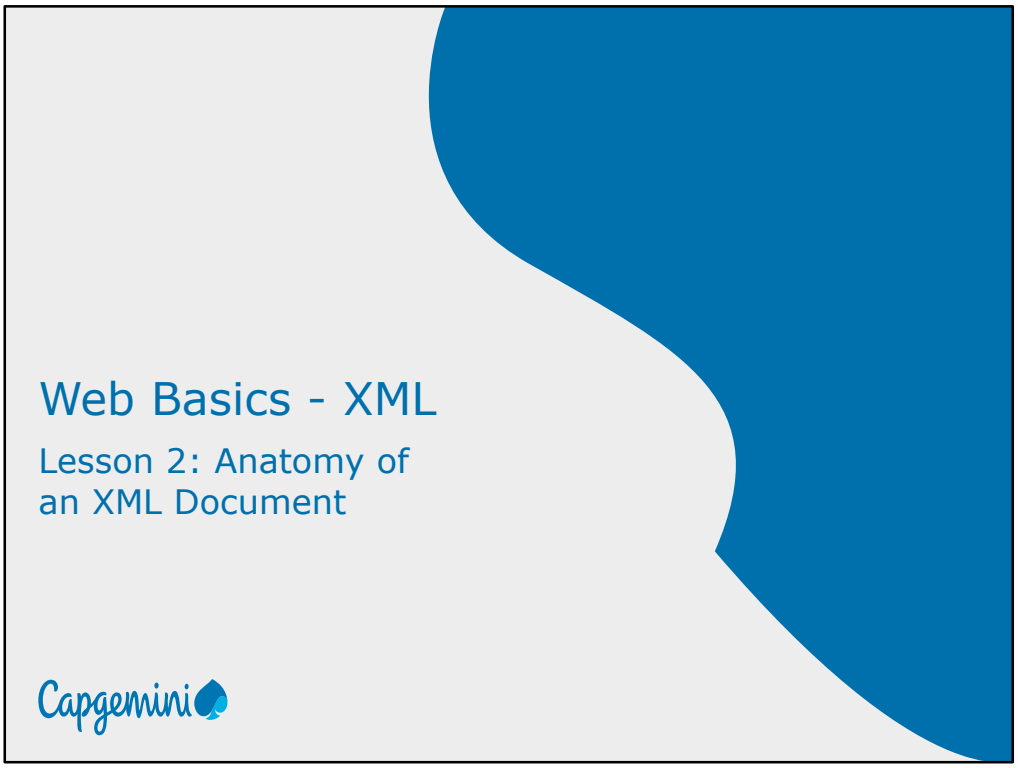
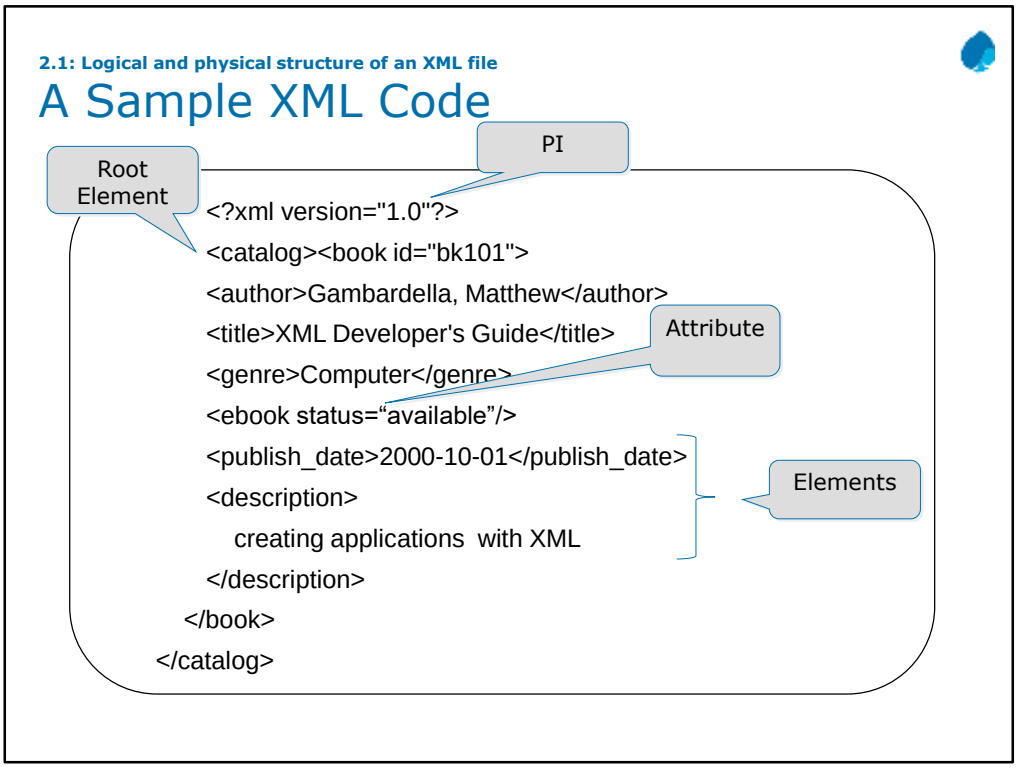

Web Basics - XML
Lesson 2: Anatomy of
an XML Document
Capgemini

Lesson Objectives

- In this lesson, you will learn:
  - Logical and physical structure of an XML file
  - Parts of XML file like:
    - Elements
    - Attributes
    - Entities
    - Processing instructions



Elements can be simple, empty, mixed
Simple : <data>value</data>
Mixed : <data> value
<sub>sub1</sub1>
</data>
Empty : <data></data>

2.1 Logical and physical structure of an XML file

Understanding the Sample XML Code

- Let us now understand the different parts of the XML file:
- XML Declaration
- Root Element
- An Empty Element
- Attributes

Understanding the Sample XML Code:
XML Declaration:
It is a processing instruction (identified by the ? at its start and end).
Root Element:
Each XML document must have only one root element, all the other elements must be completely enclosed in that element.
Line 2 (in example) identifies the start element (the start tag), and line 12 identifies the end of the element (the end tag).
Empty Elements:
Empty elements have no content and are marked up as either of the following:
<empty_element/>
<empty_element></empty_element>

Attribute Markup:
Attributes are used to attach information to the information contained in an element. The general form for using an attribute is as follows:
<element-name property=”value”>

2.2 Parts of XML file

Using XML Markup

- Tags carry the smallest unit of meaning signifying structure, format, or style of the data
- They are always enclosed within angled brackets, that is "< >". Tags are case-sensitive
- The tags <book>,<Book>, and <BOOK> carry different meanings and cannot be used interchangeably
- All tags must be paired so that they have a start <book> and an end </book>
- Tags when combined with data form elements

Using XML Markup:

XML is concerned with element markup. Instead of XML's tags being markers that indicate where a style should change or a new line should begin, XML's element markup is composed of three parts:

The start tag
The content
The end tag

Elements:

Elements contain information or content and can also contain other elements.
There is one element that contains all the other elements called the "root element".
Tags show the beginning and end of an element.
XML documents are divided into containers called "elements". People who are familiar with HTML, know that …. , <form> … </form>,
 are all elements.

2.2 Parts of XML file

Using XML Markup

- Attribute Markup:
  - It is used to attach information to the information contained in an element.
  - General form for using an attribute is as follows:
  - <element-name property="value">
  - An attribute value must be enclosed in quotation marks.
  - You can either use single quote or double quote. However, you cannot mix the two in the same specification.

Using XML Markup:
Attributes:

Attributes are element modifiers. They provide additional and more specific information about an element and its content.

Normally in HTML, attributes are used most often to provide the browser with a suggestion for formatting the display of the elements content by a browser.

For example: bgcolor attribute of <body> element or align attribute normally are used with almost all elements.

However, the same is not true with XML. The attributes are used to provide further information about the element itself. This is because the main purpose of XML is to separate markup from display, so you will rarely see formatting attributes in XML DTDs.

2.2 Parts of XML file

Using XML Markup

- Naming Rules:
  - A name consists of at least one letter: a to z or A to Z
  - If the name consists of more than one character, then it may start with an underscore (_) or a colon (:)
  - The initial letter can be followed by one or more letters, digits, hyphens, underscores, or full stops

Using XML Markup:
Naming Rules:

A name consists of at least one letter: a to z or A to Z.

If the name consists of more than one character, then it may start with an underscore (_) or a colon (:)

The initial letter can be followed by one or more letters, digits, hyphens, underscores, and full stops.

2.2 Parts of XML file

Using XML Markup

- Comments:
- Comments have the following form:
- <!- -This is comment text - ->
- Use the comment start tag and end tag correctly.
- Everything in the comment text will be completely ignored by the XML processor
- Following comment is therefore quite safe:
- <! - - These are the declaration for the <title> and <body> - ->

Using XML Markup:
Comments

In keeping with the design constraint of keeping XML simple, its comment facilities are also simple. Comments have the following form:

<!- -This is comment text - ->

Provided that you use the comment start tag and end tag correctly, everything in the comment text will be completely ignored by the XML processor. The following comment is therefore quite safe:

<! - - These are the declaration for the <title> and <body> - ->

There is only one restriction on what you can place in your comment text: the string - - is not allowed. This keeps XML backward compatible with SGML.

2.2 Parts of XML file

Using XML Markup

- Predefined Entities:

Character	Replacement
&	& or & #38
‘	' or &#39
>	> or &#62
<	< or < #60
“	" or &#34

Using XML Markup:
Predefined Entities:

The special characters for quote ("), apostrophe ('), less-than (<), greater-than (>), and ampersand (&) are used for punctuation in XML, and are represented with predefined entities: ", ', <, >, and &.

Notice that the semicolon is part of the entity. You cannot use “<“ or “&” in attributes or elements.

2.3 Well-formed XML

A Well-formed XML document

- A well-formed XML document simply includes markup pages with descriptive tags
- A well-formed XML does not need a DTD, but should conform to XML syntax
- If all tags are correctly formed and follow XML guidelines, then the document is a well-formed XML

The XML syntax is discussed on the next slide.

2.3 Well-formed XML

Syntax Rules for XML

- An XML document
  - Is case sensitive
  - Has a single root element
  - Has all matching tags
  - XML Elements should be properly nested
  - All attributes are quoted
  - White spaces are not ignored
  - May or may not have a (DTD) Document Type Description to describe the document

Demo

- A sample XML Document:
  - Example1: Note.xml
  - Example2: Greeting.xml
  - Example3:musicians.xml

Summary

- In this lesson, you have learnt the following:
  - XML has specific naming rules which describes names you can use for its markup objects, that is elements

Review Question

- Question 1: XML document must have one ___.
- Question 2: A comment in XML document is written as:
  - Option 1: <!-- ... -->
  - Option 2: /*.....*/
  - Option 3: //
- Question 3: ___ are storage units in the XML document.