

What Is XML?

XML

What Is XML?

- XML is a **standard** for creating custom **markup languages** for text documents.
- XML documents are made up entirely of text.
- Regular document text is surrounded by text-based tags that are known as **markup**.
- The markup reflects the structure and meaning of the content that it surrounds.

1.1: An XML Customer Document

```
<customer>  
  <number>12345</number>  
  <name>Organic Automotive Design</name>  
  <address>  
    <street>1234 Main Street</street>  
    <street>Suite 200</street>  
    <city>Glenview</city>  
    <state>IL</state>  
    <zip>60025</zip>  
  </address>  
</customer>
```

1.2: An XML Article Document

```
<article>
  <title>About Xampul</title>
  <section>
    <paragraph>Having long been one of the country's
    leading model makers, Xampul makes detailed scale
    models of nearly any man-made object. This includes
    models of consumer and industrial products, individual
    structures and complete building developments.</paragraph>

    <paragraph>Xampul's models are used by product designers,
    architects, engineers, urban planners, park planners and others
    to help communicate their design ideas to marketing departments,
    investors, executives, prospective buyers, city councils, citizens
    and other important decision makers.</paragraph>

    <paragraph>More information regarding Xampul can be found
    by visiting <link url="http://xampul.ligent.net">
    the Xampul Web site</link>.</paragraph>
  </section>
</article>
```

The History of Markup

Proprietary Publishing Programs

- Examples include Word, Publisher, PageMaker, FrameMaker, QuarkXPress.
- Documents are both input and output with proprietary programs.
- Content text is surrounded by proprietary markup.
- Markup reflects the formatting intent of the author.
- Authoring is done in using a WYSIWYG interface.
- Content cannot be easily re-used.
- Proprietary tools cannot be easily substituted for each other.

SGML

- Invented in the 1970's by IBM.
- Is a powerful standard for creating custom markup languages.
- Uses text-based tags to surround content text.
- Documents can be input using standard text editors.
- Markup reflects the meaning a structure of the content.
- Allows for content reuse by separating content from presentation.
- Documents can be transformed into their final presentation form.

SGML (last)

- Adopted by computer manufacturers, aerospace companies, book publishers and the Federal Government.
- Was implemented in only a few software tools that were expensive.
- Was never adopted by average users.
- Was created before the World Wide Web.

HTML

- Is the markup language used to create Web pages for a browser.
- Invented in the early 1990's by CERN, the European Particle Physics Laboratory.
- Uses text-based tags to surround content text.
- Documents can be input using standard text editors.
- Markup is mostly presentation-oriented with just a bit of structure.
- Documents can not be easily reused because content is mixed with presentation.

HTML (last)

- Was widely adopted because it was simple and easy to implement.
- Is not useful for non-publishing applications like data interchange.

XML

- Invented in 1998 by frustrated SGML users.
- Allows for the creation of custom markup languages.
- Uses text-based tags to surround content text.
- Documents can be input using standard text editors.
- Markup reflects the structure and meaning of the content.
- Documents can be transformed into their final presentation form.
- Documents may be easily reused.
- Interoperates with other Web standards.

XML (last)

- Is widely supported by a large number of affordable software tools.
- Has been widely adopted by publishers of every size and kind.
- Has also been widely adopted for use in data interchange.

Related Standards and Tools

CSS

- Stands for Cascading Style Sheets.
- Was introduced as a standard means of styling HTML pages.
- Can be used to do primitive styling of XML documents directly in the browser.
- Has not quite reached industrial strength as an XML styling tool.

XSLT

- Stands for Extensible Stylesheet Language: Transformations.
- Is the language used to transform XML content documents into a form that is suitable for presentation.
- Is typically used to transform XML content documents into XHTML or XSL-FO.
- Is an XML application itself.
- Is a complete non-procedural programming language.

XHTML

- Is the newest version of HTML that follows XML syntax rules.
- Is one of the target markup languages that is created during an XSLT transformation of an XML content document.
- Is rendered into a Web page by the browser.
- Is the best choice for standards-based Web page development.

XSL-FO

- Is a markup language used to format high quality print documents.
- Is one of the target markup languages that is created during an XSLT transformation of an XML content document.
- Is rendered into a print document by an XSL-FO print formatter.

Other Related Standards

- **XQUERY** provides a means for retrieving XML documents and their parts.
- **XPATH** is an expression language used in XSLT and XQUERY.
- **XLink** provides a means of linking XML documents.
- **XPointer** provides a means of pointing to parts of XML documents.
- **DOM** and **SAX** are processing models upon which XML processing programs are based.

Blank Slide

How is XML Used?

For Data Interchange

2.1: An XML Document for Data Interchange

```
<order>
  <customerNumber>12345</customerNumber>
  <customerName>Organic Automotive Design</customerName>
  <shiptoAddress>
    <street>1234 Main Street</street>
    <street>Suite 200</street>
    <city>Glenview</city>
    <state>IL</state>
    <zip>60025</zip>
  </shiptoAddress>
  <items>
    <item>
      <itemCode>4588A</itemCode>
      <quantity>1</quantity>
      <unitPrice>2500.00</unitPrice>
    </item>
    <item>
      <itemCode>3349B</itemCode>
      <quantity>2</quantity>
      <unitPrice>350.00</unitPrice>
    </item>
  </items>
</order>
```

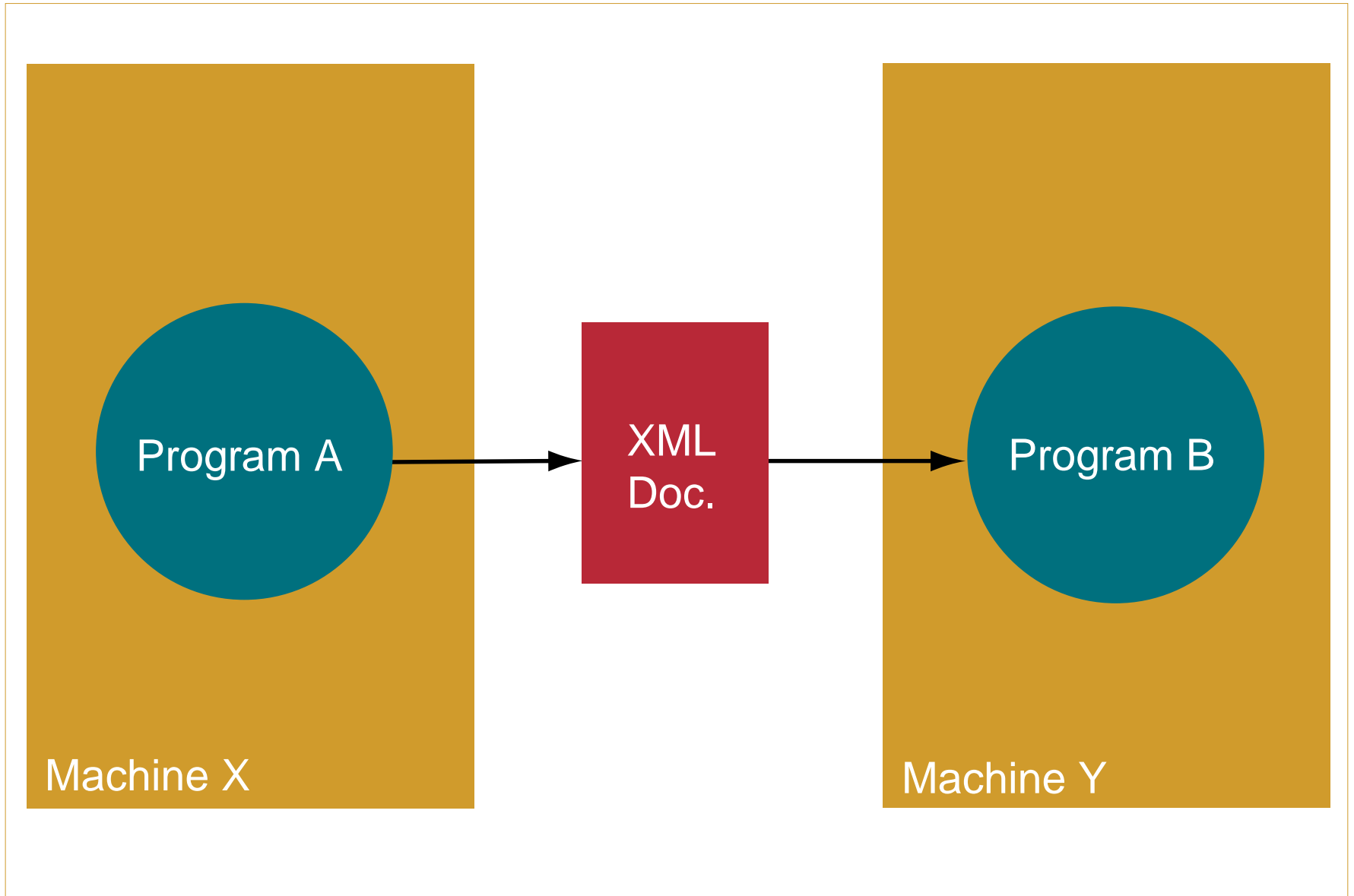
2.1: An XML Document for Data Interchange (last)

```
    </item>  
  </items>  
</order>
```

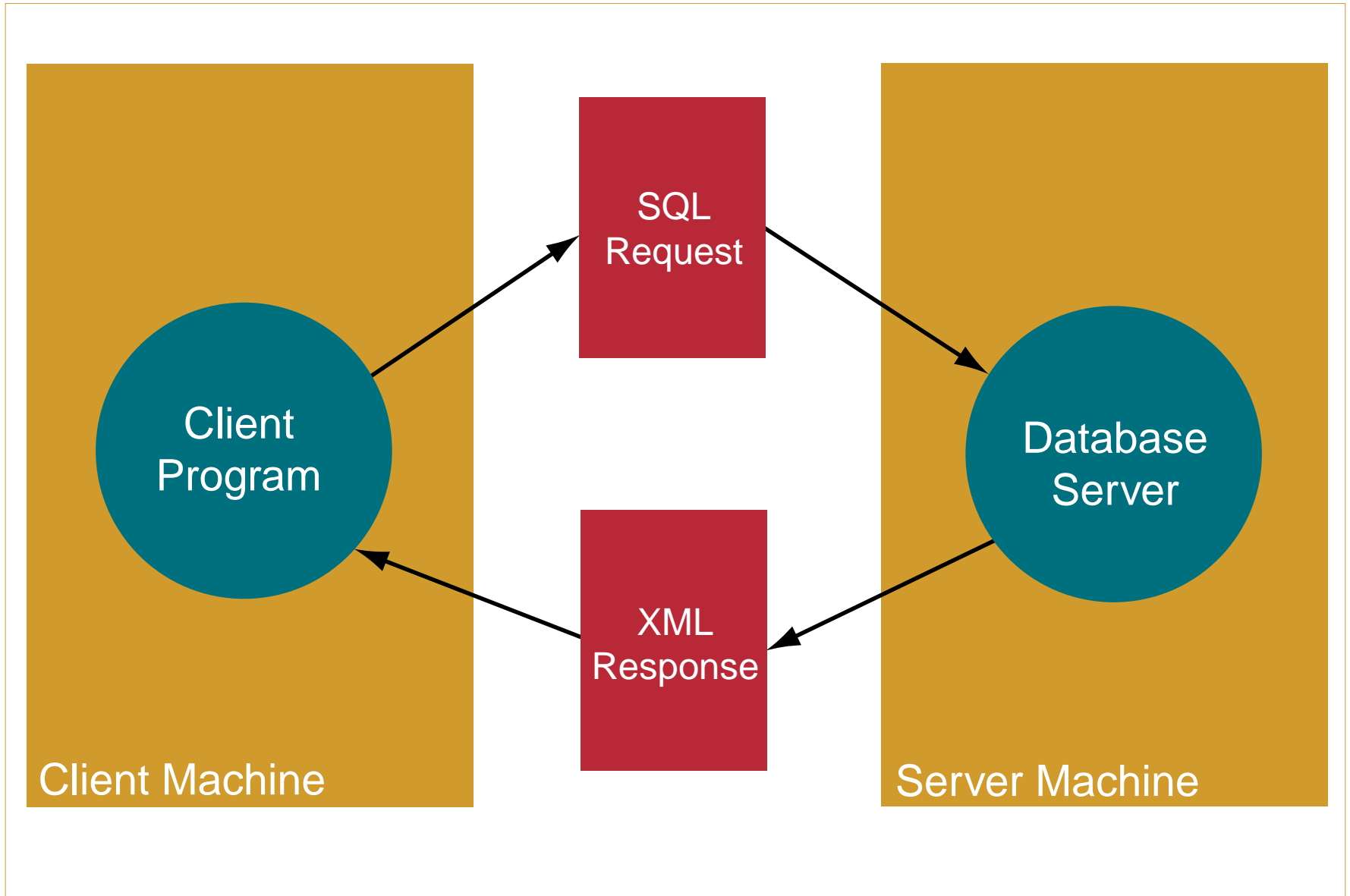

Typical Data Interchange Use Cases

- File Interchange
- Database serving
- **Web Services**
- Lightweight data entry

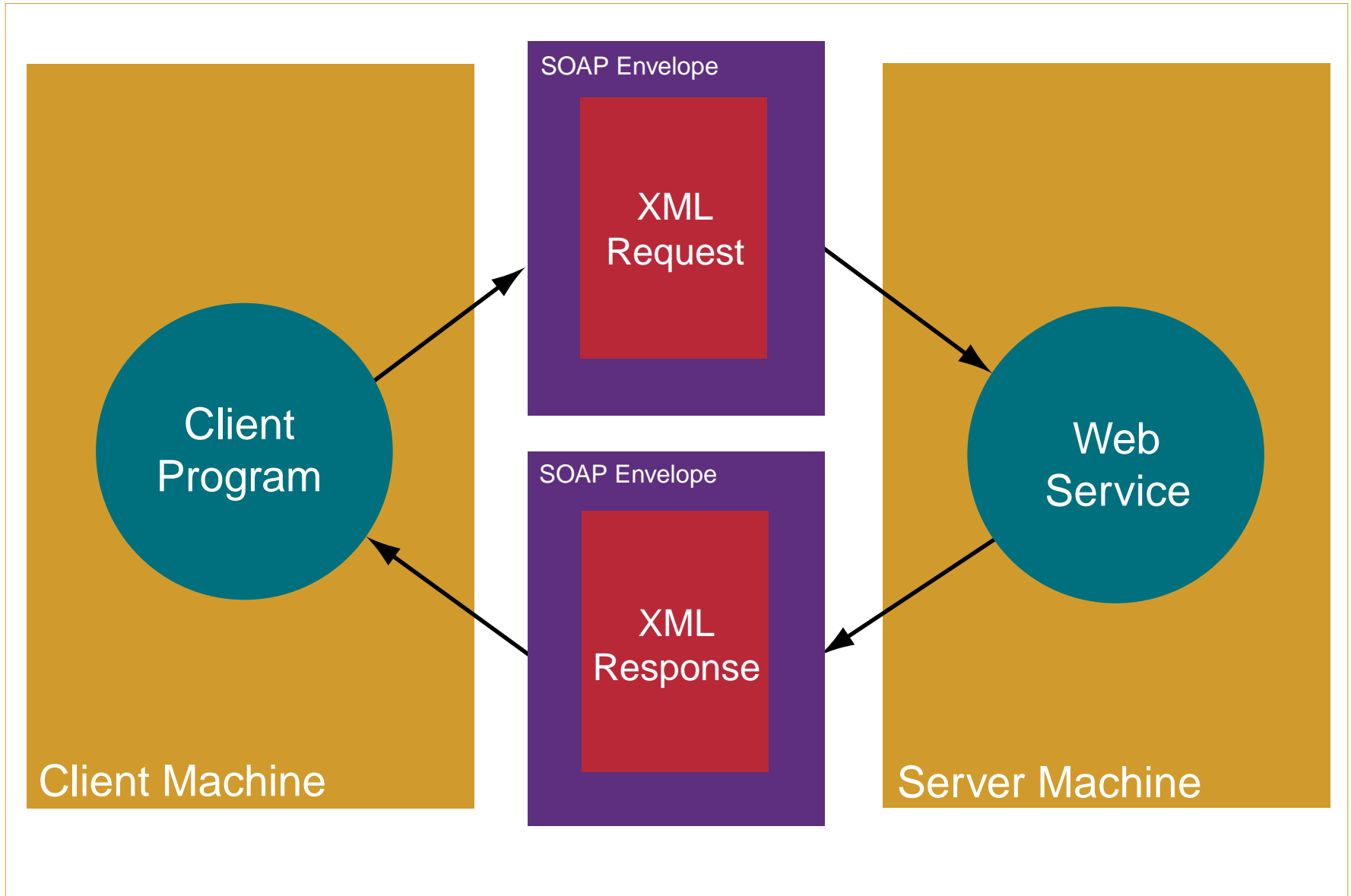
File Interchange



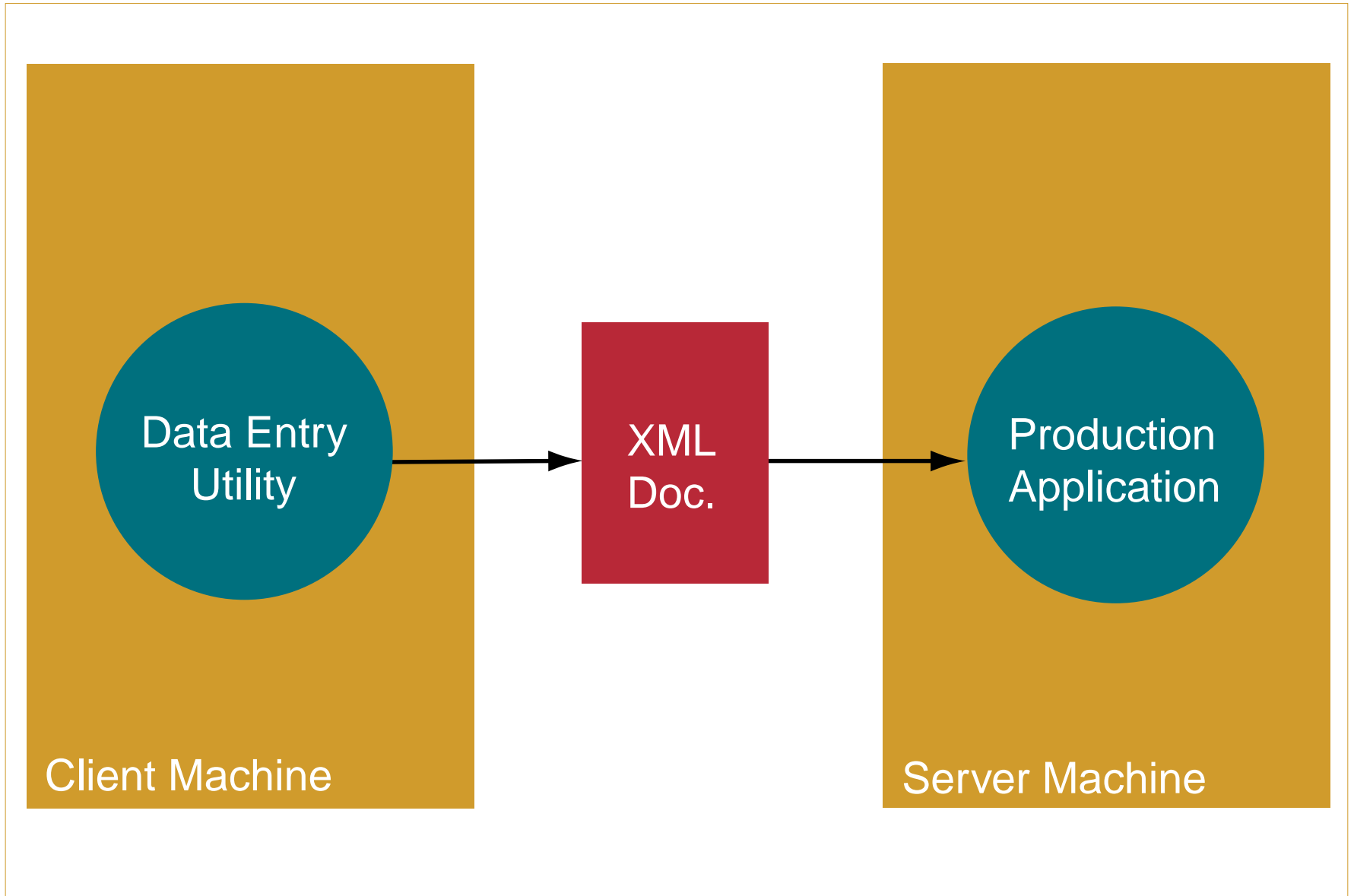
Database Serving



Web Services



Lightweight Data Entry



XML's Data Interchange Advantage

- Documents can be processed on any machine that supports XML.
- Documents can be transported over standard Web protocols like **HTTP**.
- System components can be implemented on **heterogeneous** computing platforms.
- Components can be built using standard XML tools like **parsers** and **XSLT**.
- Components can be located and used flexibly using **Service-Oriented Architecture (SOA)**.

For Publishing

2.2: An XML Document for Publishing

```
<pressrelease>
  <headline>Xavier Ampul III Named Chairman of Xampul</headline>
  <summary>Xampul announced today that Xavier Ampul III, the company
  CEO, has been named to the post of Chairman of the Board.</summary>
  <location>Glenview</location>
  <date>21 March, 2003</date>
  <detail>
    <section>
      <paragraph>
        <emphasis>Xampul</emphasis>, one of the nation's leading
        model-making companies, announced today that Xavier Ampul
        III has been named to the post of Chairman of the Board.
        Mr. Ampul has been CEO of Xampul since October of 1998
        and he will continue to hold that post as well. He succeeds
        his father, Xavier Ampul Jr., who is retiring.</paragraph>
      </section>
      <section>
        <title>About Xampul</title>
        <paragraph>Having long been one of the country's leading
        model makers, Xampul makes detailed scale models of
```


2.2: An XML Document for Publishing (last)

nearly any man-made object. More information regarding Xampul is available at the following web site:

<website><http://www.ligent.net/xampul/></website>.

</paragraph>

</section>

</detail>

<contact-name>Anna Weber</contact-name>

<contact-telephone>555-555-5555</contact-telephone>

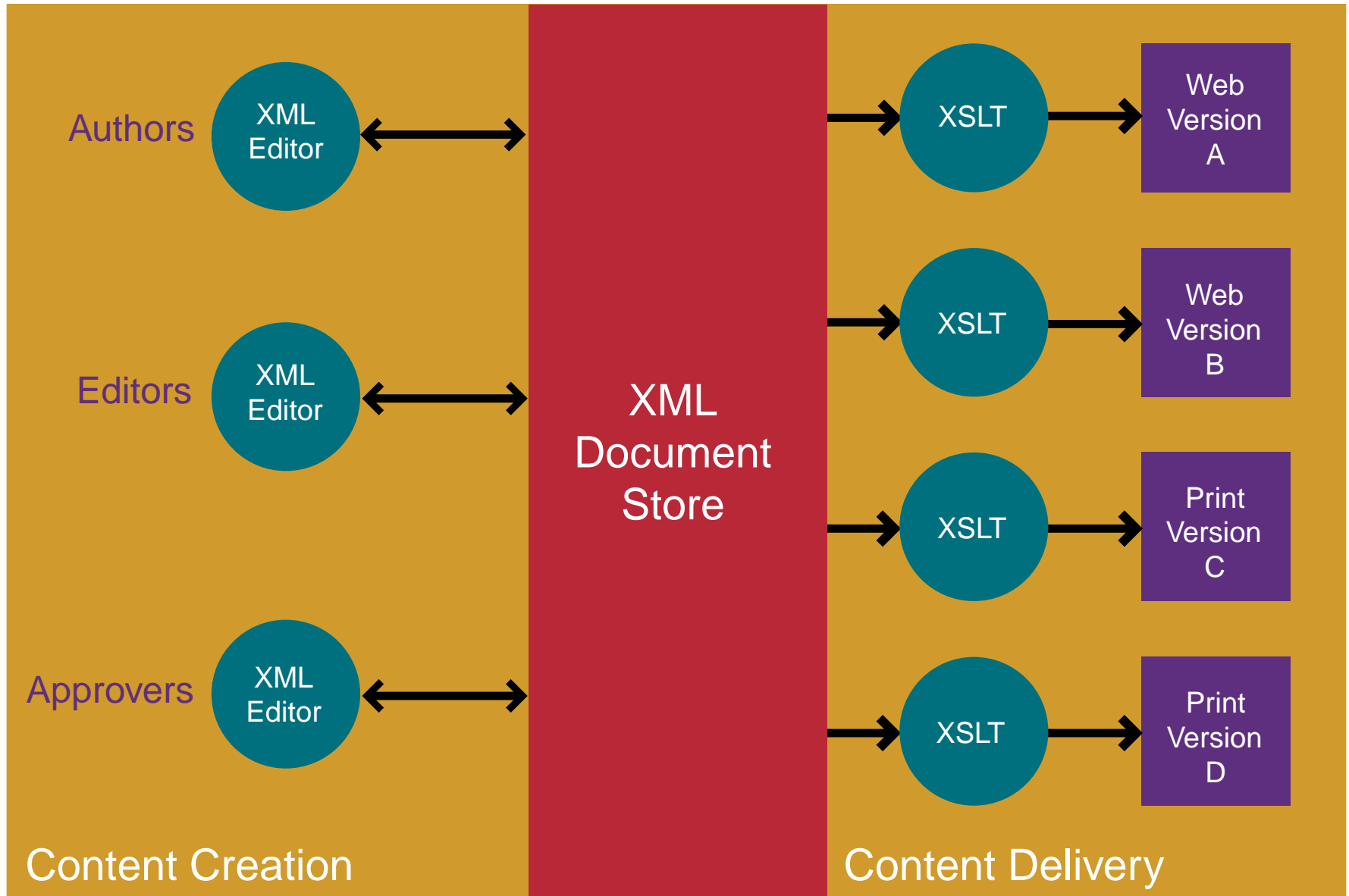
<contact-email>aweber@ligent.net</contact-email>

</pressrelease>

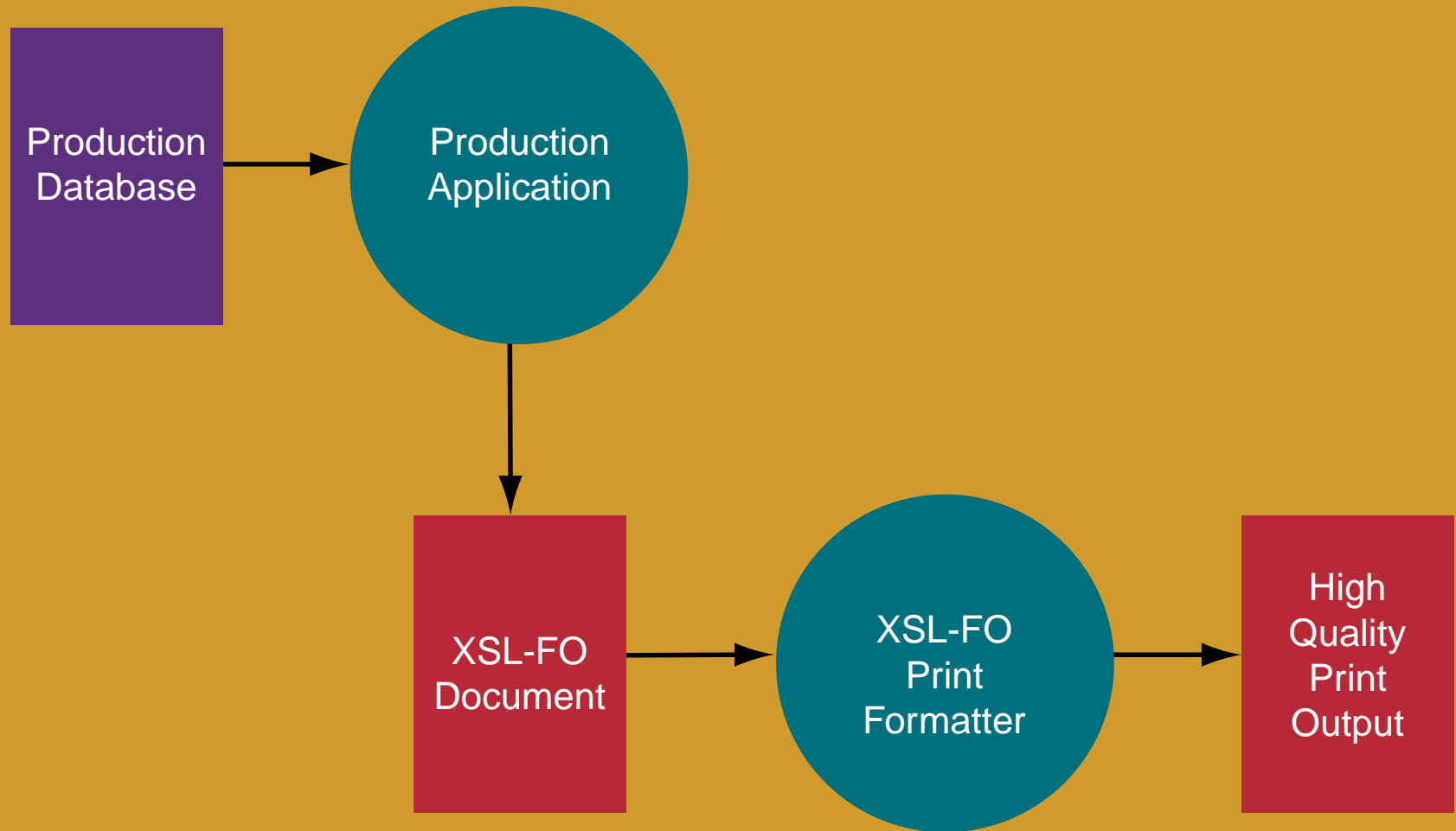
Typical Publishing Use Cases

- **Single-Source publishing**
- High quality print output
- Convenient Web page generation

Single-Source Publishing

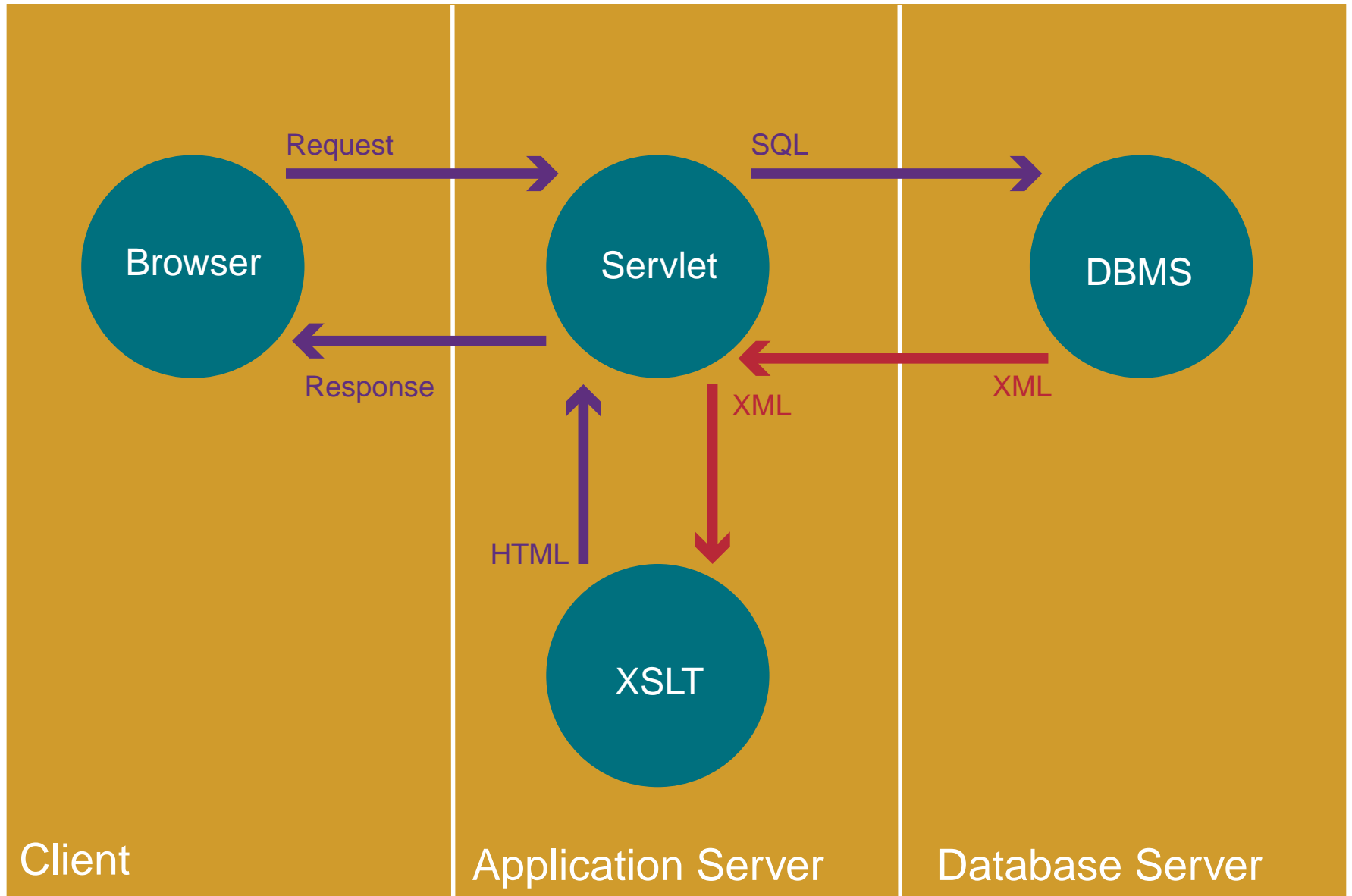


High Quality Print Output



Production System

Convenient Web Page Generation



XML's Publishing Advantage

- Content can be separated from presentation.
- Content can be reused.
- Output can be personalized for the consumer.
- System components can be built using affordable standard XML tools.
- System components can be implemented on heterogeneous computing platforms.

Blank Slide

Blank Slide