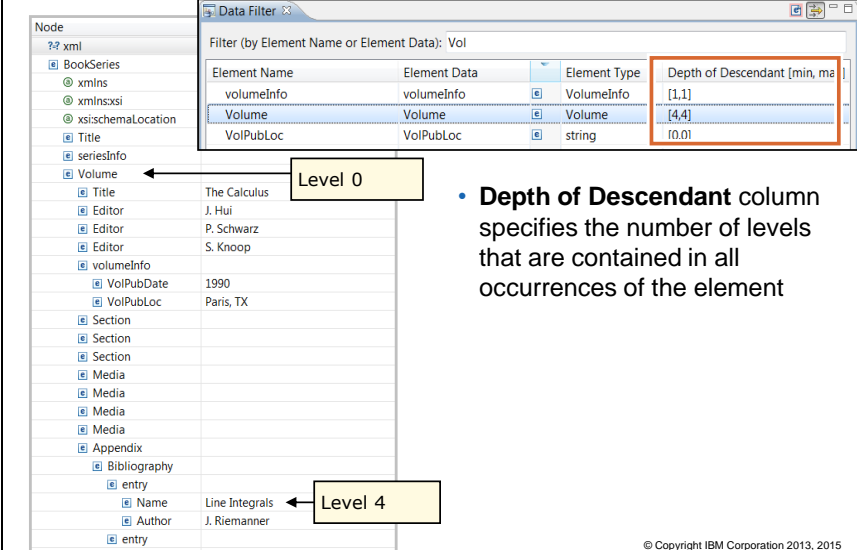
The **Data Filter** view shows the elements that are added from the sample XML documents. The elements are organized according to static cues in the structure and are listed alphabetically by **Element Name**.

You can filter the results and sort the order according to **Element Name** or **Element Data**.

When you click an element in the **Data Filter** view, the **Properties** and the **Data Paths** views update to show details of the selected element.

WebSphere Education 

Depth of descendants



Element Name	Element Data	Element Type	Depth of Descendant [min, max]
volumeInfo	volumeInfo	VolumeInfo	[1,1]
Volume	Volume	Volume	[4,4]
VolPubLoc	VolPubLoc	string	[0,0]

- **Depth of Descendant** column specifies the number of levels that are contained in all occurrences of the element

© Copyright IBM Corporation 2013, 2015

Depth of descendants

The **Depth of Descendant** column specifies the number of levels of a descendant for each of its occurrences in the model.

In this example, all examples of the element that is named **Volume** have four levels of descendants. So, the minimum and maximum values of the **Depth of Descendant** column are the same. If some occurrences of the **Volume** element had fewer descendants, you would see different values for the minimum and maximum.

The screenshot displays the 'Data Analysis Model' window in WebSphere Education. The model is derived from 4 documents. The tree structure shows the following elements and their coexistence percentages:

- Message (100%) [100%]
 - TopLevel_Element1 (75%) [66.7%]
 - TopLevel_Element2 (75%) [100%]** (Selected)
 - TopLevel_Element3 (50%) [66.7%]
 - MiddleLevel_Element1 (50%) [66.7%]
 - MiddleLevel_Element2 (50%) [66.7%]
 - MiddleLevel_Element3 (50%) [66.7%]
 - BottomLevel_Element1 (50%) [66.7%]
 - BottomLevel_Element2 (50%) [66.7%]
 - BottomLevel_Element3 (50%) [66.7%]

Annotations and actions shown:

- Click % icon to show percentages.** (Points to the percentage icon in the toolbar)
- 66.7% of the instance documents that contain TopLevel_Element2 also contain TopLevel_Element1.** (Points to the [66.7%] value for TopLevel_Element1)
- Right-click an element in the Data Analysis Model to access highlighting options** (Points to the right-click context menu)
- Highlight All Coexisting Elements** (Option in the context menu)
- Click Highlight All Coexisting Elements to show the percentage of the instance documents that contain the selected element that also includes this element** (Points to the selected menu option)

Color legend [%]: 0% [0%, 33%] [33%, 66%] [66%, 100%]

Focus Element: **TopLevel_Element2**

© Copyright IBM Corporation 2013, 2015

Highlight all co-existing elements

In this example, a fourth instance of an XML document was added to the analysis.

You can click **Highlight All Coexisting Elements** to highlight all elements that are in one or more of the same sample XML documents as your selected element. This action shows where combinations of data structures occur.

The selected element is highlighted in the Data Analysis Model view, and is also identified at the bottom of the view, as the **Focus Element**. The percentage next to each element indicates the percentage of documents that contain both that element and the focus element.

WebSphere Education **IBM**

XML Glossary File Lookup Service

Preferences

type filter text

- Integration Development
 - Build Broker Archive
 - Data Analysis
 - Diagnostic Logging
 - Glossary File Lookup Services
 - Delimited Glossary File Lookup Service
 - XML Glossary File Lookup Service**

XML Glossary File Lookup Service

File Location/Contributor Plugin ID	Glossary Reference	Description
com.ibm.hcl.sdg.bookseries	BookCodes	Book Series Terminology File

View Content

```
<?xml version="1.0" encoding="UTF-8"?>
<codeSet>
  <code>
    <codeSystem>BookCodes</codeSystem>
    <entry>
      <keyValue name='secType'>MONO</keyValue>
      <displayName>Monograph</displayName>
    </entry>
    <entry>
      <keyValue name='secType'>SURV</keyValue>
      <displayName>Survey</displayName>
    </entry>
    <entry>
      <keyValue name='mediaType'>DVD</keyValue>
      <displayName>Optical Disc</displayName>
    </entry>
    <entry>
      <keyValue name='mediaType'>SW</keyValue>
      <displayName>Software</displayName>
    </entry>
  </code>
</codeSet>
```

© Copyright IBM Corporation 2013, 2015

- Uses glossary files to transform the data element code in the sample XML documents into more readable names
- Accessed by clicking **Window > Preferences > Integration Development > Data Analysis > Glossary File Lookup Service**

XML Glossary File Lookup Service

Data element code in your sample XML documents can be difficult to understand because it often consists of strings or numbers rather than descriptive words. Glossary files contain codes for each data element and their equivalent names. The Glossary File Lookup Service retrieves the names, and shows them instead of the data element codes in your Data Analysis project.

The Glossary File Lookup Service makes it easier to use Data Analysis because it is simpler to understand the names than their equivalent data element codes.

WebSphere Education IBM

Refining a target model

When you finish editing the target model, you can generate the mapping and workflow artifacts to be deployed to the Broker. [Generate...](#)

MyTargetModel		
Editors		
Editor		
Volume		[1..1]
Title		[0..*]
Video Disc		[0..1]
mediaType		[0..1]
Title		[1..1]
VolPubDate		
AdditionalTrailer		
SummaryStatement		[0..1]

Element symbols with the shortcut arrow in the lower left corner originate directly from the Data Analysis Model

You can add your own elements to the target model


© Copyright IBM Corporation 2013, 2015

Refining a target model

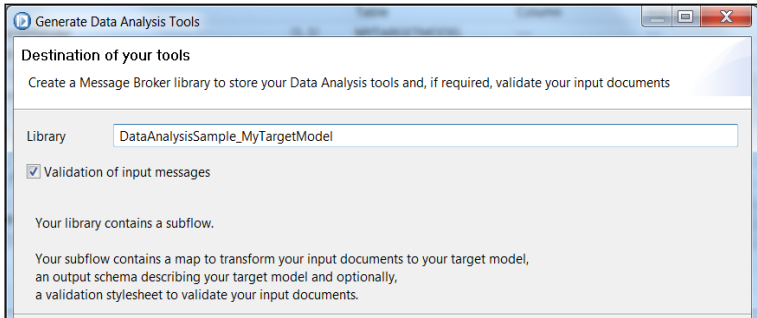
You can use the Target Model editor to alter the structure of your target model to meet the requirements of the downstream application.

With the Target Model editor you can:

- Create local elements.
- Rename elements and attributes.
- Move elements
- Delete elements and attributes.
- Restore deleted elements and attributes.

WebSphere Education 

Creating data analysis tools



- Use the target model to create a library, which contains the following Data Analysis tools:
 - Schemas
 - Maps
 - Subflows
 - Validation stylesheets
- In the Target Model editor, click **Generate** to open the Generate Data Analysis Tools wizard

© Copyright IBM Corporation 2013, 2015

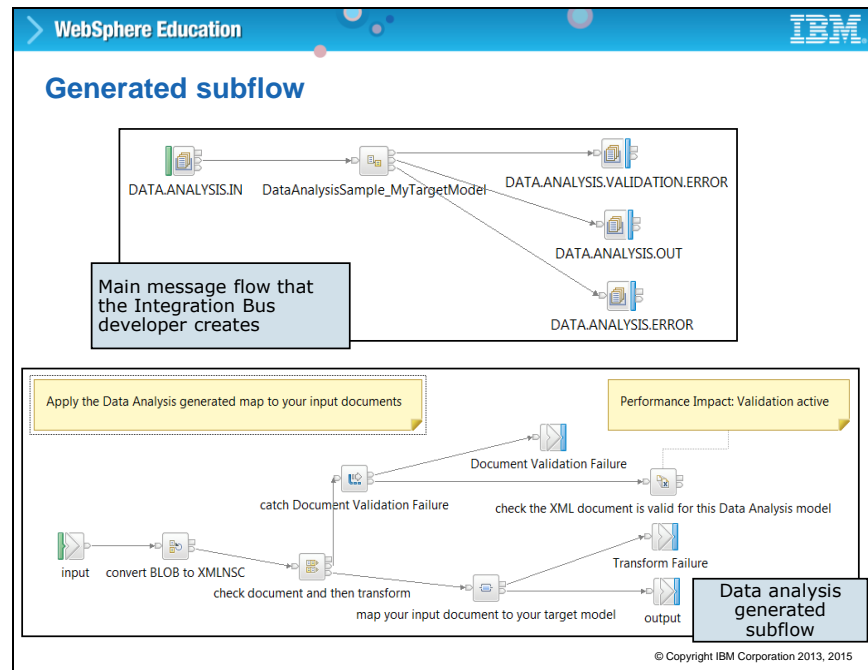
Creating data analysis tools

You can use your target model to create a library, which contains the following Data Analysis tools:

- Schemas
- Maps
- Subflows
- Validation stylesheets

To generate the Data Analysis tools, click **Generate** in the Target Model editor.


A library that contains your Data Analysis tools is created. The subflow opens in the Message Flow editor.



Generated subflow

One of the Data Analysis tools is a subflow that can be called from a message flow to validate and transform a message against the target model.

The subflow first converts the incoming message to XML. If the message is valid, it then maps the message to your target model. The subflow also includes some error handling for identifying validation failures and writing them to a queue.

WebSphere Education 

Target model: Relational database map

When you finish editing the target model, you can generate the mapping and workflow artifacts to be deployed to the Broker. [Generate...](#)

		Table	Column	SQL Data Type
MyTargetModel	[1..1]	MYTARGETMODEL	---	---
Editors	[0..1]	EDITORS	---	---
Editor	[1..*]	EDITOR	EDITOR	VARCHAR(256)
Volume	[1..*]	VOLUME	---	---
Title	[1..1]	---	TITLE	VARCHAR(256)
Video Disc	[0..*]	VIDEO_DISC	---	---
mediaType	[0..1]	---	MEDIATYPE	VARCHAR(256)
Title	[1..1]	---	TITLE	VARCHAR(256)
VolPubDate	[1..1]	---	TITLE_VOLPUBDATE	VARCHAR(256)
AdditionalTrailer	[0..1]	ADDITIONALTRAILER	---	---
SummaryStatement	[0..1]	SUMMARYSTATEMENT	---	---

To view the target model in a database format:

1. Click the **Enable relational database mapping** icon.
2. Complete the **Destination Database** and **Destination Schema** fields, and then select the product and version of your database.
3. To generate the default SQL table names for the target database, click the **Relational mapping options** icon and then click **Generate Default SQL Table Naming**.

© Copyright IBM Corporation 2013, 2015

Target model: Relational database map

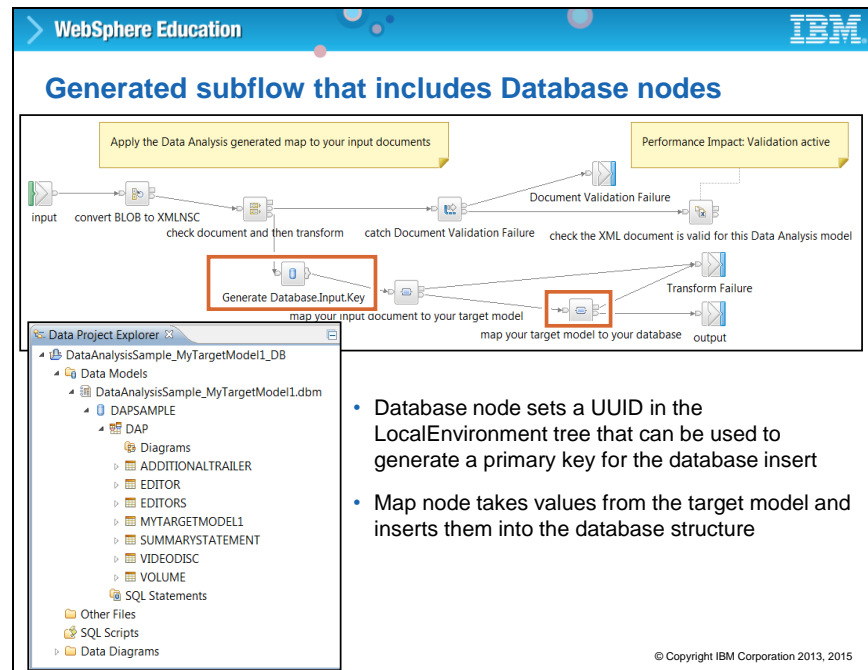
As an option, you can view your target model in a database format.

The slide shows the steps for viewing the target model in a database format.

1. Click the **Enable relational database** mapping icon. The Destination Database Platform Selection wizard opens.
2. Complete the **Destination Database** and **Destination Schema** fields. Select the product and version of your database. Click **OK**. The blank **Table**, **Column**, and **SQL Data Type** columns are shown.
3. To generate the default SQL table names for your target database, click the **Relational mapping options** icon and then click **Generate Default SQL Table Naming**.

The Target Model editor is populated with the details of your target model, in a database format. In the populated model, you can:

- Rename tables and columns.
- Assign different element types to elements.
- Delete cell contents.



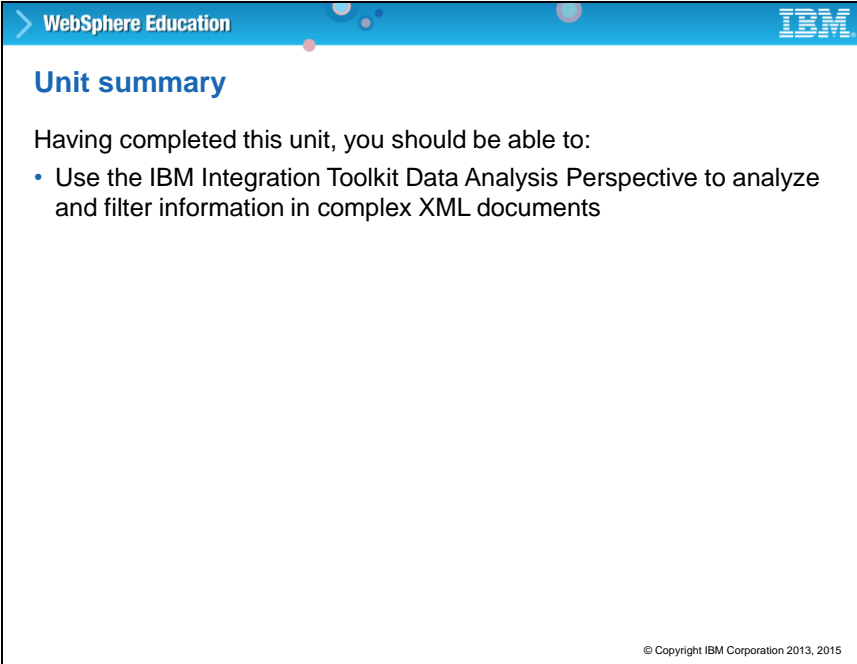
Generated subflow that includes Database nodes

You can use Integration Bus to insert the target model information into an existing database.

The Data Analysis perspective generates a subflow that maps the target model to the database. The message flow contains a Map node that sets a UUID in the LocalEnvironment tree. The UUID can be used to generate a primary key for the database. Another Map node inserts the target model into the database structure.

Integration Bus currently supports DB2, Oracle, and Microsoft SQL Server for a Data Analysis database. The IBM Knowledge Center for IBM Integration Bus contains instructions for setting up a Data Analysis database.

For a video demonstration of Data Analysis, see the IBM Integration Bus V9 Data Analysis videos on YouTube.



The slide features a blue header bar with the text 'WebSphere Education' on the left and the IBM logo on the right. Below the header, the title 'Unit summary' is displayed in blue. The main content area contains a paragraph stating 'Having completed this unit, you should be able to:' followed by a bulleted list with one item: 'Use the IBM Integration Toolkit Data Analysis Perspective to analyze and filter information in complex XML documents'. A small copyright notice '© Copyright IBM Corporation 2013, 2015' is located in the bottom right corner of the slide.

> WebSphere Education IBM

Unit summary

Having completed this unit, you should be able to:

- Use the IBM Integration Toolkit Data Analysis Perspective to analyze and filter information in complex XML documents

© Copyright IBM Corporation 2013, 2015

Unit summary

The IBM Integration Bus Data Analysis function analyzes and filters information in complex XML documents. You can use the analysis to create a library that contains Data Analysis tools to quickly and easily transform the data in IBM Integration Bus. This unit described the Data Analysis tools that are available in the IBM Integration Toolkit for analyzing XML data.

Having completed this unit, you should be able to use the IBM Integration Toolkit Data Analysis Perspective to analyze and filter information in complex XML documents.