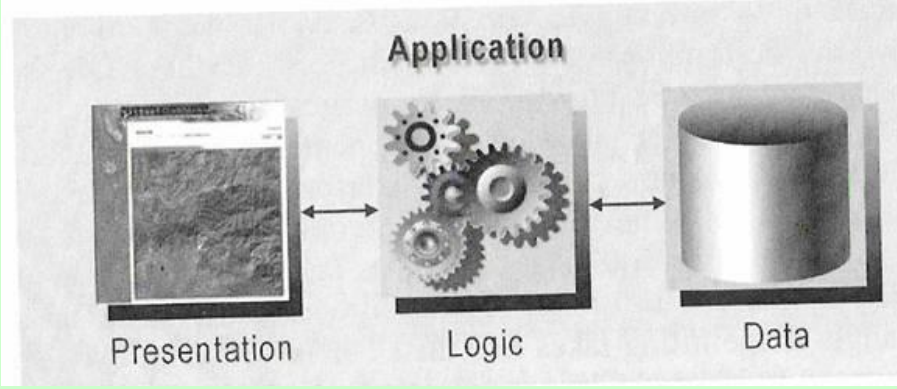


Zhong-Ren Peng and Ming-Hsiang Tsou, 2003 *Internet GIS, Wiley*

GEOG 482/582 : GIS Data Management

Lesson 11: Web-enabled Database Systems

Overview

Learning Objective Questions:

1. What are web-enabled database systems?
2. What is a four-tier architecture for web-enabled systems?
3. What is the family tiering of eXtensible Markup Language (XML) and how is it used?
4. What is web data processing called web services?
5. What are elements of an application architecture?
6. What are four types of client-server architectures?

Lesson Preview

Learning objective questions act as the lesson outline.

Questions beg answers.

Foundations of WWW-enabled Systems

1. What are web-enabled database systems?

- Database systems with an architecture suited for Internet data communications
- Systems with both basic and unique characteristics
- Distinct architectures that integrate spatial database software tools, server-to-database middleware, application programming languages, Internet Markup Languages
- Used with diverse data collection, processing, and dissemination technologies
- Organizational, local, national, and global information infrastructure
- Mobile and participatory business functions
- Widest possible audience

Key terms
Web-enabled systems

Advantages and disadvantages of web-enabled database management

Advantages (opportunities)

- Interoperability in a distributed computing environment
- Hardware and software independence
- Rapid deployment and universal accessibility at a management cost
- Lowering the cost of using spatial information technology
- Improved spatial information customer services

Disadvantages (challenges)

- Inherent limitations of the web (html/xml) services, not geodatabase
- Inherent limitations of spatial databases, 'too large a packet'
- Security and privacy concerns
- Copyright control of abuse and misuse

Intranet and Extranet using the Internet

Internet…the ‘wired physical character’ as a foundation for the WWW.

WWW are the (HTTP) data protocols layered (tiered) on top of the Internet

Intranet – Internet for private use within an organization
Usually contains a security “firewall” server to check incoming and outgoing packets

Extranet – Internet for public use, beyond the organization

Intranet and Extranet use the same technologies basically
… difference involves who has access to information.

Key terms
Intranet
Extranet
firewall

2. What is a four-tier architecture for web-enabled systems?

1st Tier from a users perspective (but 4th from systems capability perspective): Presentations tier, e.g. browser or other client

2nd Tier: Communication infrastructure tier – network connections

3rd Tier: Business logic tier – rules and capabilities;

- web server – Internet connection between client and server
- web activity engine – Internet mapping or processing

4th tier: Data management tier

Spatial database management

Data warehouse

Other database connections

See next slide - Y&H Figure 10.1 Architecture of a web-enabled database system

Key terms
4-tier architecture

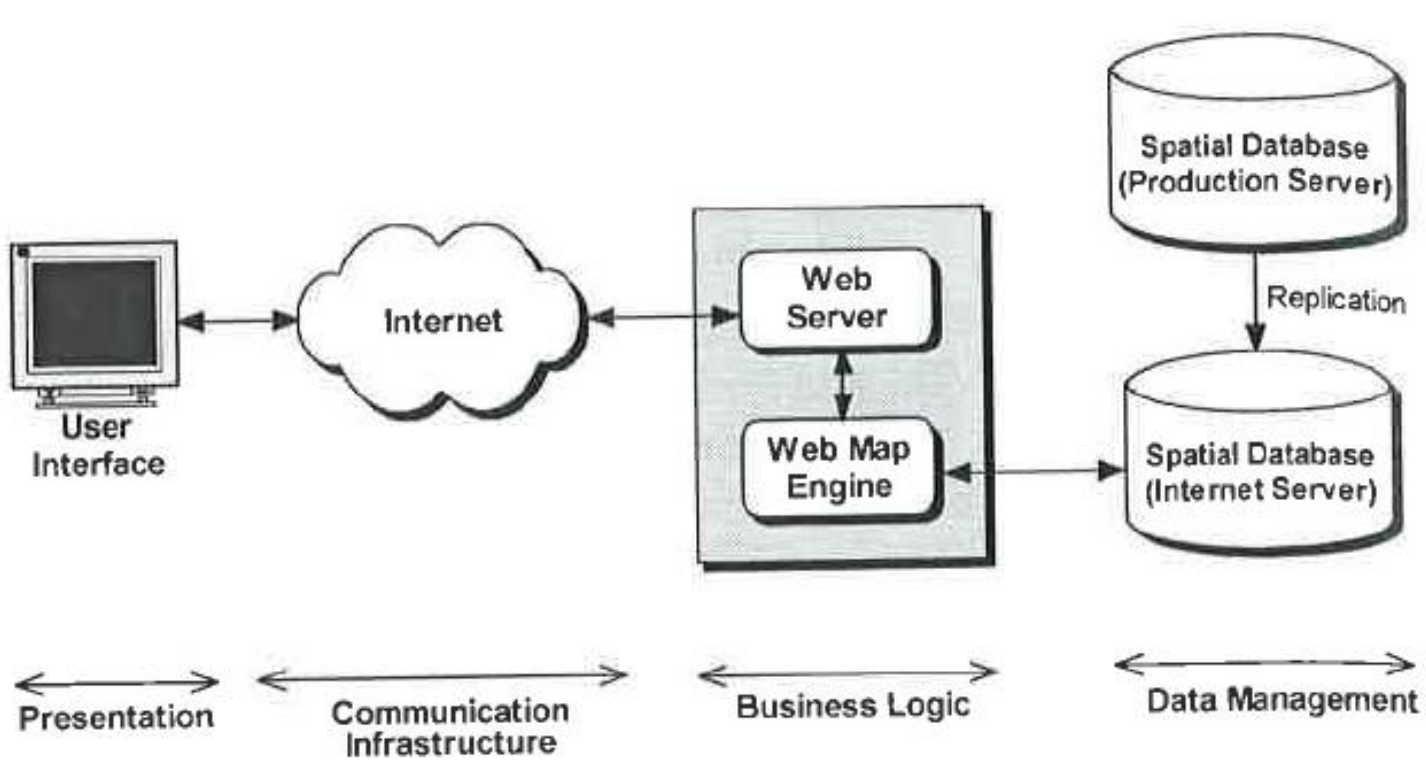


Figure 10-1. Generic Architecture of a web-enabled spatial database system

Web technologies and standards - 1

- Data and information communication over the Internet
- Performed at Tier 2 as Data Communications Infrastructure implemented using communication standards

Four abstraction layers, each with its own protocols. From highest to lowest the layers are:

1) application layer (for example WWW's HTTP) contains all protocols for specific data communications services on a process-to-process level (for example how a web browser communicates with a web server).
2) transport layer (TCP) handles host-to-host communication.
3) interprocess layer (IP) connects local networks, thus establishing internetworking.

Transmission Control Protocol / Interprocess Protocol (TCP/IP) provides end-to-end connectivity specifying how data are formatted, addressed, transmitted, routed and received at the destination. (description at http://en.wikipedia.org/wiki/Tcp/ip)

4) link layer (commonly Ethernet) contains communication technologies for a local network.

Key terms
TCP/IP

Web technologies and standards - 2

- What is an Internet (Interprocess) Protocol (IP) Address?

  (every computer has a unique one – whether dynamic or static at time of processing - on the Internet)
  - Network/host/node address identifies particular computer
  - 32-bit number written as four (byte) octets, i.e. 8 bits (0-255)
    - Octet binary number 11111111 = decimal number 255
  - Each octet separated by a decimal point
  - Two components: domain address & network/host/node address
- Domain address – assigned by International Corporation of Names and Numbers (ICANN)
- Domain name is made up of parts
  - Generic code (.edu, .org, .com or country code .nz, .ca, .nl)
  - Secondary code is a name, developed by system administrator, e.g., as in geography.uw.edu (subdomain and domain names within .edu)
- Network/host/node – assigned by local server administration

Key terms
ICANN

Characteristics of the World Wide Web

Information stores on the WWW

- Website – an information store based on display capability with one or more themes
- Web Portal – a website with generic interactive applications
- Web portlet – an appliance client for dedicated info (weather)
- Web Server – computer host for one or more of above

World Wide Web Consortium (W3C) – 400+ member organizations worldwide for setting standards

Interest in data transmissions – started out in 1993 as Physics dataset transmission

Created HTTP, HTML, URL specifications

- HTTP - HyperText Transfer Protocol
- HTML - HyperText Markup Language
- URL - Uniform Resource Locator

Key terms
W3C
HTTP
HTML
URL

Web pages

- Web page – in simplest form, ASCII text with HTML tags
- HyperText Markup Language (HTML) tags form basis of HTML
- Open any web page in an ASCII editor (e.g. Notepad++)
- View the Tags – pairs of tags are used; open with a tag command, and close command with a slash <tag> <\tag>
- Hyperlinks – URL expressions used to jump from one page URL address to another

Uniform Resource Identifier (URI) – string of characters for unique address

- URI : can be either URL or URN or both

- URL = HTTP + TCP/IP

e.g., http://en.wikipedia.org/wiki/Main_Page identifies a resource, in this case English Wikipedia's home page, whose representation, in the form of the home page's current HTML and related code, as encoded characters, is obtainable via the HyperText Transfer Protocol from a network host whose domain name is en.wikipedia.org.

- URN :Uniform Resource Name – string of characters for unique name
  For example an ISBN as book naming string for Yeung and Hall *Spatial Database Systems, 2007 (paperback)*
  - *urn:isbn:1-4020-5393-2*

Hypertext transfer protocol (HTTP) runs on top of TCP/IP

See Y&H Figure 10-2 for sequence of steps

1) Client browser forms request using URL (HTTP on TCP/IP)

Note: "1" does not appear in figure 10-2; it is a workstation client creating a request for information

2) Link to host server is made using node portion of URL (HTTP and TCP/IP)

3) Server processes request into packet(s)

4) Returns results of request packet(s) using HTTP and TCP/IP (on the network)

5) HTML page packets are decoded in the browser and displayed on screen

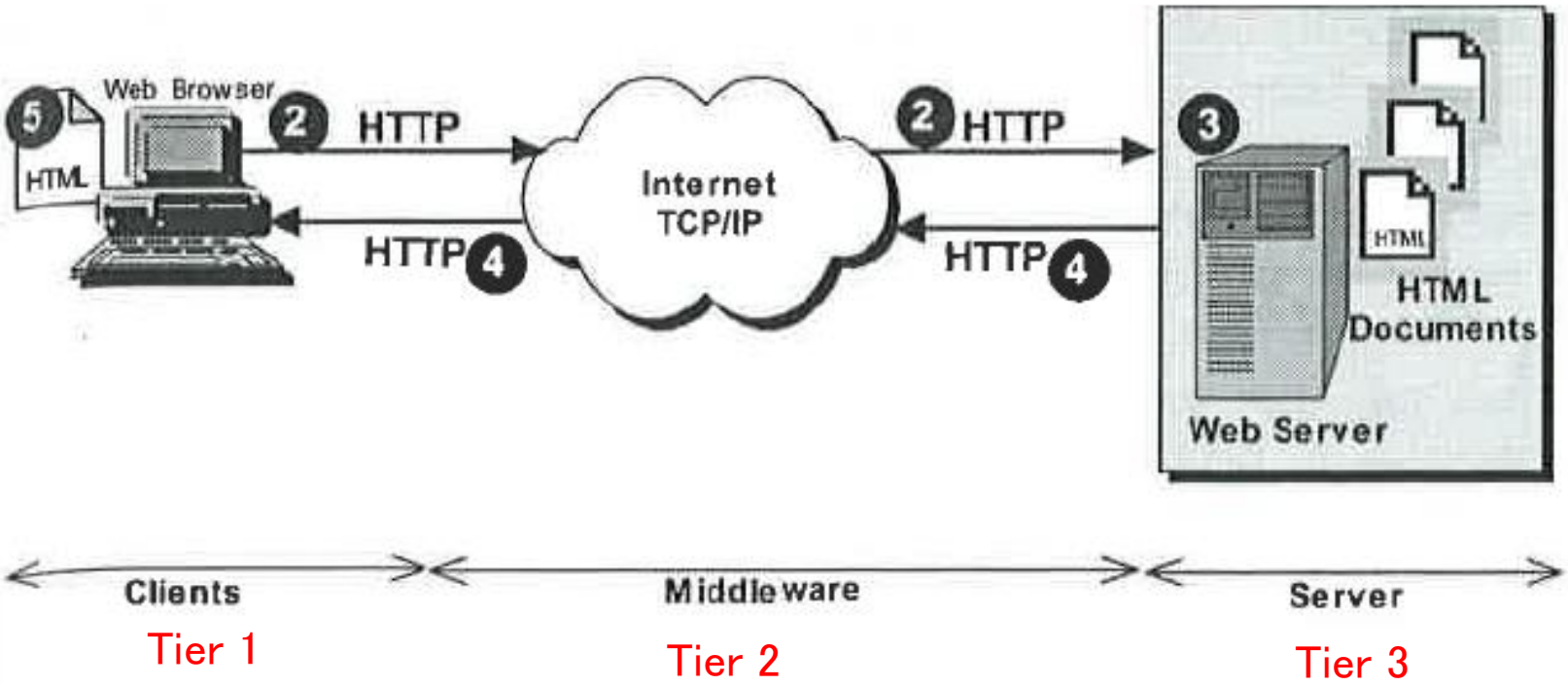


Figure 10-2. Web client/server interaction using HTTP

Step 1: Person invokes client

Tier 4 DBMS not depicted on slide; see slide 7 Figure 10-1

Yeung and Hall *Spatial Database Systems, 2007*

3. What is the family tiering of eXtensible Markup Language (XML) and how is it used?

eXtensible Markup Language (XML) is an expansion of HTML; NOT an extension

See Y&H Figure 10-4 XML family of technologies and standards

Key terms
XML

Three-tiers for implementing the XML family:

3) Foundation

2) Tools

1) Dialects and applications

XML family foundation

3) Foundation - Meta Markup Language

Collection of rules for defining elements

Attributes and tags enclosed within XML documents

Documents contain XML expressed within a dialect

XML family tools

2) Tools

Tool library for creating and interpreting XML document

Describe and validate structure - Document Type Definitions (DTD)

Schema content modeling –XML Schema (like in ArcGIS Diagrammer)

Format data for display - Stylesheet Language (XSL)

Link to a database – Linking Language (XLL)

Query a database – XML Query (XQuery) and XML Structured Query (XSQ)

Convert XML from one dialect to another - XML Stylesheet Language Transform (XSLT)

Key terms

DTD
XSL
XLL
XQuery
XSQ
XSLT

XML family dialects

1) Dialects

X3D - 3D gaphics

GML – Geography Markup Language

MathML – Mathematics

SVG – Scalar Vector Graphics for 2D graphics

XHTML WWW extensible hypertext markup language

CML – Chemistry Markup Language

SML – Multimedia Markup Language

XFORM – form generation

WSDL – web service description language

SOAP – Simple Object Access Protocol / Service Oriented Access Protocol

ArcXML – Esri schema language using Schema tool (2)

For databases, two ways to use XML

Document-centric model – collection of documents

- XML creates semi-structured documents with irregular content
- XHTML was created for this purpose
- Spatial Metadata is a good application

Data-centric model – collection of 'raw' data

- Storage or interchange format for data
- Transferred by middleware application (e.g. MS ADO.NET)
- Or, XML enabling the database (e.g. Oracle 9i, MS SQL Server 20XX)
- GML dialect used for interoperable data transfer

4. What is web data processing called web services?

Services – software components used to build comprehensive applications

- Simple Object Access Protocol (SOAP) – information rich and flexible, thus difficult to implement
- Web Service Description Language (WSDL) – information simple and more structured (constrained), thus easy to implement

Key terms
SOAP
WSDL

Individual services are invoked through API or RPC running on desktop or mobile

Vendor-independent, standards-based activation of remote cross-platform procedure

See Y&H Figure 10-10 Spatial Database Application Web Services Model

Web service provider offers the service

Provider Develops and Deploys service on a Web Server and makes it discoverable and invokable through an online registry

Directory used for discovery of service: for SOAP universal directory, for WSDL narrow directory; WSDL is most common

Registry provides info to define the "contract" for invoking

- name of service
- service inputs
- services outputs
- how to communicate with service

Service follows rigidly defined document specifying a XML Schema

Binding: Discovery and Invoking a service

Web service works using an architecture built from stack of software layers based on what we considered earlier

Components loosely interact with each other using standard protocols

Building block foundation of stack(s) are standards

HTML announced with and wrapped by HTTP

HTTP announced with and wrapped by TCP/IP'

Software tools and interfaces enable client to server communication:

- XML – basic foundation
- SOAP – collection of XML-based rules defining format of communication between Web service and clients
- WSDL – XML based rules defining web service interface, data and message types, interaction patterns, and protocol mapping
- Universal Description, Discovery and Integration (UDDI) – web services registry/discovery mechanism for storing/categorizing information and retrieving pointers to Web service interfaces
- Open Geospatial Consortium Web Services – OGC family of standards specifications

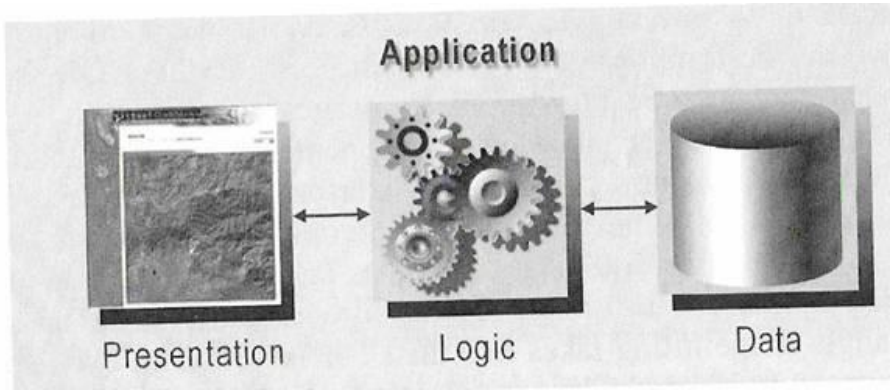
Next…Web-enabled Application Architectures

Basic considerations for software implementation in application architecture

- Presentation, business logic, data
- Client, server
- Stand-alone vs. client/server systems
- Thin client vs. thick client
- Middleware (application server)

5. What are elements of an application architecture?

Three key elements used to describe application workload.



- Presentation work: provides display of information
- Logic work: performs operations on data to create information
- Data work: performs access to data management

- Where is the computer data processing work performed?
- Distribution (partitioning) of workload is key to understanding architecture?
- Three application activities can be performed on one computer or many computers. What is the requirement for system configuration?

Strategies for Application Implementation

Information-centric versus Application-centric

Information-centric approach – very short-term transactions that provide information, e.g. weather, traffic, and road conditions information provided this way, as in Google Maps, and MS Bing (implemented using a thin client)

Rapid access is the primary consideration

Application-centric approach – long term transactions that provide domain-specific data to applications for use by professional users for deeper information and knowledge construction (implemented using a thick client)

Sophisticated information processing is the primary consideration

Stand-alone GIS vs. Client-server GIS

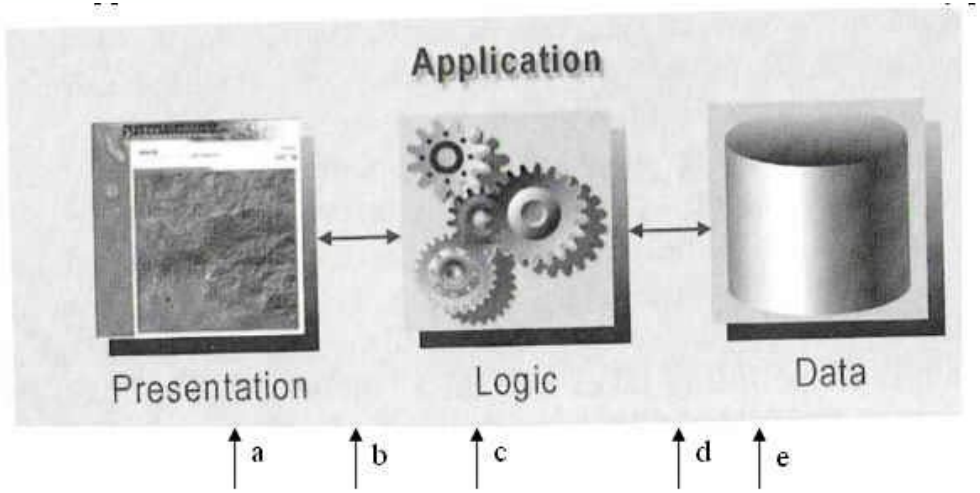
If the three elements reside in the same machine (computer), the application is called stand-alone

- e.g. ArcGIS Desktop or Laptop

If the three elements reside in different machines (computers), the application is called client-server

- When web browser acts as client remote from the server where data and analysis logic resides
- Can be implemented with different network configurations (e.g. local area network (LAN) or wide area network (WAN)
- Difference in LAN or WAN is the nature of the data transmission from computer to computer (material coming within a few slides)

Thin client vs. thick client processing



Thin client Thick client

Three application elements (workload) can be partitioned into client and server at any point (a to e)

The system whose partitioning point is closer to "a" is said to have thin client (i.e. light work load in client, the client is not doing much of the data processing, e.g., b and c

The system whose partitioning point is closer to "e" is said to have thick client (i.e. heavy work load in client), the client is doing most of the data processing, e.g. d and e.

Middleware

For client to communicate with the server, we need translator to link them together.

This intermediate connection component is called “middleware” or application server

e.g. ArcSDE middleware links ArcGIS Desktop (client) to PostgresQL DBMS (server) within enterprise-wide database GIS environment

ArcGIS implements web service support for OGC and ISO standards

Two basic network types

Difference between LAN and WAN is in the standards stack and wiring used to implement data transmission

LANs (Local Area Networks)

- Support communications within a building or localized environment
- Generally speaking, support high-bandwidth communications over short distances
- Provide high-speed access data

WANs (Wide Area Networks)

- Support communications between remote locations
- Generally speaking, support lower-bandwidth than LAN
- The internet is a global WAN

Key terms
LAN
WAN

6. What are four types of client-server system architectures?

What does a configuration look like?

Which one should you choose?

How do you choose them?

Four types of configurations implement thin-thick clients…

1. Central file server with workstation (thick) clients
2. Central DBMS server with workstation (thick) clients
3. Centralized application processing with terminal (thin) clients
4. Web transaction processing with browser (thin) or workstation (thick) clients

1. Central file server with workstation (thick) clients

Data is retrieved from within files on a server and processed on the workstation

- Client contains presentation and business logic (i.e. thick clients are commonly involved)
- Server contains data management

Requires the transfer of large amounts of data from the server to the client

High demand for bandwidth since a lot of data is transmitted

Rather thick client processes a large data file

2. Central DBMS server with workstation (thick) clients

Data is retrieved from the server by a DBMS and map rendering is processed on the workstation

Only the data required to support the client display be transferred because DBMS filters data needed

Reduce demands on the network data transmission

In the middle of thick and thin range, logic processing shared

This is the configuration we use in Lab 4, 5, and 6

3. Centralized application processing with terminal (thin) clients

Data and application software are both stored and run on servers

Significantly reduces network bandwidth requirements

Will be best deployed over WANs

In the middle of thick and thin, logic processing is shared

4. Web transaction processing with browser or workstation

Application software and data files reside on servers

Web browser display information products via the Internet or intranet
(i.e. thin clients)

Requires sequential support to a large number of user transactions

Can be best deployed over WANs

A thin client, wherein the server is doing most of the processing.

This is ArcGIS Online configuration

Comparing four types of client-server architectures

Thick clients

Large amount of data transfer

Small # users

Thin clients

Small amount of data transfer

Large # users

Arranged by client thickness (i.e., workload activity):

1. Central file server with workstation clients doing most of work
2. Central DBMS server with workstation clients doing some of work
3. Centralized application processing with terminal clients do little work
4. Web transaction processing with browser or workstation clients doing little work

Decision considerations about client-server communication technologies

The relation between system interface and network configuration involves

- **Data volume**: use file server or DBMS server?
- **Level of computing complexity with number of users**: thick client or thin client? Web transaction?
- **Wait tolerance**: emergency dispatch requires low wait tolerance, and thus demand high bandwidth
- **Technology life cycle**: is new technology always better?
- **Organization policies and standards**: share common approach

Summary

In this lesson, you learned about…

1. The basics of web-enabled database systems
2. A four-tier architecture for web-enabled systems
3. eXtensible Markup Language (XML) and how it is used
4. Web data processing called web services
5. Elements of an application architecture
6. Four types of client-server system architectures

Contact me at nyerges@uw.edu if you have questions or comments about this lesson.

GEOG 482/582: GIS Data Management

END Lesson 11: Web-enabled Database Systems