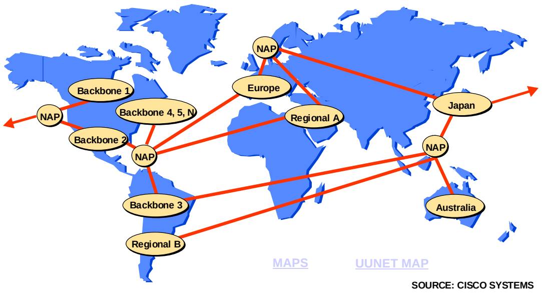
http + servers outline

Administriva: rooms, readings.

- **Networking wrap up.**
- **Big Picture**
- **What happens when you click?**
  - http, cookies, logging.
- **Servers**
  - CGI, servlets, JSP, architecture.

1.Structure of the Internet



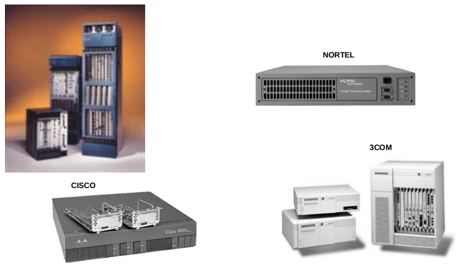
MAPS UUNET MAP

SOURCE: CISCO SYSTEMS

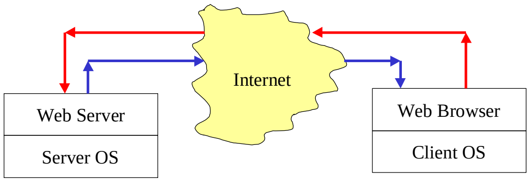
Internet Backbone Structure

- **Level 1 (interconnect level, NAPs)**
  - billions of pages per day
- **Level 2 (national backbone, MAE, FIX)**
  - Federal Internet eXchange Points
  - Peering agreements: connect, share routing info)
- **Level 3 (regional providers, state level)**
- **Level 4 (local ISP)**
- **Level 5 (companies, individuals)**
- **Level 6 (routers, see picture)**

Routers



2. Connecting on the WWW



Server-Side View

Database-driven content

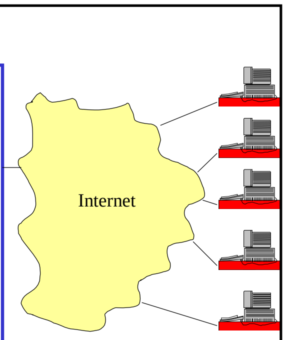
Lots of Users

- **Scalability**
- **Load balancing**

Often implemented with cluster of PCs

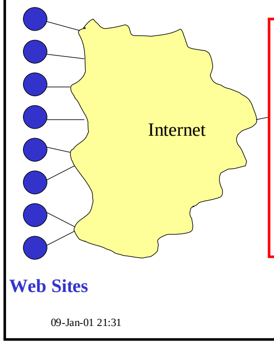
24x7 Reliability

Transparent upgrades



Clients

Client-Side View



Content rendering engine
- Tags, positioning, movement

Scripting language interpreter
- Document object model
- Events
- Programming language itself

Link to custom Java VM

Security access mechanisms

Plugin architecture + plugins

Trade-offs in Client/Server arch.

- **Compute on servers?**
  - + easy to control, debug.
  - – Peak load, reliability, cost.
- **Compute on clients?**
  - – Many different browsers
  - – {Netscape, IE, Lynx, …} × Version × OS
  - – Each supports different tags, languages…
- **Peer-to-peer? Gnutella?**

2. What happens when you click?

- **Suppose**
  - – You are at **www.yahoo.com/index.html**
  - – You click on **www.grippy.org/mattmarg/**
- **Browser uses DNS => IP addr for** www.grippy.org
- **Opens TCP connection to that address**
- **Sends HTTP request:**

```
Get /mattmarg/ HTTP/1.0                          Request
User-Agent: Mozilla/2.0 (Macintosh; I; PPC)
Accept: text/html; */*
Cookie: name = value                             Request
Referer: http://www.yahoo.com/index.html         Headers
Host: www.grippy.org
Expires: …
If-modified-since: ...
```

HTTP Response

```
                                                  Status
HTTP/1.0 200 Found                                Response
Date: Mon, 10 Feb 1997 23:48:22 GMT               1st header
Server: Apache/1.1.1 HotWired/1.0
Content-type: text/html                           Image/jpeg, ...
Last-Modified: Tues, 11 Feb 1999 22:45:55 GMT
```

- **One click => several responses**
- **HTTP1.0: new TCP connection for each page elt**
- **HTTP1.1: KeepAlive - several requests/connection**

Response Status Lines

- **1xx Informational**
- **2xx Success**
  - – 200 Ok
- **3xx Redirection**
  - – 302 Moved Temporarily
- **4xx Client Error**
  - – 404 Not Found
- **5xx Server Error**

HTTP Methods

- **GET** -Bring back a page
- **HEAD** -Like GET but just return headers
- **POST**
  - – Used to send data to server to be processed (e.g. CGI)
  - – Different from GET:
    - A block of data is sent with the request, in the body, usually with extra headers like **Content-Type:** and **Content-Length:**
    - Request URL is not a resource to retrieve; it's a program to handle the data being sent
    - HTTP response is normally program output, not a static file.
- **PUT, DELETE, ...**

Cookies

- **Small piece of info**
  - Sent by server as part of response header
  - Stored on disk by browser and returned in request header
  - May have expiration date (deleted from disk)
- **Associated with a specific domain & directory**
  - Only given to site where originally made
  - Many sites have multiple cookies
  - Some have multiple cookies per page!
- **Most Data stored as name=value pairs**
- **See**
  - **C:\Program Files\Netscape\Users\default\cookies.txt**
  - **C:\WINDOWS\Cookies**

Logging Web Activity

- **Most servers support “common logfile format” or “extended logfile format”**
- **Apache lets you customize format**
- **Every HTTP event is recorded**
  - Page requested
  - Remote host
  - Browser type
  - Referring page
  - Time of day
- **Applications of datamining logfiles !**

3. Servers

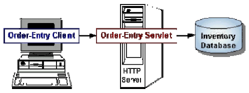
- **We want to do more via an http request – we’d like to invoke code to run on the server.**
- **Initial solution: Common Gateway Interface (CGI) programs.**
- **Example: web page contains form that needs to be processed on server.**

CGI Code

- **CGI scripts can be in any language.**
- **A new process is started (and terminated) with each script invocation (overhead!).**
- **Improvement I: run some code on the client’s machine (e.g., catch missing fields in the form).**
- **Improvement II: Server APIs (but these are server-specific).**

Java Servlets

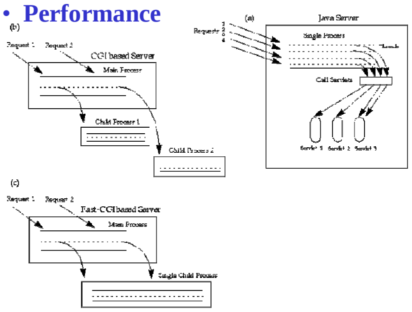
- **Servlets : applets that run on the server.**
  - Java VM stays, servlets run as threads.
- **Accept data from client + perform computation**
- **Platform-independent alternative to CGI.**



- **Can handle multiple requests concurrently**
  - Synchronize requests - use for online conferencing
- **Can forward requests to other servers**
  - Use for load balancing

Servlet Advantages

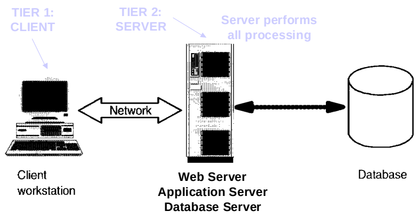
- **Java advantages (platform indep, memory alloc)**
- **Performance**



Java Server Pages (JSP)

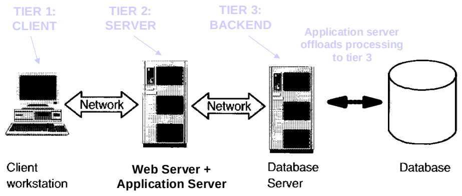
- **Sun's answer to Microsoft's ASP.**
- **Allows mixing static HTML w/ dynamically generated content.**
- **JSP is more convenient than servlets for the above purpose.**
- **Run the example at** **http://java.sun.com/products/jsp/html/jspbasics.fm.html**

Two-Tier Architecture



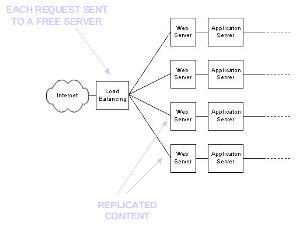
Server does too much work. Weak Modularity;

Three-Tier Architecture

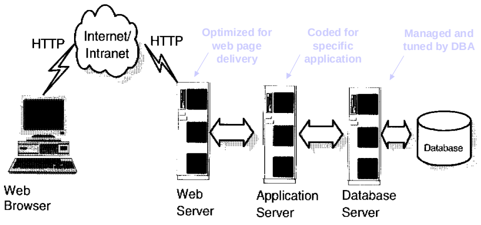


Note: Using 2 computers instead of 1 can result in a huge increase in simultaneous clients. Depends on % of CPU time spent on database access. While DB server waits on DB, Web server is busy!

Server Clusters (Farms)

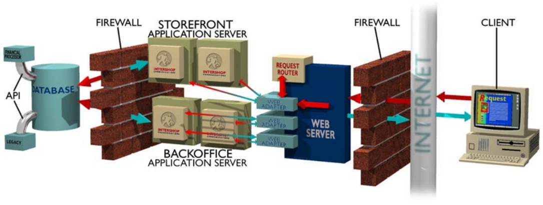


N-Tier Architecture

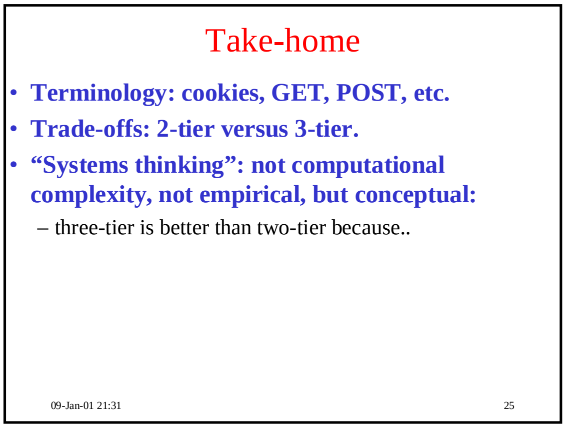


Achieves full separation of function and administration
Huge number of simultaneous clients

InterShop Architecture



SOURCE: INTERSHOP



Take-home

- **Terminology: cookies, GET, POST, etc.**
- **Trade-offs: 2-tier versus 3-tier.**
- **“Systems thinking”: not computational complexity, not empirical, but conceptual:**
  - three-tier is better than two-tier because..