Session 11
Under the hood of a commercial website

Acknowledgments:
Adapted from Chris Dellarocas, U. Md.
Outline

• Issues for building enterprise-class systems

• Typical commercial website architecture: Travelocity

• How much does it cost?
The story so far…

Web browser

Internet

Web server

CGI/ASP script

Database

Static content
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

Distributes client requests among replicated servers

Web browser
Internet
Local Load balancer
Replicated Web servers
Intelligent content routing
Intelligent content routing

STEP 1
User enters standard URL

STEP 2
Web Server returns HTML with embedded URLs pointing to Akamai Network

STEP 3
Rich content served locally

STEP 4
User browser request embedded objects

Companies: Akamai, Sandpiper

Figure by MIT OCW.
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

Manager notices department intranet page hasn’t been updated in a while and requests that it be refreshed with new information.

Request automatically enters queue of editor responsible for intranet page content.

Editor checks out intranet page, locking it from other edits, and begins adding new content.

REJECTED (cycle begins again)

Page with updated content is posted to the live intranet page.

Manager with approval rights checks updated content. Manager can approve content or reject it pending further edits.

After updating page, editor checks it back into the system.
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

- **Front-end**
  - Web Server
  - Load Balancer
  - Web interface/
    Manage page hits

- **Middleware**
  - Database intelligence
  - Session management software
  - Customer and session
    Management
  - Transaction and requests
    management

- **Back-end**
  - Inventory Database
  - Content Database
  - Profile Database
  - Read/
    Store data

- **INTERNET**
  - Firewall
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

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Servers</td>
<td>$129,000</td>
<td>33%</td>
</tr>
<tr>
<td>Network Hardware</td>
<td>$115,050</td>
<td>29%</td>
</tr>
<tr>
<td>Labor</td>
<td>$124,800</td>
<td>32%</td>
</tr>
<tr>
<td>Facilities</td>
<td>$26,000</td>
<td>7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$394,850</strong></td>
<td></td>
</tr>
</tbody>
</table>

### Annual Operating Costs

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance</td>
<td>$10,400</td>
</tr>
<tr>
<td>Hosting</td>
<td>$65,000</td>
</tr>
<tr>
<td>Labor</td>
<td>$204,750</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$280,150</strong></td>
</tr>
</tbody>
</table>
Typical Software Costs

Startup Costs

<table>
<thead>
<tr>
<th>Description</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Applications</td>
<td>$310,250</td>
</tr>
<tr>
<td>Web design</td>
<td>$1,105,000</td>
</tr>
<tr>
<td>Enterprise Integration</td>
<td>$364,000</td>
</tr>
<tr>
<td>Other labor</td>
<td>$403,650</td>
</tr>
<tr>
<td>Facilities</td>
<td>$78,000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$2,260,900</strong></td>
</tr>
</tbody>
</table>

Annual Operating Costs

<table>
<thead>
<tr>
<th>Description</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintenance</td>
<td>$7,150</td>
</tr>
<tr>
<td>Labor</td>
<td>$890,500</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$897,650</strong></td>
</tr>
</tbody>
</table>
## Total Costs

<table>
<thead>
<tr>
<th></th>
<th>Software</th>
<th>Hardware</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Startup</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>$2,260,900</td>
<td>$394,850</td>
<td>$2,655,750</td>
</tr>
<tr>
<td>Hardware</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>$2,655,750</td>
</tr>
<tr>
<td><strong>Operating</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>$897,650</td>
<td>$280,150</td>
<td>$1,177,800</td>
</tr>
<tr>
<td>Hardware</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>$1,177,800</td>
</tr>
</tbody>
</table>
Moral of this lecture

Computer Systems can be Fast, Cheap, or Reliable

– Choose any two