Web System Architecture

Introduction

- The World Wide Web (WWW) can be viewed as a huge distributed system with millions of clients and servers for accessing linked documents.

- Servers maintain collections of documents while clients provide users an easy-to-use interface for presenting and accessing those documents.

- A document is fetched from a server, transferred to a client, and presented on the screen. To a user there is conceptually no difference between a document stored locally or in another part of the world.

Contd…

- Now, Web has become more than just a simple document based system.

- With the emergence of Web services, it is becoming *a system of distributed services* rather than just documents offered to any user or machine.

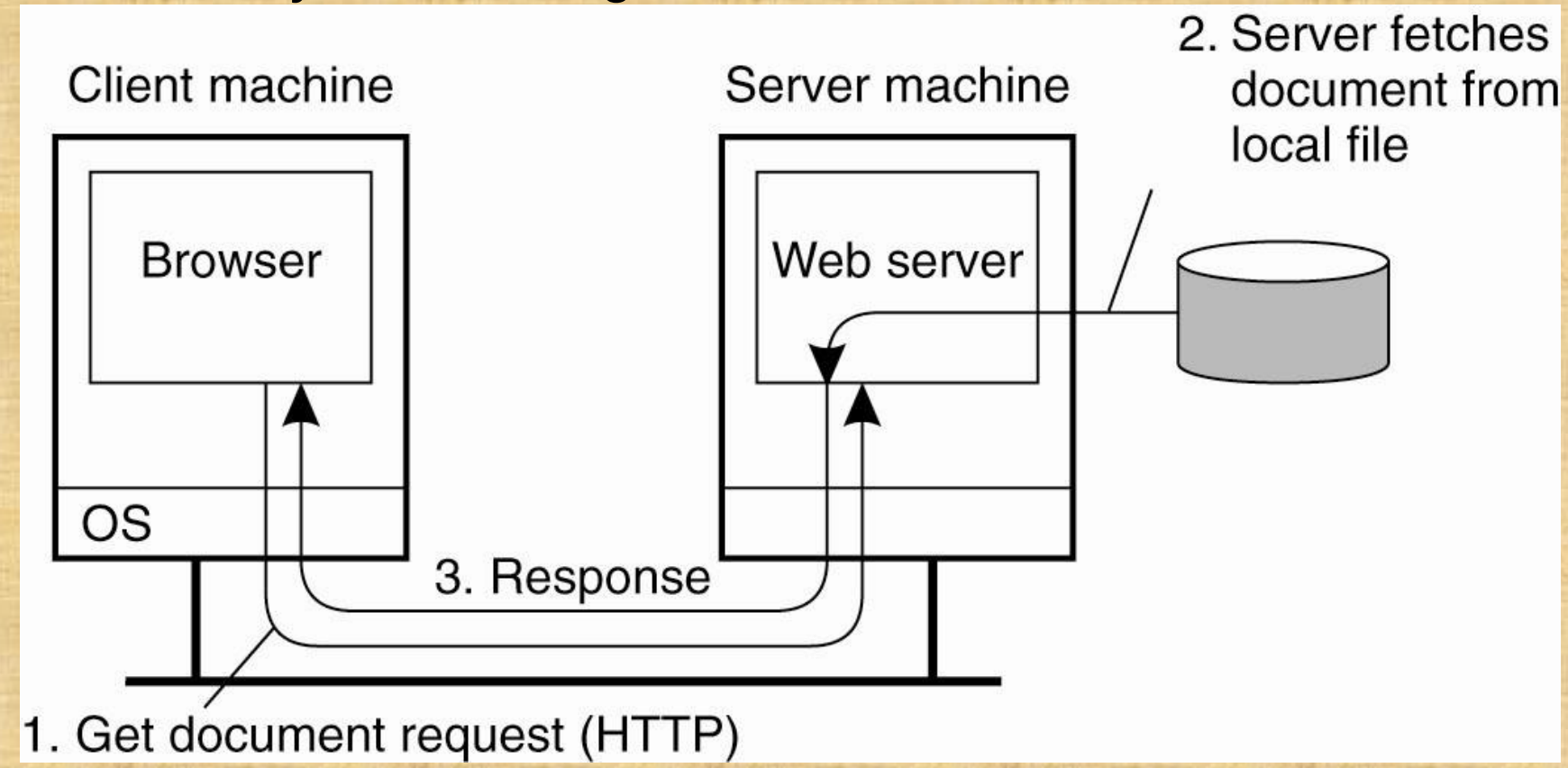
- What can we get from WWW?
  - Read news, listen to music and watch video;
  - Buy or sell goods such as books, airline tickets;
  - Make reservations on hotel room, rental car, restaurant, etc.;
  - Pay bills and transfer money from one bank account to another

- In 1989, Tim Berners-Lee had suggested a way to let all users, but particularly scientists, browse each others' papers on the Internet.
- He developed HTML, URLs, and HTTP.

Traditional web-based systems

1. Many Web-based systems are still organized as simple client-server architectures.

2. The core of a Web site: a process that has access to a local file system storing documents.



Uniform Resource Locator

- A reference called Uniform Resource Locator (URL) is used to refer a document.
- The DNS name of its associated server along with a file name is specified.
- The URL also specifies the protocol for transferring the document across the network.
- Example:

  http://www.cse.unl.edu/~ylu/csce855/notes/web-system.ppt

Browser

- A client interacts with Web servers through a special application known as browser.

  - Responsible for displaying documents.

- A Web document does not only contain text, but it can include all kinds of dynamic features such as audio, video, animations, etc.
- In many cases special helper applications (interpreters) are needed, and they are integrated into the browser.
  - E.g., Windows Media Player and QuickTime Player for playing streaming content
- The variety of document types forces browser to be extensible. As a result, plug-ins are required to follow a standard interfaces so that they can be easily integrated with the browsers

General Architecture

- Historically, a web browser
- But also:
  - Mobile application
  - Desktop application
  - Other server application

The web is a two-tiered architecture.
A web browser displays information content,
and a web server that transfers information to the client.

Courtesy:Ambient Intelligence Technologies

Contd…

Web Language

HTML is the standard markup language for information publishing.

- Elements of WWW

-Client/server

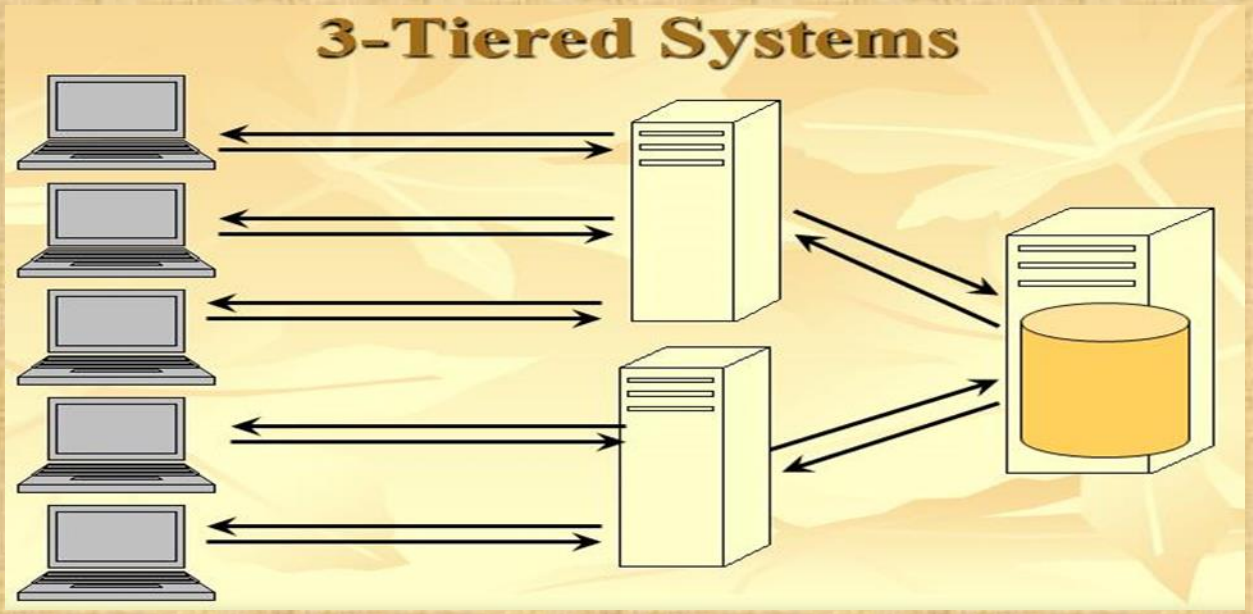
- Web Languages and Protocols

- Web Pages

- Home Page

- Web Browsers

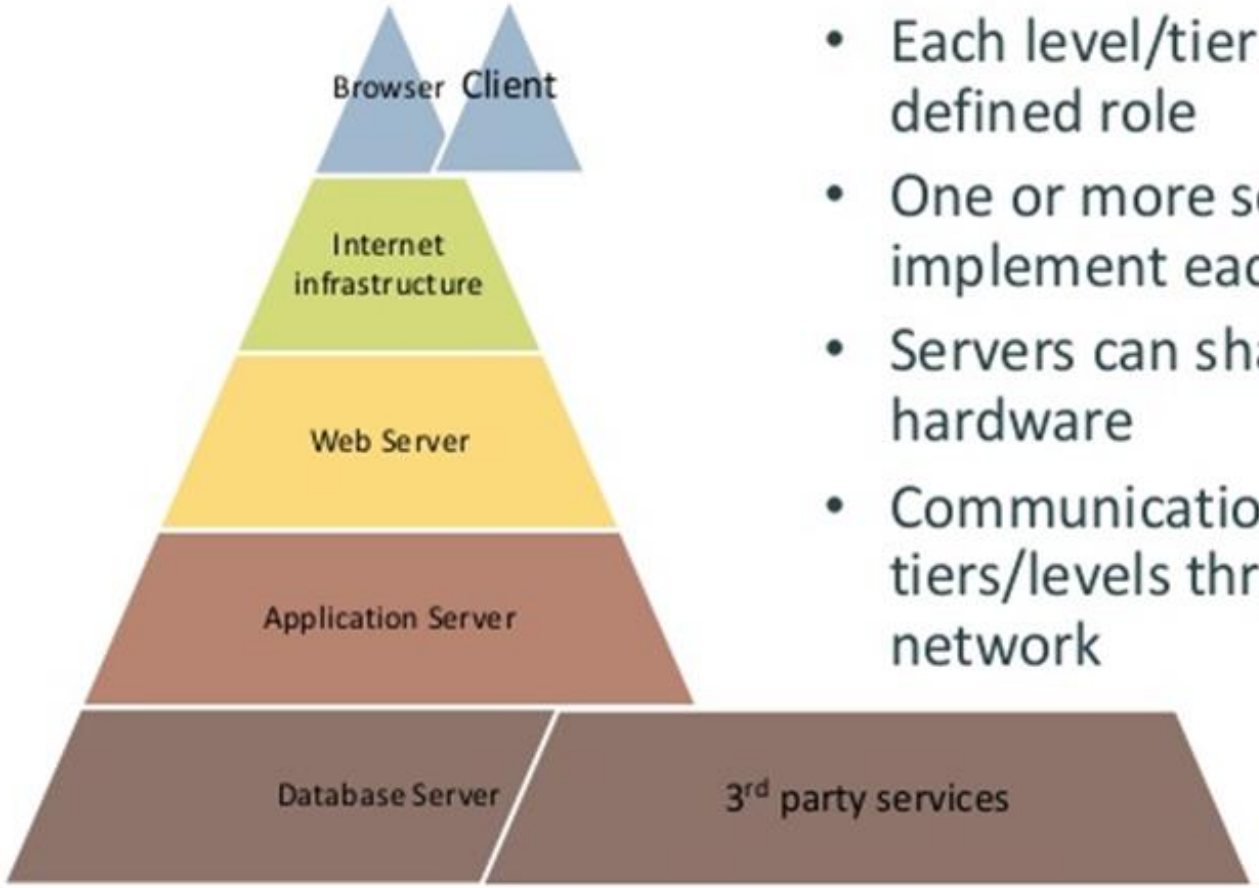
- Web Sites



- A web browser is the **first tier** (presentation),
- An engine using some dynamic Web content technology (such as ASP, ASP.NET, CGI, ColdFusion, JSP/Java, PHP, Perl, Python, Ruby on Rails or Struts2) is the **middle tier**
- A database is the **third tier** (storage).

- Multi-tier application (*n*-tier application)
  - Information tier (data or bottom tier)
    - Maintains data for the application
    - Stores data in a relational database management system (RDBMS)
  - Middle tier
    - Implements business logic and presentation logic
    - Control interactions between application clients and application data
  - Client tier (top tier)
    - Application's user interface
    - Users interact directly with the application through the client tier

N-tier (N-level) architecture



- Each level/tier has a well defined role
- One or more servers implement each tier/layer
- Servers can share hardware
- Communication between tiers/levels through the network

Courtesy:Ambient Intelligence Technologies

Components

- connections to the Internet - Internet Service Provider (ISP)
- servers - implement each tier/level of the architecture.
- physical networks - interconnecting the servers
- One or more network devices (router, firewall, switch) - implement communication and security policies

Web Browser

- The primary purpose is to bring information resources to the user.
- An application for retrieving, presenting, and traversing information resources

Web server

- Manages the HTTP protocol (handles requests and provides responses)
  - Receives client requests
  - Reads static pages/contents from the file system
  - Activates the application server for dynamic pages/content generation (server-side)
  - Provides an file (HTML, or other) back to the client
- One HTTP connection for each request
- Multi-process, Multi-threaded or Process pool

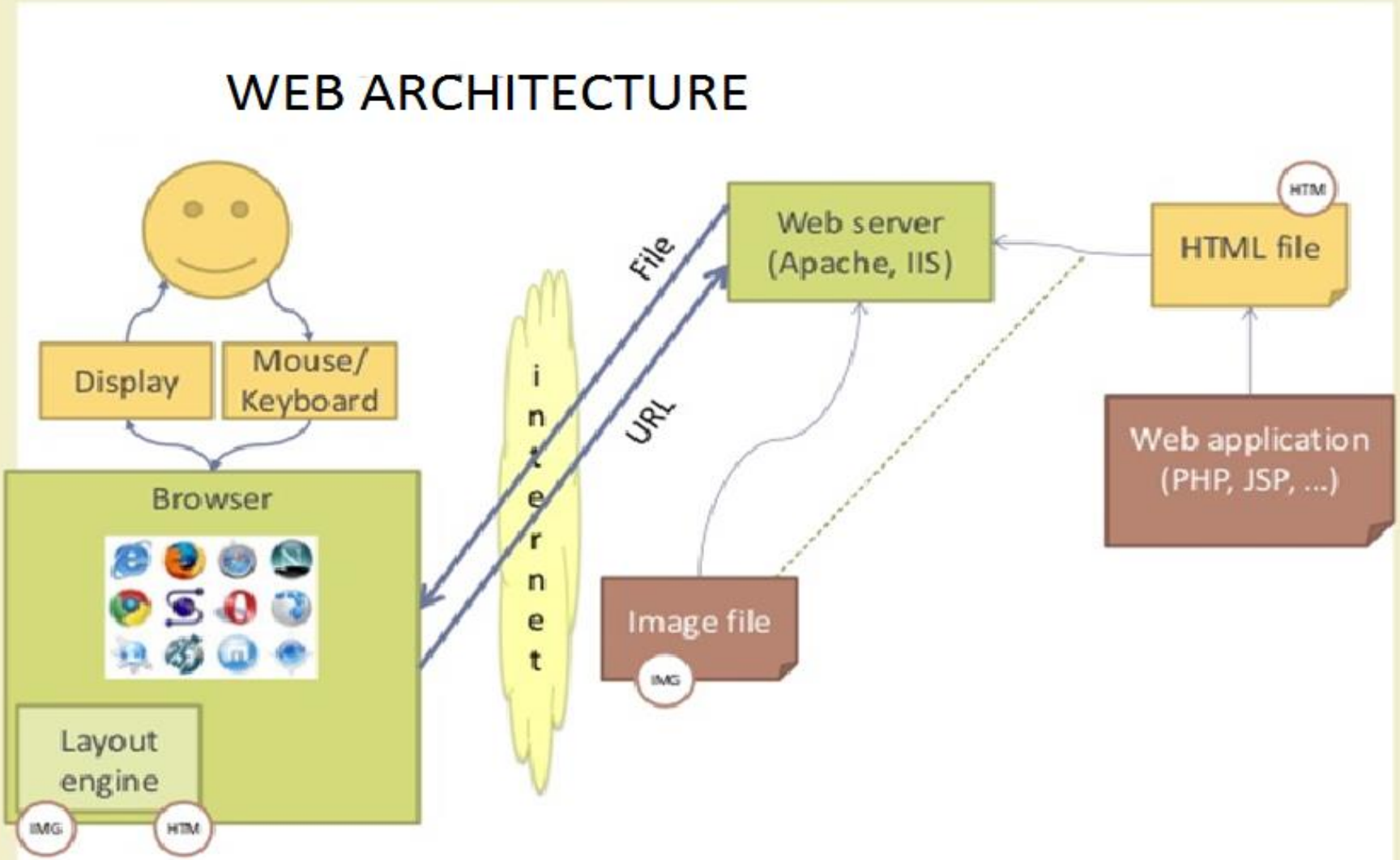
The most adopted web server - Apache

HTML

- HyperText Markup Language
- Document layout language (not a programming language)
- Defines structure and appearance of Web pages

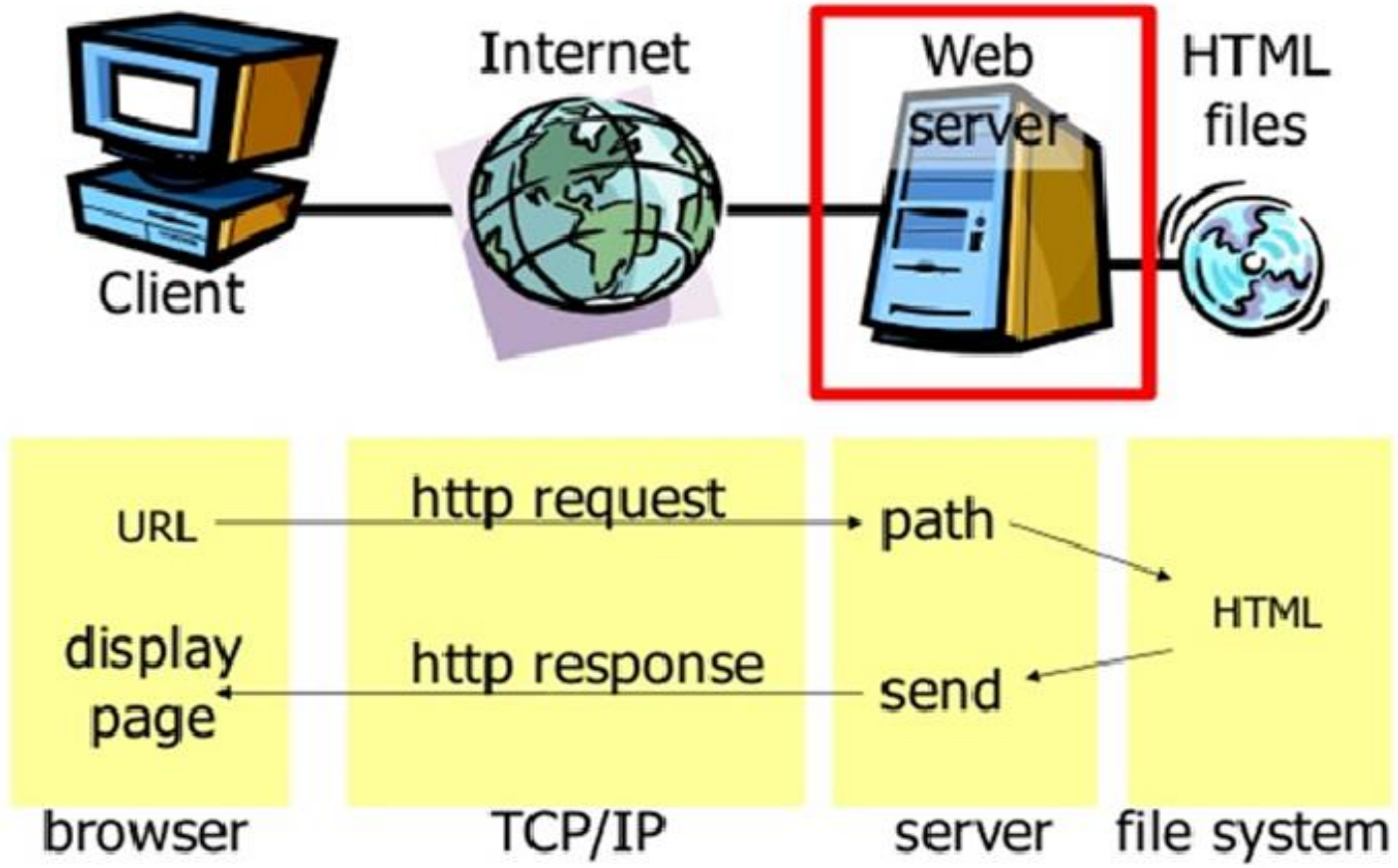
URI

- Universal Resource Identifier
- URLs are location dependent
- It contains four distinct parts: the protocol type, the machine name, the directory path and the file name.
- There are several kinds of URLs: file URLs, FTP URLs, and HTTP URLs



Courtesy:Ambient Intelligence Technologies

Example



Courtesy:Ambient Intelligence Technologies

Multitiered architectures

- Web documents can be built in two ways:

Static	Dynamic
On request locates and returns the object	Request is forwarded to an application system
Includes predefined HTML pages & GIF or JPEG files	On request reply is generated dynamically
Does not require web server to communicate	Done through server side execution(program)

- Although Web started as simple two-tiered client-server architecture for static Web documents, this architecture has been extended to support advanced type of documents.

Contd…

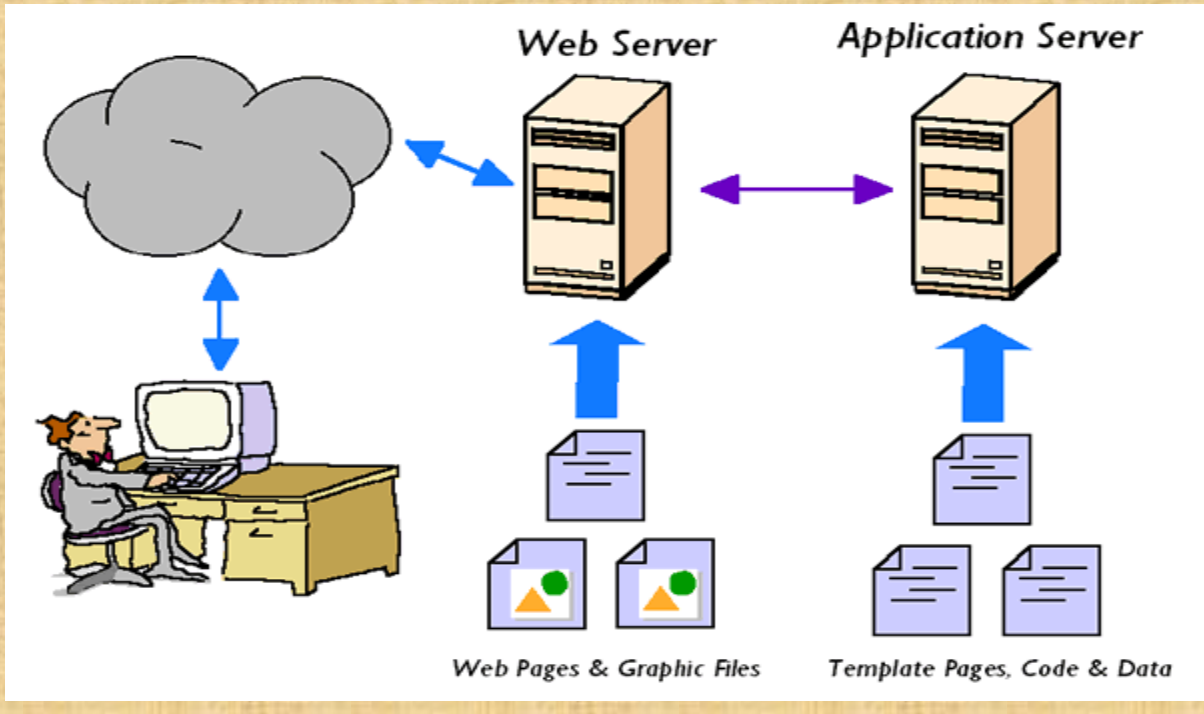
- Most of the Web sites are now organized as three-tiered architectures consisting of a Web server, an application server and a database server.

- User data comes from an HTML form, specifying the program and parameters.

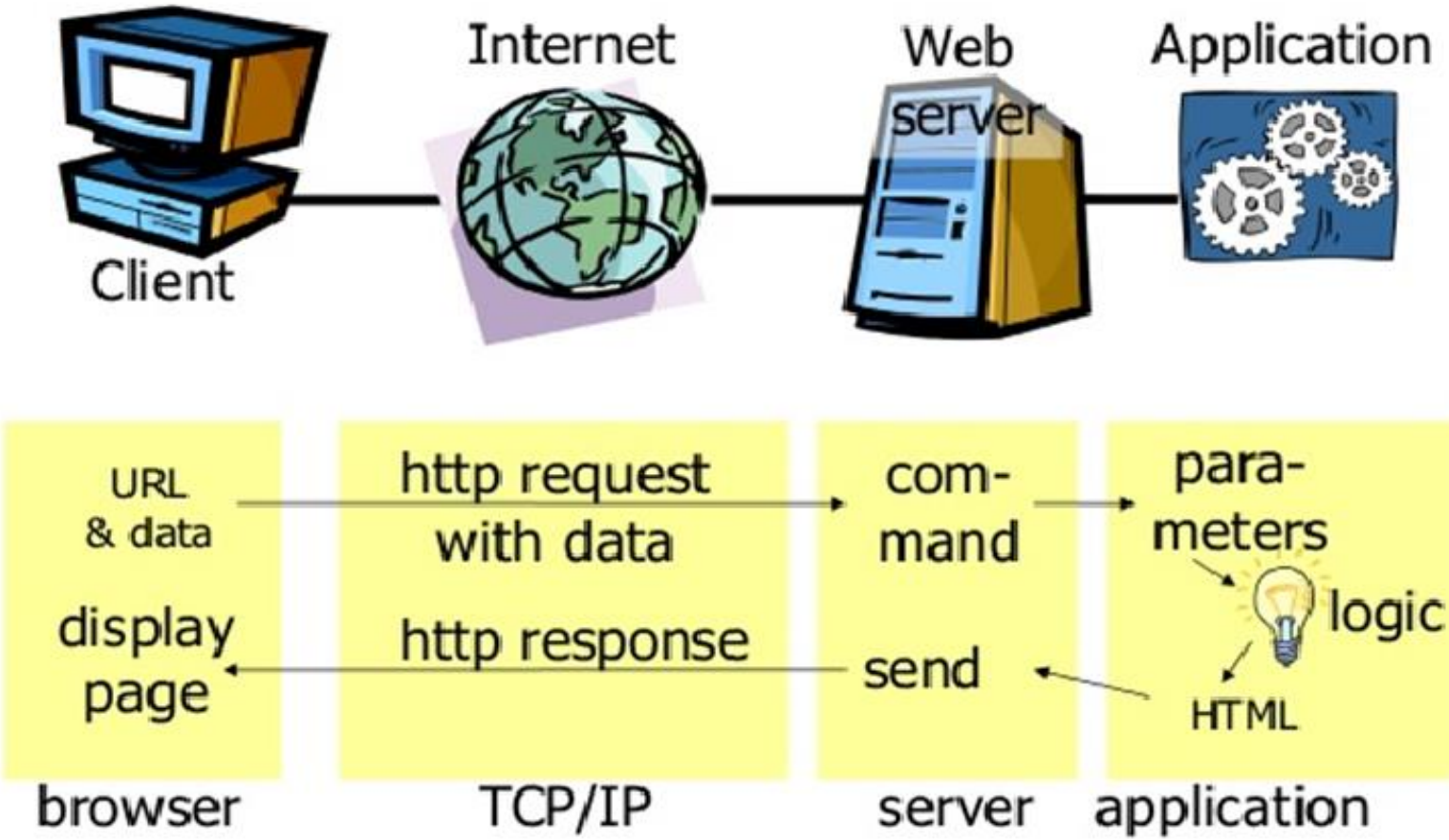
- Server-side scripting technologies are used to generate dynamic content:
  - Microsoft: Active Server Pages (ASP.NET)
  - Sun: Java Server Pages (JSP)
  - Netscape: JavaScript
  - Free Software Foundation: PHP

Application server

- Dynamic page generation and content generation
- Manages the site business logic
- It's the middle tier between the client browser and the data residing on a database
- Implements the session mechanisms
- Different technologies and architectures are available.



Dynamic web transaction



Courtesy:Ambient Intelligence Technologies

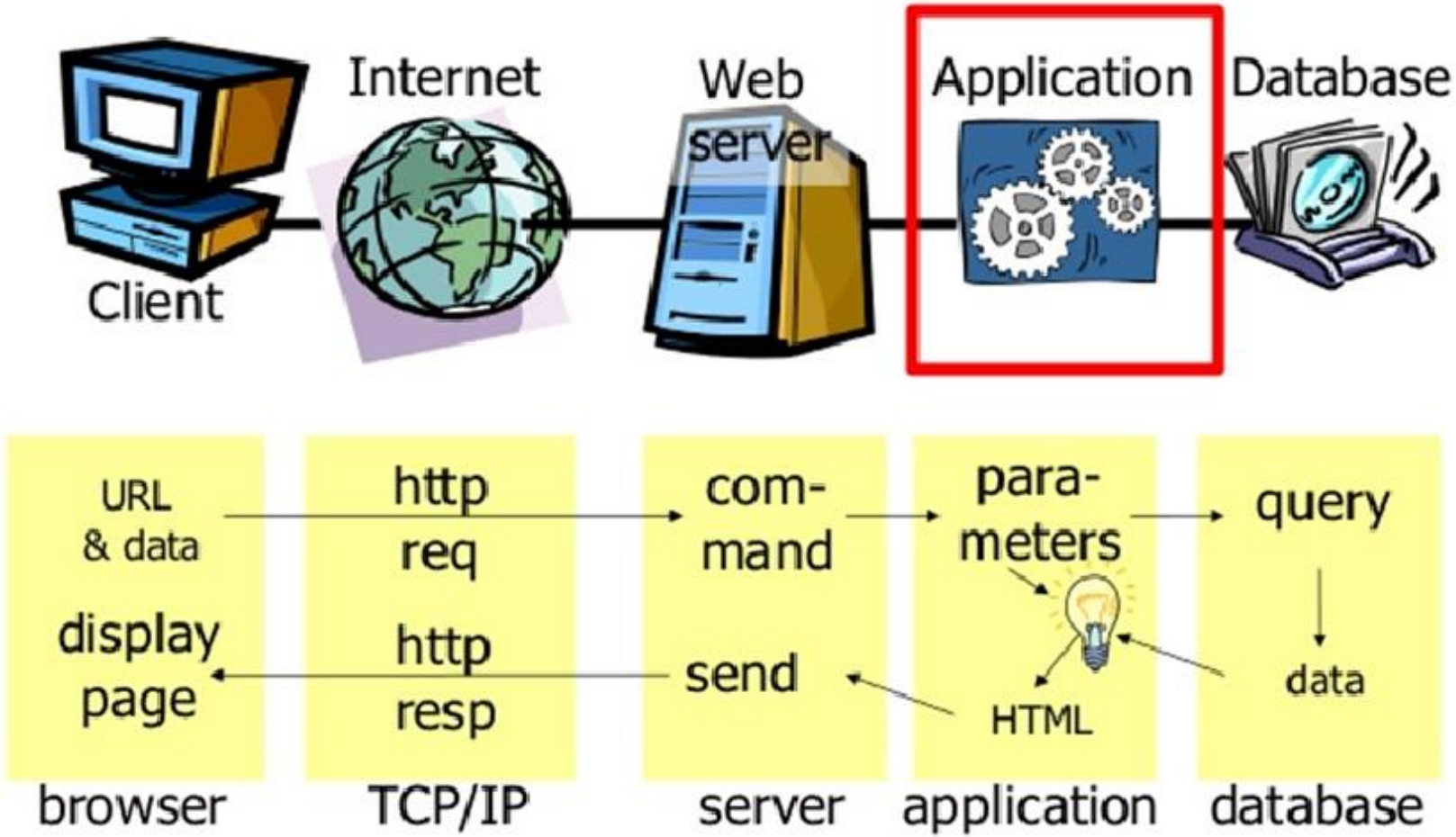
Adopted standards

- URL (uniform resource locator) for finding web pages
- HTML (hyper text markup language) for writing web pages
- GIF (graphics interchange format), PNG (portable network graphics), JPEG, … for images
- HTTP (hyper text transfer protocol) for client-server interaction
- TCP/IP (transmission control protocol over internet protocol) for data transfer

Database server

- Stores the data on which the application server works.
- Executes the queries issued by the application server:
  - Updates the stored data
  - Inserts new data
  - Provides back query results
- The most frequent/complex queries can be implemented internally as stored procedures (pre- compiled queries with parameters)

Example

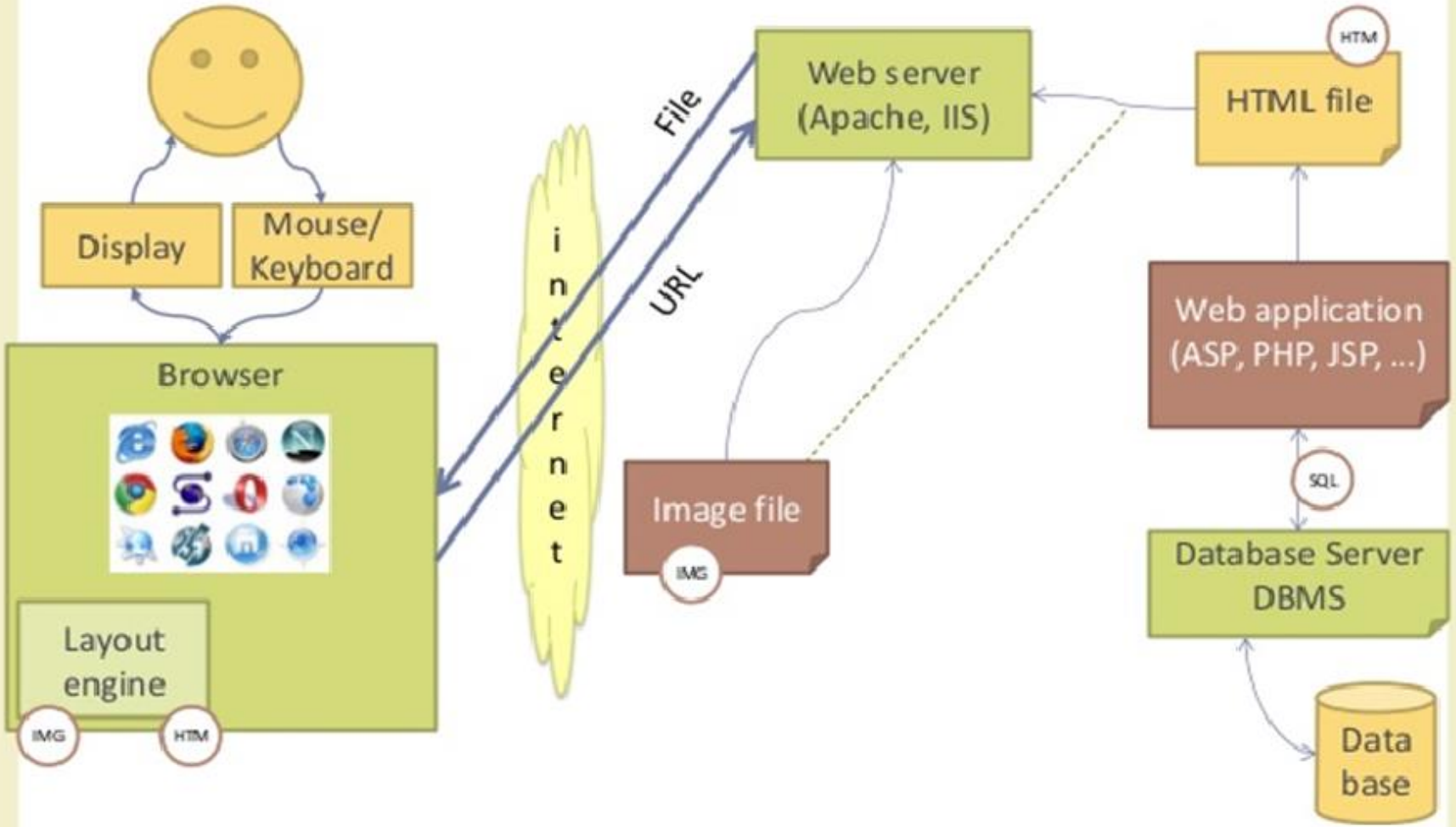


Courtesy:Ambient Intelligence Technologies

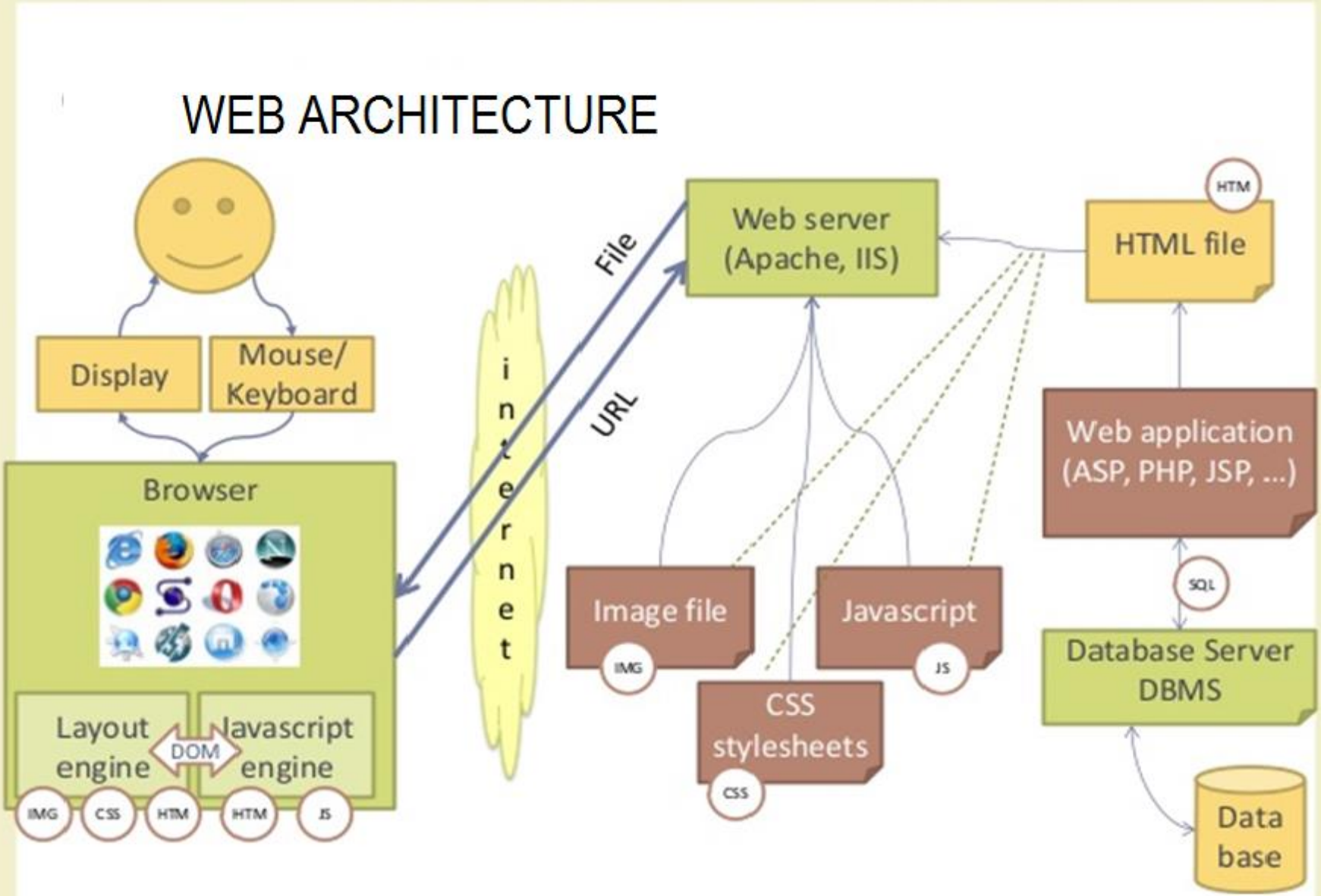
Adopted standards

- Cookies for storing the state of a session
- Java, JavaScript, ActiveX, Flash to program the user interface on the browser
- SQL (structured query language), ODBC (open database connectivity) to access data bases

General Architecture



Courtesy:Ambient Intelligence Technologies



Courtesy:Ambient Intelligence Technologies

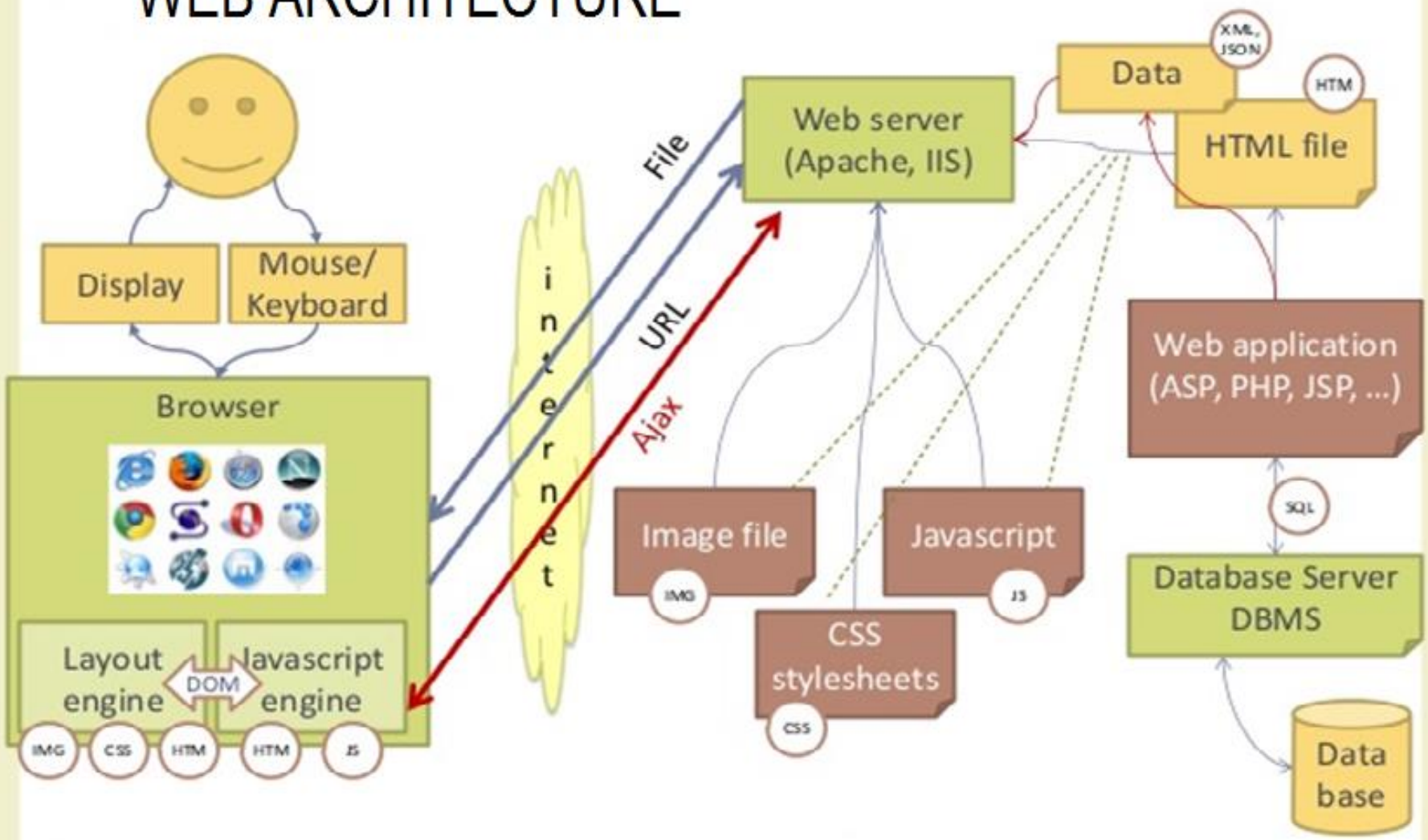
Web 2.0

- Web applications support social interaction models
- Peer exchange and user-contributed content instead of rigid publisher/reader pattern
  - Online communities
- Rich, dynamic, interactive user interfaces
- Integration of contents across web sites (mashups)

Adopted standards

- Dynamic HTML: DOM, Javascript, CSS
  - JavaScript, Flash to handle a runtime environment on the browser
  - DOM (XHTML Document Object Model) to allow on-the fly modification of the web page
  - CSS 2.1 to modify attribute and handle objects
- AJAX: Asynchronous Javascript and XML
  - XMLHttpRequest for asynchronous communication to the server
  - Data transfer formats: JSON, XML, RDF, RSS, Atom, FOAF, ...
- Mash-up technology

WEB ARCHITECTURE



Courtesy:Ambient Intelligence Technologies

Web Security Issues

- The Web has become the visible interface of the Internet

  - Many corporations now use the Web for advertising, marketing and sales

- Web servers might be easy to use

but…

- Complicated to configure correctly and difficult to build without security flaws
  - They can serve as a security hole by which an adversary might access other data and computer systems

-Web Security issues are with respect to Integrity, Confidentiality, DoS, Authentication