

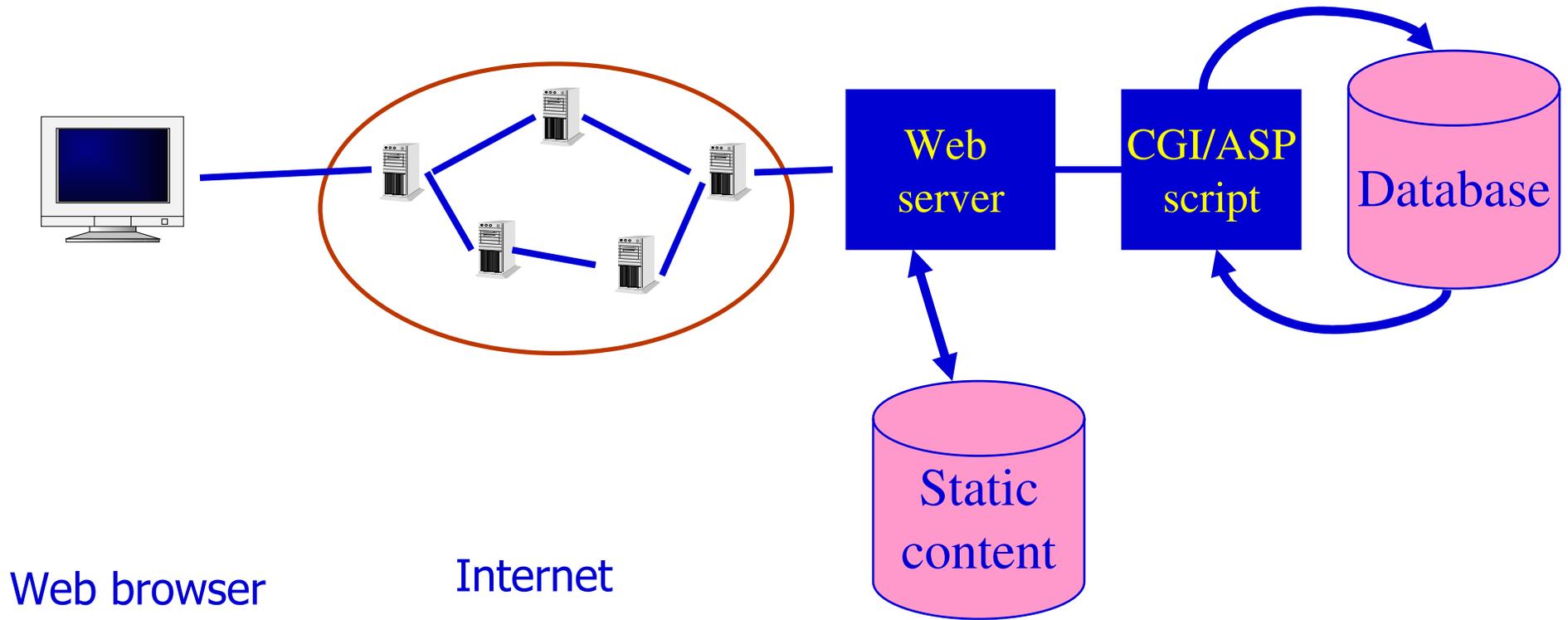
15.561  
Information Technology Essentials

**Session 11**  
**Under the hood of**  
**a commercial website**

# Outline

- Issues for building enterprise-class systems
- Typical commercial website architecture: Travelocity
- How much does it cost?

# The story so far...



# Issues for building enterprise-level eCommerce sites

- **Security**
  - Physical security
  - Access control
- **Availability/Fault Tolerance**
  - Ensure the computer services remain available to users in the face of partial failures
- **Performance/Scalability**
  - Ensure that response time remains acceptable as site traffic grows
- **Content management**
  - Ensure that content development is done in a streamlined and orderly fashion

# Enterprise-level eCommerce sites: Summary of Technologies

- **Security**
  - Physical security: Hosting
  - Access control: Firewalls
- **Availability/Fault Tolerance**
  - Replication
- **Performance/Scalability**
  - Replication
  - Load Balancing
  - Web Caching
- **Content management**
  - Content Management Software

# **Issue #1: Network and Physical Security**

# Types of firewalls

- **Packet filter:** Looks at each packet entering or leaving the network and accepts or rejects it based on user-defined rules.
- **Application gateway:** Applies security mechanisms to specific applications, such as FTP and Telnet servers.
- **Proxy server:** Intercepts all messages entering and leaving the network. The proxy server effectively hides the true network addresses

# Web Hosting

- Ensure 24x7 site operation
- Provide access to network bandwidth
- Provide physical site security

# The various flavors of hosting

- **Free hosting**
  - Your site on a shared server. Usually a shared domain name. Usually supported by advertising
- **Shared hosting**
  - Your site (and domain name) on a shared server in a special facility with staff to respond when machines go down, etc. Often bundled with email.
- **Dedicated hosting**
  - Your site(s) on a separate, dedicated server in a special facility. Varying amounts of additional services (such as design of server configuration, firewalls, replication, etc.)
- **Collocated hosting**
  - Similar to dedicated hosting except you own the server rather than renting it.

# There is no single right answer

- **Select level of hosting based on what kind of company you are**
  - basic eCommerce presence → shared hosting
    - » Century 21
  - some in-house expertise, high volume → dedicated hosting
    - » Land'sEnd, Vanguard
  - eCommerce pioneer → collocated hosting
    - » Yahoo, Amazon

## Issue #2: Ensuring Availability and Fault Tolerance

- Why do computers crash?
  - Hardware errors
  - Operating system errors
  - Application errors
  - Human errors
- Use redundancy to restore normal operation after crashes
  - Data redundancy
  - Active Replication

# Data Replication

- Keep several copies of same data (replicas)
- If one server is down, query next server
- Can improve response when load is heavy
- Problem: How to synchronize replicas?

# Active Replication

- Establish redundant copies of vital programs and servers
  - process groups
  - every group member operates on its own replica
- Every message is processed by all group members
  - members remain in mutually consistent states
- If one member fails, other members can still respond

## **Issue #3 Scalability: Why you should care**

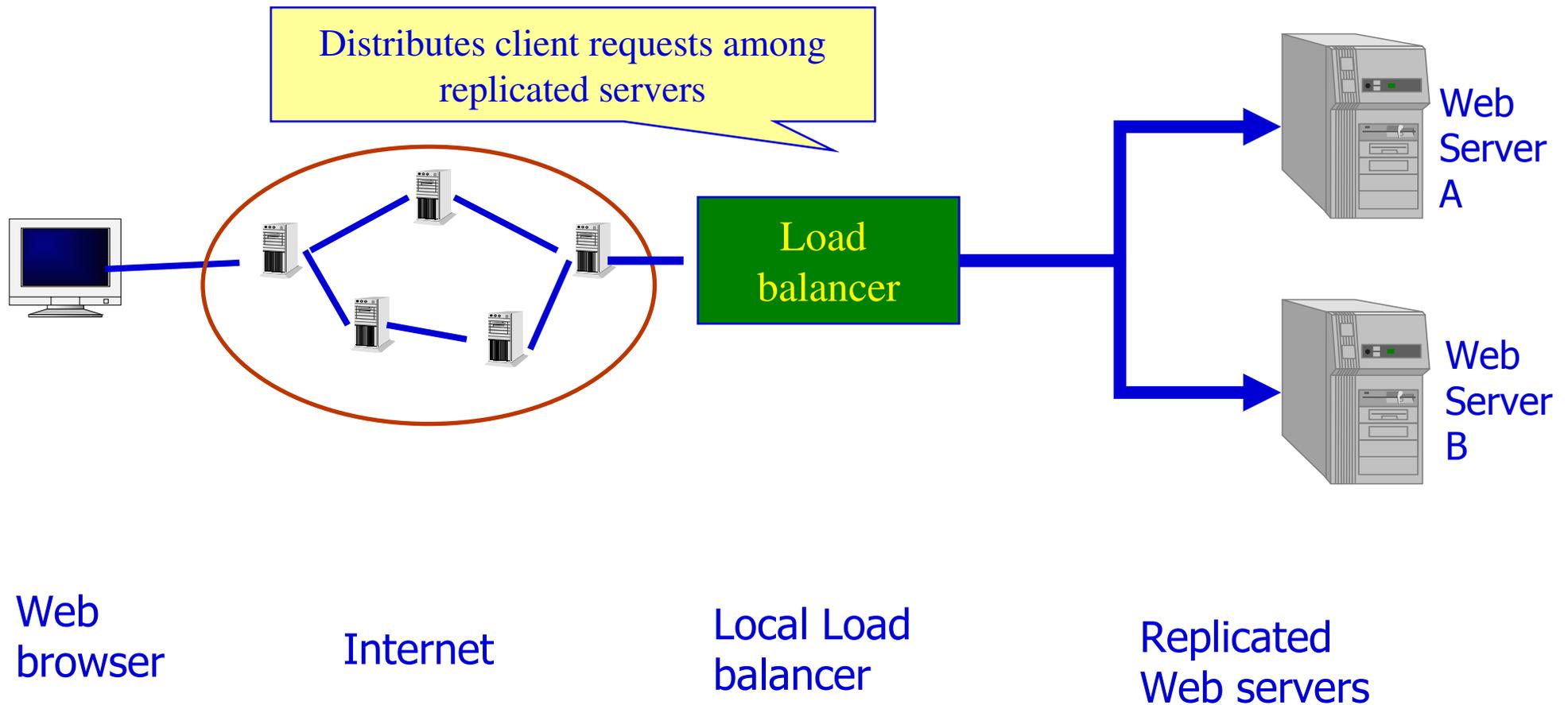
- **... the No.1 reason that customers got fed up and took their business elsewhere was technical problems, including unacceptably slow response times.**
  - Fortune magazine
- **... 28% of Netizens that encountered glitches, left the site never to return**
  - Business Week

# Technological Alternatives

- Local load balancer
- Site mirroring
- Content routing

# Load balancing solutions

Companies: Cisco, HydraWeb



# Intelligent content routing

The image shows a screenshot of the Creative Commons website in a Microsoft Internet Explorer browser window. The address bar shows <http://creativecommons.org/>. The page features the Creative Commons logo, a navigation menu with categories like Audio, Images, Video, and Text, and several main content blocks. A blue callout bubble labeled "Rich content" points to the "Find" section, which includes an icon of a document with a CC logo and a person. Another blue callout bubble labeled "Main page (Text)" points to the "Recent Features" section, which includes a news item about the 1st annual Fall Fundraiser and a mention of "The Lonely Island".

Address: <http://creativecommons.org/>

Donate today to the [Creative Commons 2005 Fundraising Campaign](#)

**creative commons** Worldwide

Creative Commons is a nonprofit organization that offers a flexible copyright for creative work.

- Audio**  
music, sounds, speeches...
- Images**  
photos, illustrations, designs...
- Video**  
movies, animations, footage...
- Text**  
blogs, essays, articles...

**Find**  
Music, photos, and more

**Publish**  
Your stuff, safely and legally

[Learn](#)

Creative Commons offers a [flexible range of protections and freedoms](#) for authors and artists. We have built upon the "reserved" of traditional copyright to create a voluntary "some rights reserved" copyright. We're a nonprofit. All of our tools are free.

**Recent Features**

**Creative Commons announces its 1st annual Fall Fundraiser**  
OCT 2005 [Creative Commons needs your help](#) to reach our 2005 goal today!

**The Lonely Island**  
OCT 2005 Andy Samberg, Jorma Taccone and Akiva Schaffer are the members of [The Lonely Island](#) comedy collective, who have released much of their music and video shorts online under a Attribution-NonCommercial-ShareAlike license.

"Rich content"

Main page (Text)

"Rich content"

# Intelligent content routing

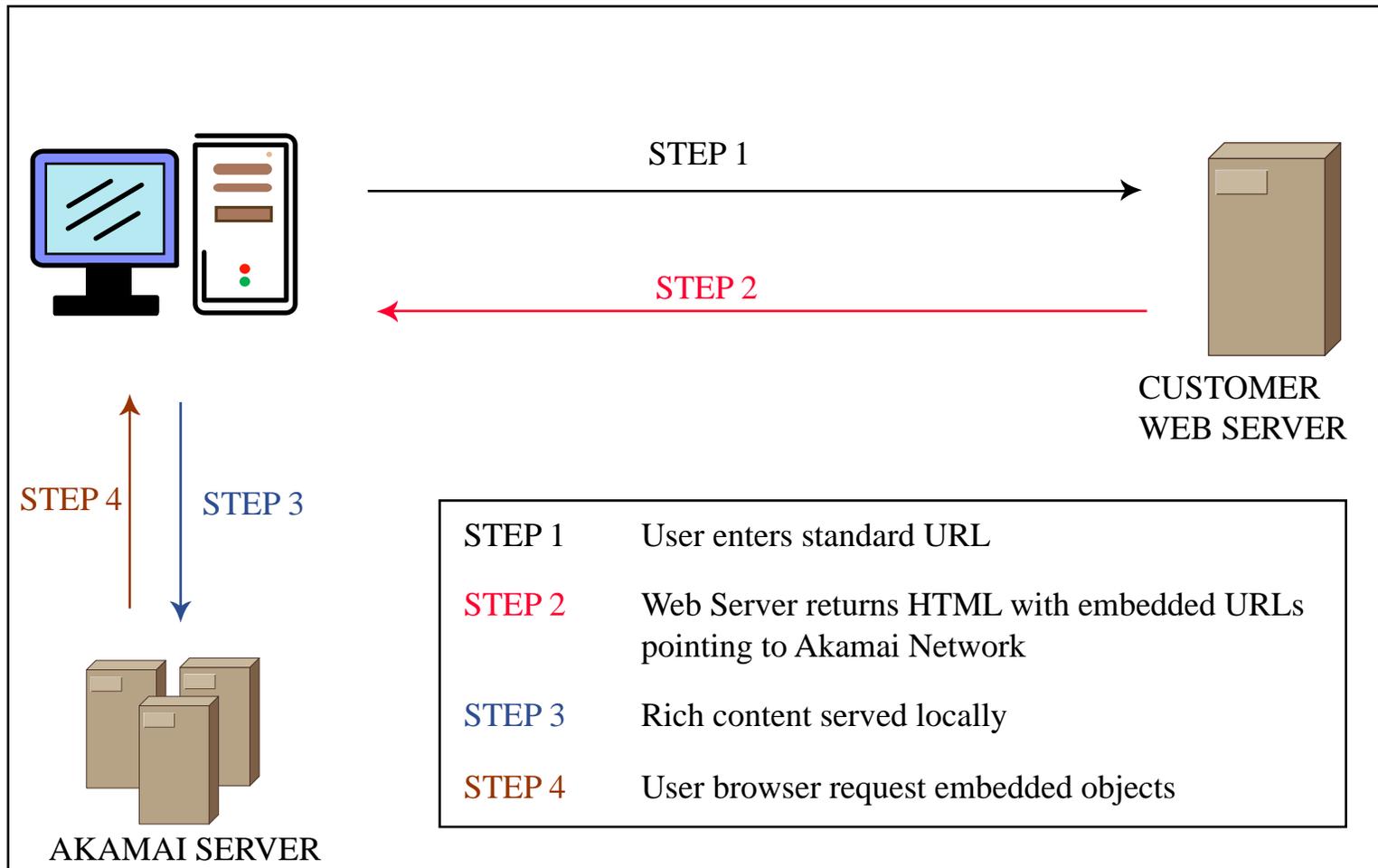


Figure by MIT OCW.

Companies: Akamai, Sandpiper

# Akamai network

- Over 15,000 servers in over 65 countries

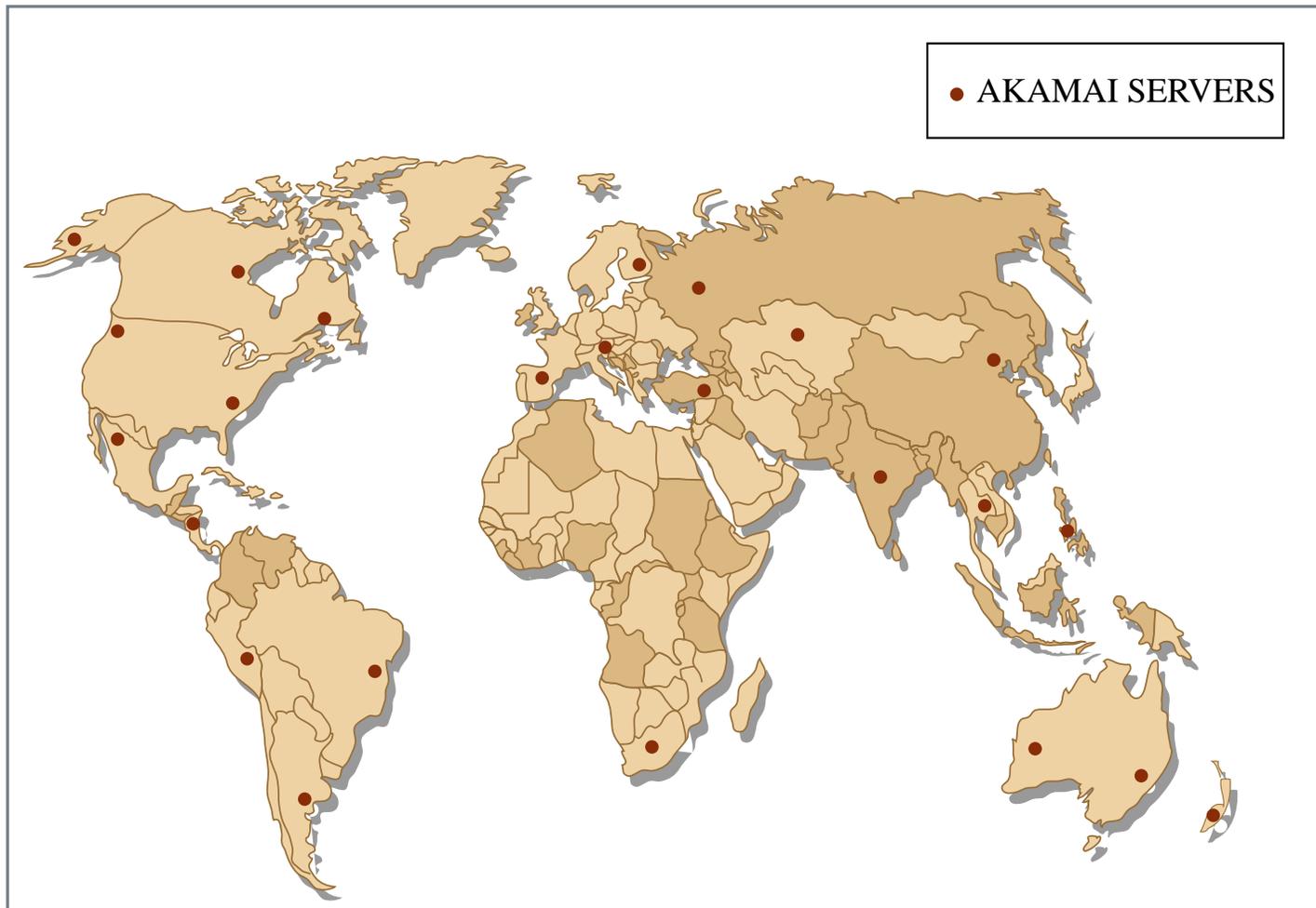


Figure by MIT OCW.

## **Issue #4: Large-scale content development**

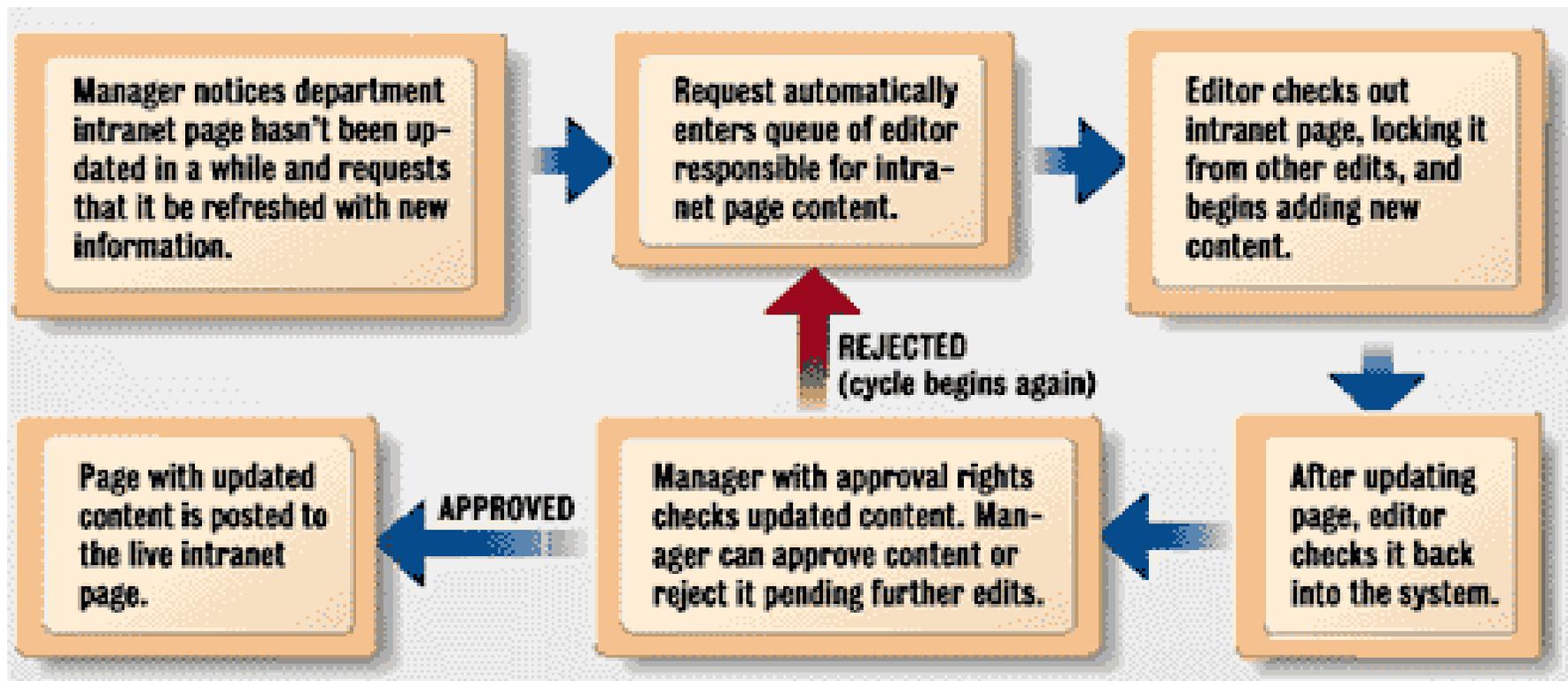
- **Large number of authors contributing site content**
- **Diverse types of content (e.g., image, video, and other media files)**
- **Need for regular content posting and replacement (i.e., weekly sales promotions)**
- **Often one or more approvals are required before content is posted**
- **Some content needs to be personalized or tailored to match the needs and interests of a site visitor**

## **Additional issues**

- **Often support for multiple languages and time zones is required**
- **Content presentation must consistently conform to branding and appearance standards**
- **Version archiving and an audit trail**
- **Content must be viewable across a variety of browsing devices, not just PCs**

# Solution: Content management systems

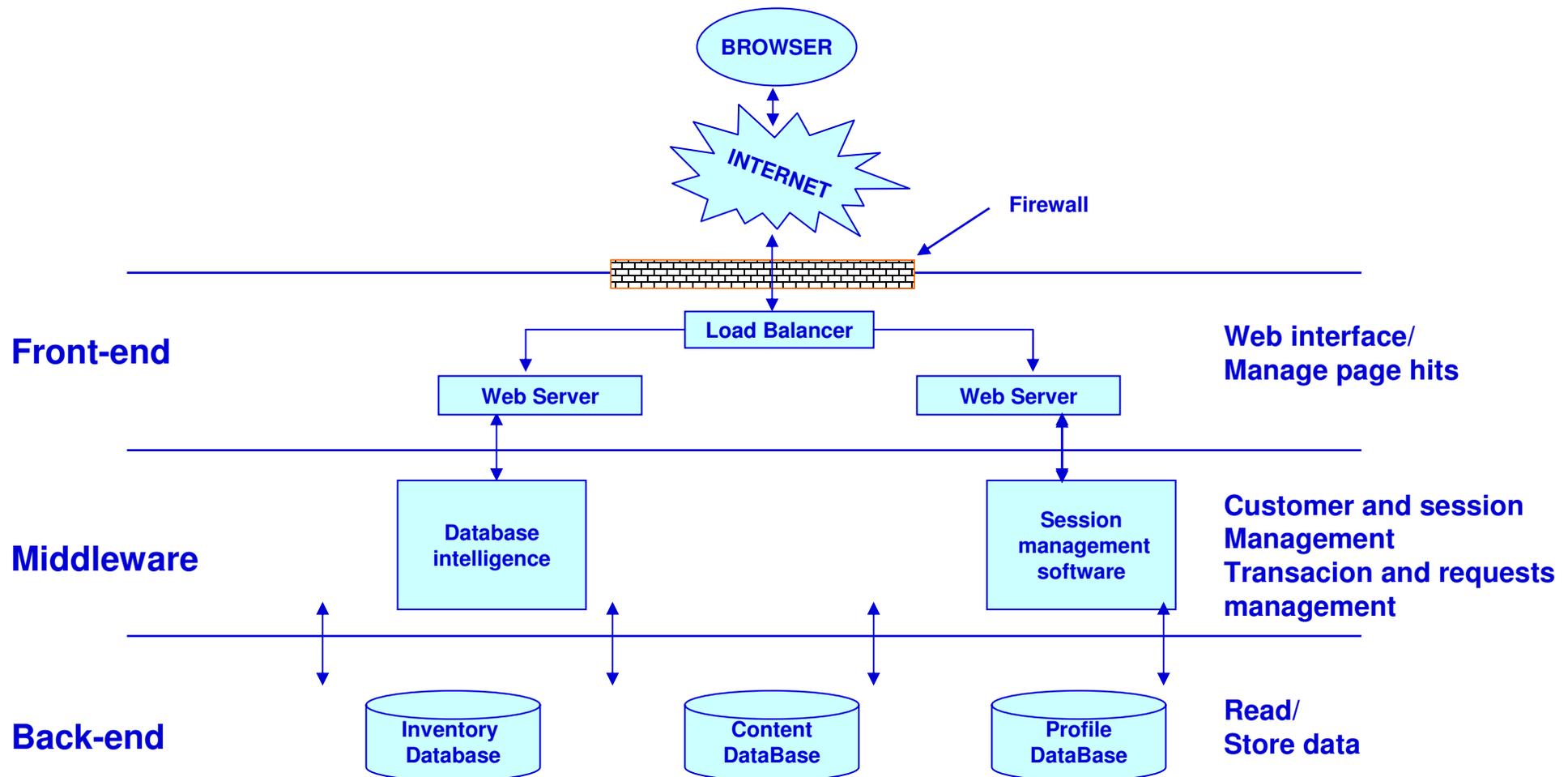
- Automatic support for content management workflows



# Enterprise-level eCommerce sites: Summary of Technologies

- **Security**
  - Physical security: Hosting
  - Access control: Firewalls
- **Availability/Fault Tolerance**
  - Replication
- **Performance/Scalability**
  - Replication
  - Load Balancing
  - Web Caching
- **Content management**
  - Content Management Software

# A Three Layer Website Architecture



# How much does it cost?

- Typical e-Commerce corporate site
- 600,000 unique visitors per month
- Growth aimed at 3 million visitors per month in two years

# Typical Hardware Costs

## Startup Costs

Servers	\$129,000	33%
Network Hardware	\$115,050	29%
Labor	\$124,800	32%
Facilities	\$26,000	7%
<b>Total</b>	<b>\$394,850</b>	

## Annual Operating Costs

Maintenance	\$10,400
Hosting	\$65,000
Labor	\$204,750
<b>Total</b>	<b>\$280,150</b>

# Typical Software Costs

## Startup Costs

Software Applications	\$310,250
Web design	\$1,105,000
Enterprise Integration	\$364,000
Other labor	\$403,650
Facilities	\$78,000
<b>Total</b>	<b>\$2,260,900</b>

## Annual Operating Costs

Maintenance	\$7,150
Labor	\$890,500
<b>Total</b>	<b>\$897,650</b>

# Total Costs

## Startup

Software	\$2,260,900	85%
Hardware	\$394,850	15%

**Total** **\$2,655,750**

## Operating

Software	\$897,650	76%
Hardware	\$280,150	24%

**Total** **\$1,177,800**

## **Moral of this lecture**

**Computer Systems can be Fast, Cheap, or Reliable**

**– Choose any two**