In e-commerce sites, “success is measured by the number of users accessing the site and the ability of the site to satisfy users requests promptly”

“30% of users will leave a Web site if the downtime is 7 seconds and the abandonment rate can reach 70% as times exceed 12 seconds”
Introduction

- Three-tier architecture
- Load Balancer
  - DNS Round Robin
  - DNS Round Robin with AAlarms
  - Hardware
- Page Caching Techniques
  - Simple
  - Hot Page
  - Size Limit
DNS Round Robin

<table>
<thead>
<tr>
<th>URL (SITE NAME)</th>
<th>IP ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>299th.topcities.com</td>
<td>203.45.67.1</td>
</tr>
<tr>
<td>299th.topcities.com</td>
<td>203.45.67.2</td>
</tr>
<tr>
<td>299th.topcities.com</td>
<td>203.45.67.3</td>
</tr>
</tbody>
</table>

DNS Round Robin URL Mapping
DNS Round Robin with Asynchronous Alarms

DNS Round Robin URL Mapping

<table>
<thead>
<tr>
<th>URL (SITE NAME)</th>
<th>IP ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>299th.topcities.com</td>
<td>203.45.67.1</td>
</tr>
<tr>
<td>299th.topcities.com</td>
<td>203.45.67.2</td>
</tr>
</tbody>
</table>

Utilization Script

AALARM Message

Overloaded Message

Client

URL 299th.topcities.com

NETWORK

ROUTER

DNS SERVER

203.45.67.1

203.45.67.2
Hardware Load Balancers
Server Side Page Caching

- Page caching could reduce traffic loads on WS and possibly reduce the response time

- Page Validation
  - Most common validator is “Last-Modified”
    - Contained in the file header
    - If-Modified-Since request

- Caching Policies
  - Simple, Size Limit and Hot Page
Simple Caching Policy

- Caches all pages whenever possible
- “Can only yield a hit ratio up to 30%-50%”
- Most frequently accessed pages not always cached
Size Limit Page Caching

- Caching pages with a size lower than a “pre-determined threshold”
- Improves the hit ratio because more small files are able to be cached
Hot Page Caching

- DNS Server calculates page access averages
- DNS Server calculates “hot threshold”
- Pages exceeding the “hot threshold” become “hot pages” and are stored in the cache
- “Only a small portion of files are accessed frequently and around one-third are accessed just once”
Overall Performance of Caching and Load Balancing

- Page caching techniques can reduce the loads on web servers while possibly reducing the response time
  - Caching Hot Pages can significantly improve the load balance
  - Caching with a Size Limit does not significantly improve the load balance, and may skew the load
Conclusion

- Usability Study
  - Current User-base
  - Prospective User-base
- Consider Scalability during design phase
  - Load Balancing Solutions will integrate seamlessly
Works Referenced


