HP technical case study: the design and architecture of the 2007 Boston Marathon website

Executive summary

This white paper details the technical architecture and configurations implemented for the public website supporting the 2007 Boston Marathon race. It provides background information on the 2006 website's architecture and highlights changes made to the 2007 website. The paper also provides detailed configuration diagrams as well as a sampling of race day performance data.

Target audience: This white paper is a technical document designed to be read by system architects and other professionals wishing to understand how the 2007 Boston Marathon website was architected.

Introduction

Boston Marathon background

The Boston Marathon is the world's oldest annual marathon, and is run by the Boston Athletic Association (BAA). The Boston Marathon website, http://www.baa.org/, is an Internet website designed to provide real time information about the race and its participants to many thousands of concurrent web users during the week of the actual race. Several organizations and companies were involved with the design and support of the website:

- **The Boston Athletic Association (BAA)**: Overall responsibility for the management of all Boston Marathon activities including the Boston Marathon website.
- **Information-Overload**: Contracted by the BAA and responsible for the Boston Marathon website development, day-to-day project management and the coordination of computer-related race week activities.
- **Versatile Communication Inc.**: Responsible for architecting and implementing the computer network backbone that supports the Boston Marathon website.
- **HP**: Responsible for the design of the multi-tiered hardware and networking architecture used by the 2007 Boston Marathon website. HP was also responsible for proof testing the 2007 Boston Marathon website in its Houston labs prior to the actual race to ensure the website had sufficient scaling and failover capacity. Once the website had been thoroughly tested by HP, it was migrated to the production site.
- **Northeast Data Vault (a division of Conversent Communications):** Provided co-location services, backups, disaster recovery, migration services, management services and was the secure hosted site.
- **F5 Networks**: Responsible for architecting the load balancing, firewall and virtual private network (VPN) support for the website.

After having excellent and successful years from 2004 through 2006, it was time for the website to consider upgrading the architecture with newer system technology. The 2007 architecture would consist of the latest HP BladeSystem servers, HP ProLiant servers and HP StorageWorks 1000 Modular Smart Array (MSA1000) storage array hardware. Software upgrades were also considered, tested and implemented.

The HP Dynamic Internet Solutions Architecture (DISA) would continue to be implemented along with minor changes in the external storage for the back-end clustered database because it provides availability, reliability, scalability and manageability. DISA is a well established architecture that has been implemented all over the world by both small companies and large enterprises.

Internet backbone

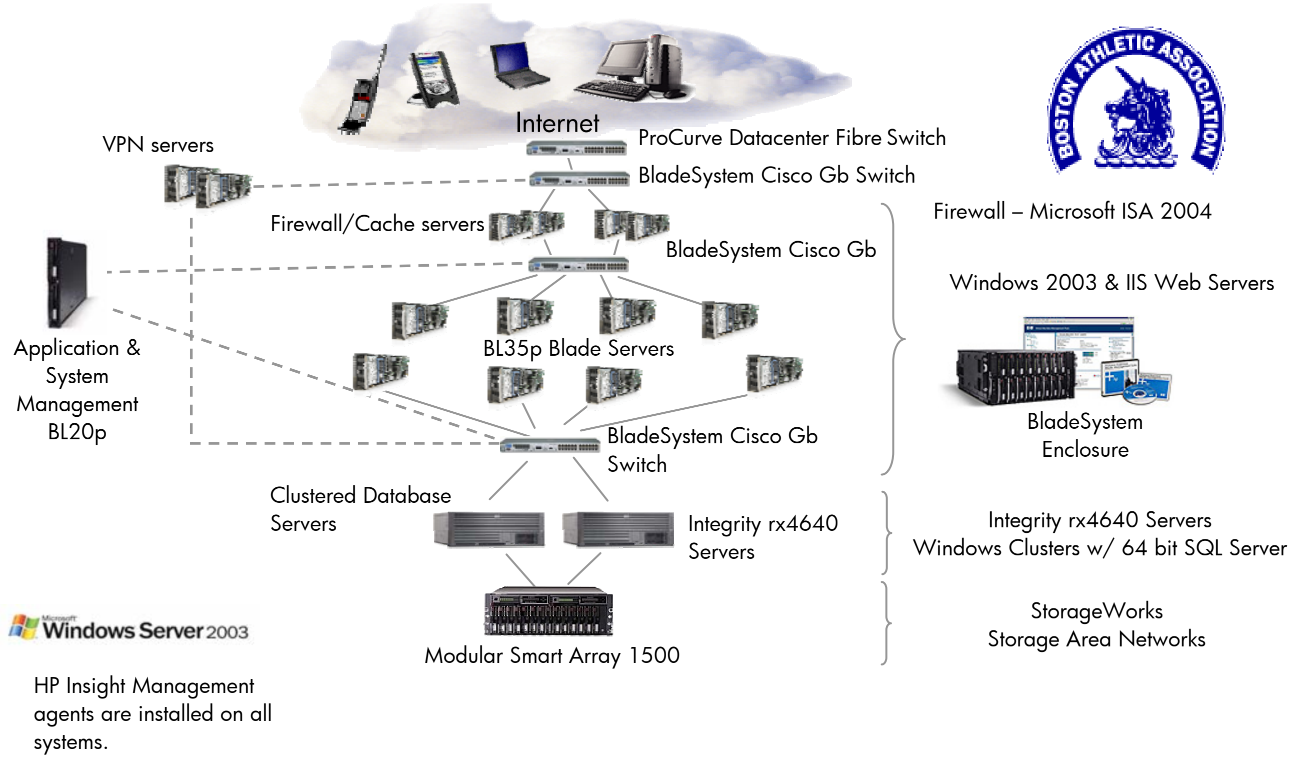
Redundant 1Gb Fibre Channel (FC) trunks connect the rack to the Internet at Conversant. The FC trunks were configured as Active/Passive, and they fail back if the primary fails and then become active again. Testing showed fail-over at 1-2 seconds and fail-back at 4-6 seconds.

Redundancy

- 1Gb Fibre Trunk (Connection to the World Wide Web [WWW])
- NIC Teaming
- Power Supplies (All hardware capable of supporting redundant power supplies were so configured)
- Microsoft® Windows® Cluster Service (Back-end databases were clustered)

2006 Boston Marathon website architecture

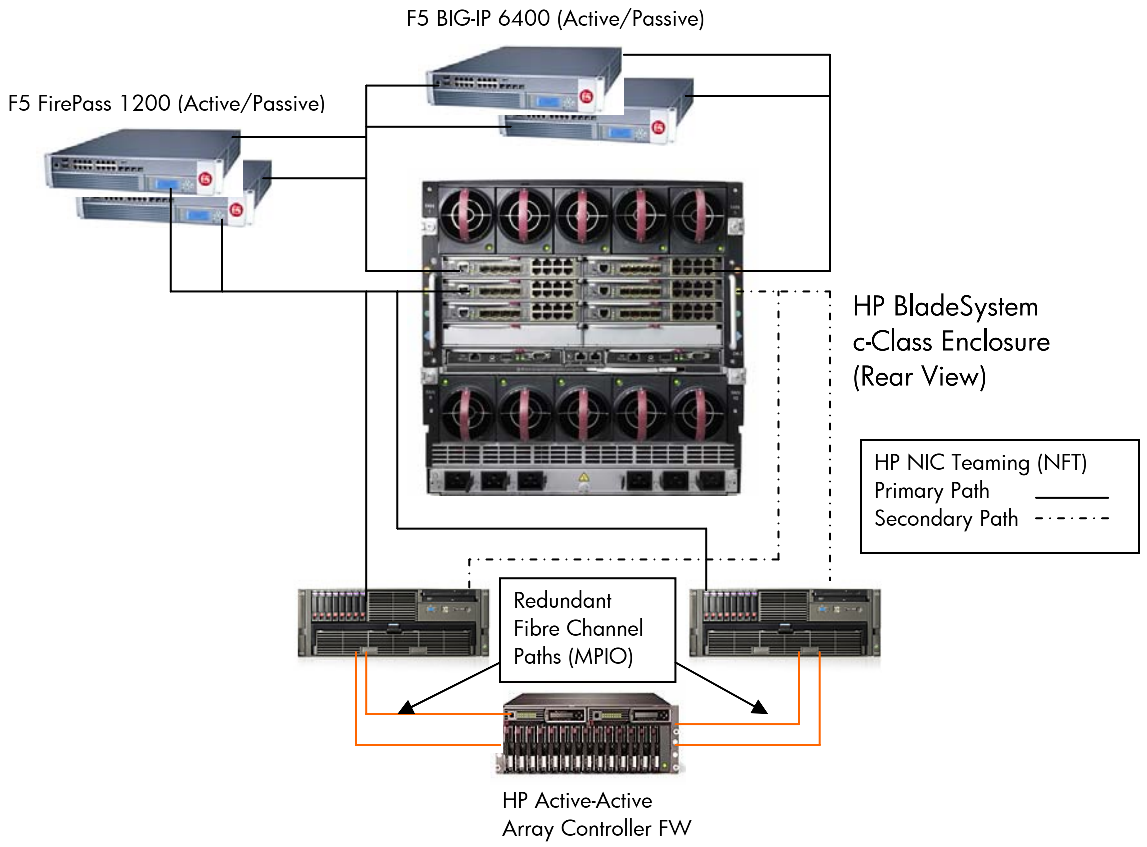
Figure 1. Boston Marathon 2006 website architecture


Internet
ProCurve Datacenter Fibre Switch
BladeSystem Cisco Gb Switch
BOSTON ATHLETIC ASSOCIATION
VPN servers
Firewall/Cache servers
BladeSystem Cisco Gb
Firewall – Microsoft ISA 2004
Windows 2003 & IIS Web Servers
Application & System Management BL20p
BL35p Blade Servers
BladeSystem Enclosure
BladeSystem Cisco Gb Switch
Clustered Database Servers
Integrity rx4640 Servers
Integrity rx4640 Servers
Windows Clusters w/ 64 bit SQL Server
Microsoft Windows Server 2003
Modular Smart Array 1500
StorageWorks
Storage Area Networks
HP Insight Management agents are installed on all systems.

2007 Boston Marathon website architecture

Figure 2. Boston Marathon 2007 website architecture

Website Networking Diagram



Website configuration details

Database

The back-end database consisted of two HP ProLiant DL585 G2 servers and one StorageWorks 1000 Modular Smart Array (MSA1000) storage device. The ProLiant DL585 G2 hardware configuration consisted of 4 dual-core CPUs, 8GB RAM, 2 internal 72GB SAS RAID1 disks, two HBAs, two integrated NICs, one NC360T PCI Express Dual Port NIC and redundant power supplies. The MSA1000 hardware configuration consisted of 10 72GB SCSI 15K disks configured with Advanced Data Guarding (ADG) and two on-line spares, redundant array controllers configured as Active-Active, redundant power supplies, and redundant Fibre Channel (FC) switches.

The two ProLiant DL585 G2 servers utilized Microsoft clustering services for high availability, and Microsoft SQL Server 2000 was configured Active/Active. The first instance of SQL ran on NODE1 of the cluster, and the second SQL instance ran on NODE2; each node of the cluster provided backup or redundancy to the other in the event of a failure. Database failover time in this case was 40-45 seconds. Both databases contain identical data as participant updates are written back-to-back.

Participant updates were first logged to a database at the Fairmont Copley Plaza Hotel – limited numbers of people had access to this, including the Media. Updates were then made to the two clustered SQL instances which were hosted by this site for public access.

The two ProLiant DL585 G2 servers were equipped with two HBAs, each providing redundant fibre channel (FC) paths to the MSA1000. The redundant HBAs were connected to the redundant FC 2/8 switches of the MSA1000 storage array providing redundant FC paths. HP Multipath Input/Output (MPIO) Basic Failover v1.0 for MSA1000/1500 was installed to fully automate and utilize the redundant FC paths in the event of a failure at the HBA, MSA FC controller, FC cable or FC switch.

The DL585 G2 servers each utilized the integrated NICs and one port of the dual-port NC360T NIC; the NC360T active port was dedicated to the cluster heartbeat, while the integrated NICs were configured with HP NIC teaming and connected to two HP GbE2c Ethernet Blade Switches for c-Class BladeSystem in bays 3 and 4 of the HP BladeSystem c7000 enclosure.

Note: When implementing an Active-Active SQL cluster, for performance reasons, neither of the nodes in the cluster should ever exceed 50% of its resources to allow resource capacity in the event a fail-over does occur.

Web

The web layer consisted of seven HP ProLiant BL460c servers, each with 2 processors, 2GB RAM, two 72GB SAS hot plug disks configured as RAID1 and four NICs (two integrated and one dual port Ethernet mezzanine card). The integrated NICs were connected to the Active and Passive BIG-IP 6400s. HP NIC Teaming was used to team the NICs of the mezzanine card and connected to the GbE2c switches in bays 3 and 4 of the c7000 blade enclosure.

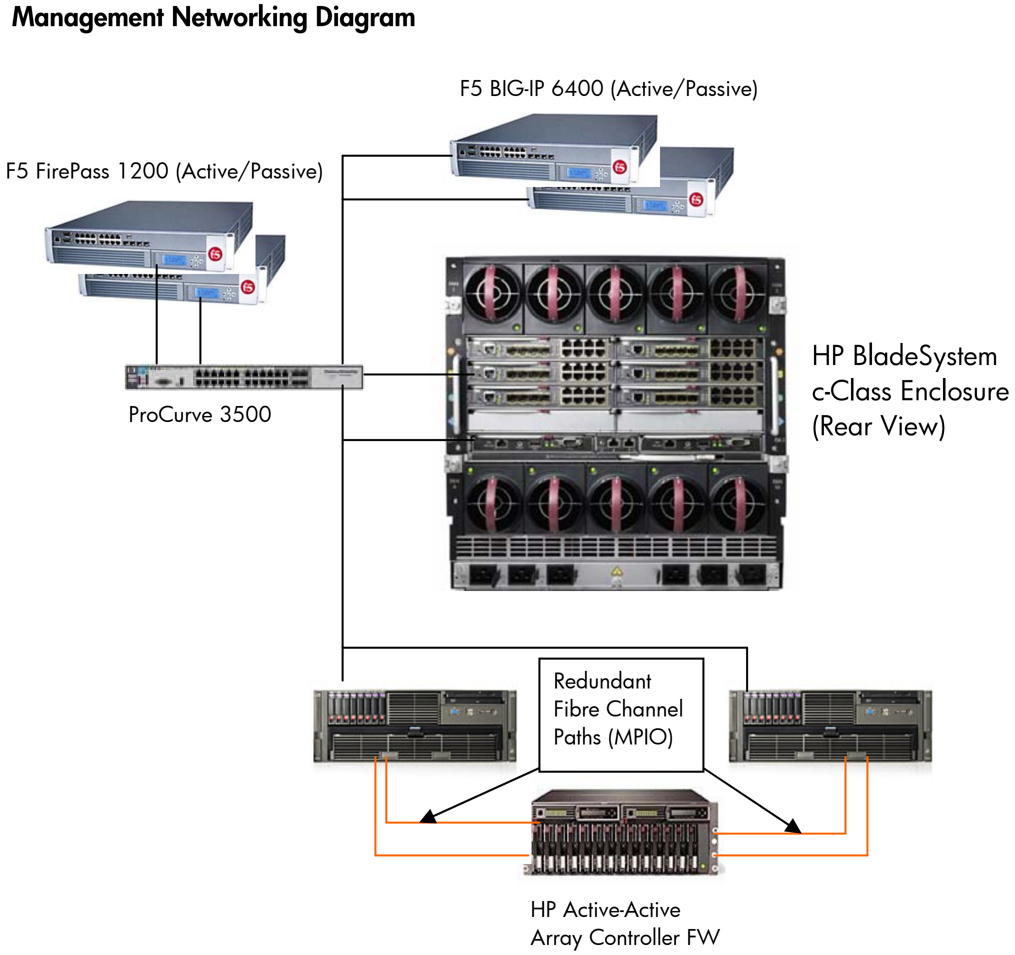
The ODBC implementation for web servers 1-4 was configured to access database instance 1 (DB01) while web servers 5-7 were configured to access database instance 2 (DB02).

Firewall/load balancer

The firewall/load balancing layer consisted of two F5 BIG-IP 6400 devices in an Active-Passive configuration. The front side of the BIG-IP was connected to the datacenter (public side), and the back side was connected to the web layer.

2007 Boston Marathon management architecture

Figure 3. Boston Marathon 2007 Management Architecture


Management Networking Diagram
F5 BIG-IP 6400 (Active/Passive)
F5 FirePass 1200 (Active/Passive)
HP BladeSystem c-Class Enclosure (Rear View)
ProCurve 3500
Redundant Fibre Channel Paths (MPIO)
HP Active-Active Array Controller FW

VPN

The two VPN systems were F5 FirePass 1200 series implemented in an Active-Passive configuration. The front side of the FirePass was connected to BIG-IP 6400 (public side), and the back side was connected to the database layer. The VPN uses Secure Sockets Layer (SSL) and was configured to allow only four specific logins with appropriate credentials to utilize the VPN.

The primary pre-race function was to remotely manage and monitor the health of the site. In addition, the VPN connection provides access to the web and database layers, allowing the synchronization of this site with the production BAA.org site before cutting over and going live on March 30th, with plans to continue through 2011.

Management

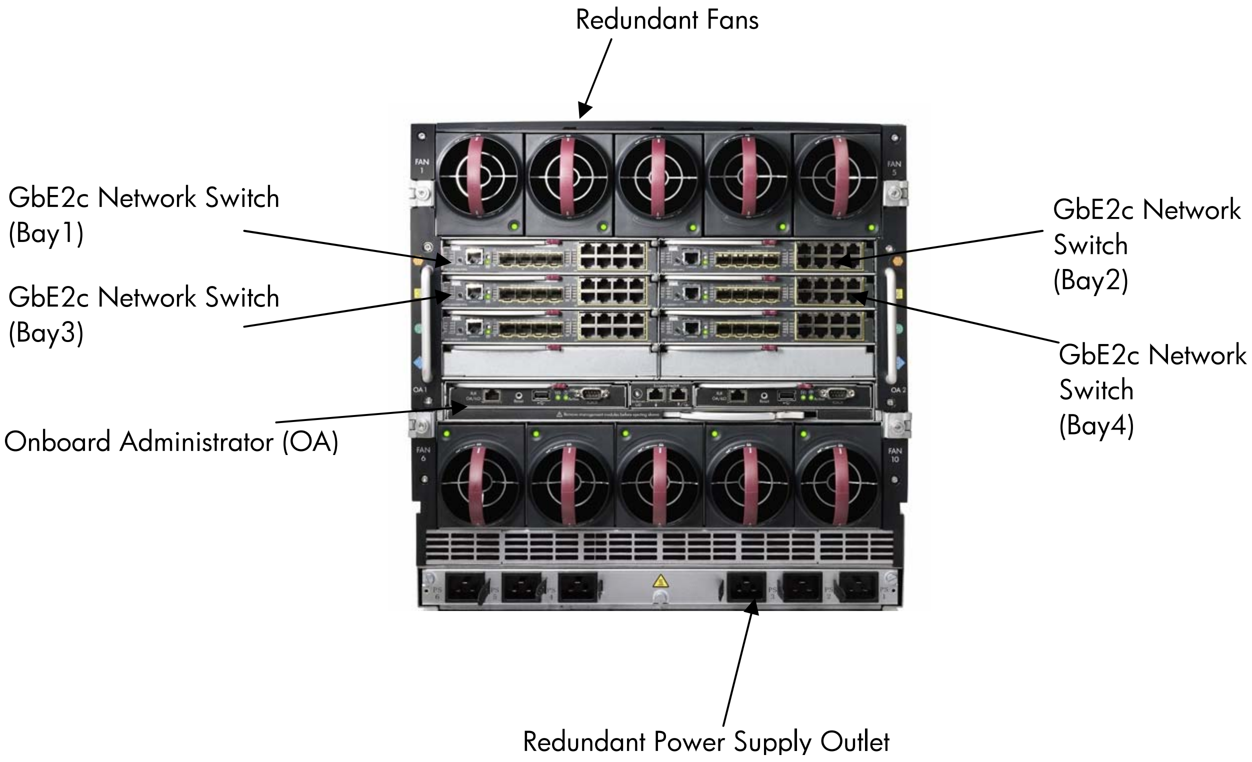
The management system consisted of an HP ProLiant BL460c server with 2 CPU, 2GB RAM, 2 x 72GB SAS disks (RAID1), 2 integrated NICs and one dual port Ethernet mezzanine card. The two integrated NICs were connected to GbE2c switches located in the c7000 bays 1 and 2, and the mezzanine card NICs were connected to switches located in bays 3 and 4. This allowed for management at all layers and at the physical switches within the c7000 enclosure.

The HP Integrated Lights-Out 2 (iLO 2) management ports of the BladeSystem enclosure's Onboard Administrator (OA) and of the ProLiant DL585 G2 servers were connected to the management network. The management system utilized HP Systems Insight Manager (HP SIM) 5.1 with Service Pack 1 to monitor both the management agents and iLO of all systems. The management system also utilized HP ProLiant Essentials Rapid Deployment Pack (RDP) to deploy the system during build-out.

To monitor for availability at the application and database layers, a simple batch routine was created using a small executable from Microsoft called "tinyget5.exe". The routine performs an HTTP GET to a small Adobe® ColdFusion *.cfm file created by the ColdFusion developer that utilized Microsoft Internet Information Services (IIS), the WWW, ColdFusion and the database. If the expected result is not returned, an error message is logged to a file and displayed on the screen. An additional Microsoft Visual Basic (VB) script was written to send an e-mail alert in the event of an error message from the "tinyget5.exe" executable. For this project, manual monitoring and reaction in the event of an error was used, but this could be further automated to isolate and eliminate the failed application server.

c7000 BladeSystem enclosure

Figure 4. c7000 BladeSystem Enclosure (Rear view)


Redundant Fans
GbE2c Network Switch (Bay1)
GbE2c Network Switch (Bay3)
Onboard Administrator (OA)
GbE2c Network Switch (Bay2)
GbE2c Network Switch (Bay4)
Redundant Power Supply Outlet

Hardware summary

Table 1. Server hardware information

Server function	# of servers	Server type	CPU	Memory	Disk/RAID	NICs	Redundant power supplies	Additional hardware
Firewall/ Cache	2	F5 BIG-IP 6400	2	2GB	1	1 x 1GB Integrated switch	No	NONE
ColdFusion Web application	7	BL460c	2 x 3.2 GHz	2GB	2 x 72GB SAS/RAID0	2 x Integrated 1 x Mezzanine (dual port)	YES (BladeSystem Enclosure)	NONE

Server function	# of servers	Server type	CPU	Memory	Disk/RAID	NICs	Redundant power supplies	Additional hardware
Clustered database	2	ProLiant DL585 G2	4 x dual-core 2.8 GHz	8GB	2 x 72GB SAS/ RAID1	2 x Integrated 1 x NC360T (dual port)	YES	1 x HP Smart Array P400 Controller 2 x Fibre Channel Host Bus Adapters
Application & System Management	1	BL460c	2 x 3.2 GHz	2GB	2 x 72GB SAS/ RAID1	2 x Integrated 1 x Mezzanine (dual port)	YES (BladeSystem Enclosure)	NONE
VPN	2	F5 FirePass 1200 series	1	512MB	1	2 x Integrated	No	NONE

Table 2. External attached storage information

Device	Type	Number of disks	RAID configuration	On-Line Spares	Fibre Channel switch(es)	Redundant power supplies
External shared storage	MSA1000	10 x 72GB SCSI	ADG (Advanced Data Guarding)	2	2	YES

Summary 2006 vs. 2007 websites

2006 website		2007 website	
Software	Hardware	Software	Hardware
ColdFusion MX 7.0 (Web Application)	8 x BL35p	ColdFusion MX 7.0 (Web Application)	7 x BL460c G1
HP SIM 5.0	1 x BL20p G2	HP SIM 5.1 w/SP1	1 x BL460c G1
HP ProLiant Essentials Rapid Deployment Pack (RDP) 2.1	1 x BL20p G2	HP ProLiant Essentials Rapid Deployment Pack (RDP) 3.0	1 x BL460c G1
Microsoft SQL Server 2000 64-Bit Active-Active Cluster	2 x Integrity rx4640 MSA1500	Microsoft SQL Server 2000 Active-Active Cluster	2 x ProLiant DL585 G2 1 x MSA1000

2006 website		2007 website	
Software	**Hardware**	**Software**	**Hardware**
HP MPIO Basic Failover for MSA1000/1500, 1.0	Redundant Fibre Channel Paths	HP MPIO Basic Failover for MSA1000/1500, 1.0	Redundant Fibre Channel Paths
N/A	HP BladeSystem Cisco Network Switches	N/A	HP BladeSystem GbE2c Switches

Performance

Load balancing/cache layer

The performance statistics were pulled from live data on the day of the race. During the peak hours of race day, the BIG-IP handled 84,501 concurrent connections, 73.5 Gbits of inbound data and 845.8 Gbits outbound data, and 4,500 requests per second with a cache hit ratio of 83.3%.

The following figures graph the BIG-IP Percent CPU Utilization, Active Sessions and Requests per Second between the hours of 9:30 AM and 5:30 PM.

Figure 5. BIG-IP % CPU Utilization

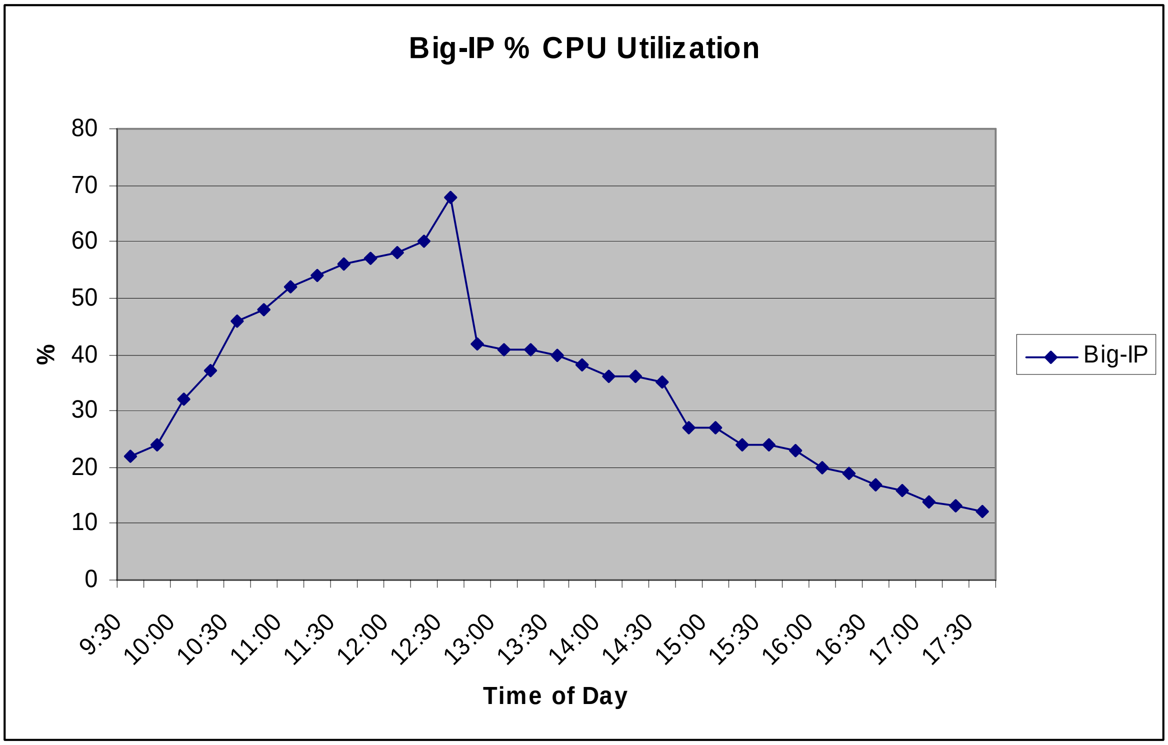


Figure 6. BIG-IP Active Sessions

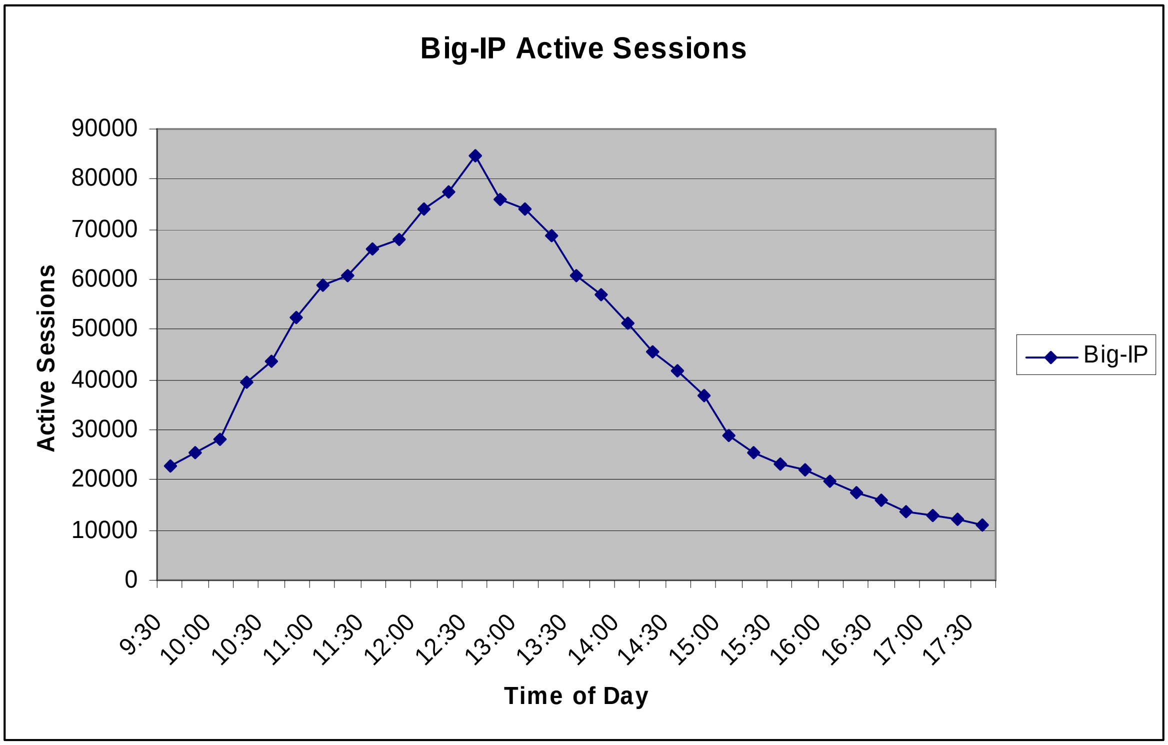
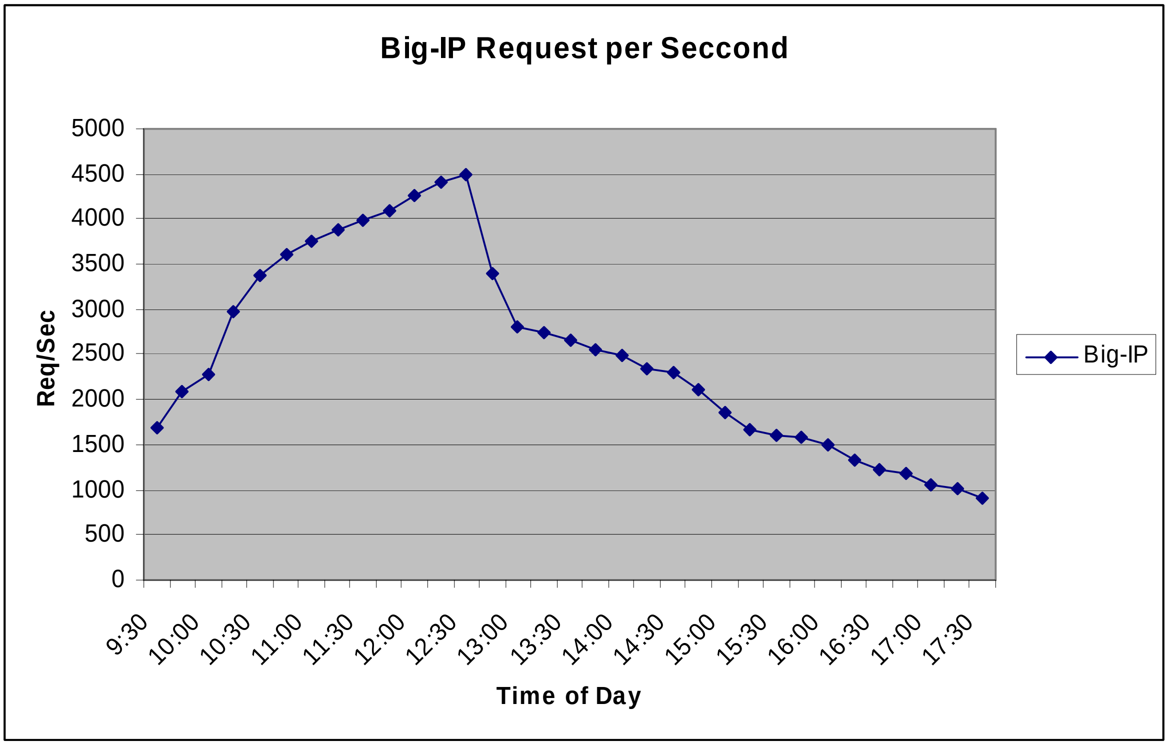


Figure 7. BIG-IP Requests per Second



This year, due to weather, it is believed more spectators viewed the race from indoors via the World Wide Web than in person alongside the course. With the large increase of web-based viewing of the race, the datacenter network segment servicing the BAA website became saturated between 11am and 3pm. The BIG-IP maximum CPU utilization peaked at 70% which was largely attributed to a 30% packet loss and retransmission of packets.

Peaks and valleys are typical of utilization statistics for web applications. There are often peak levels of utilization that are 10 to 20 times or more than the typical utilization rates; website design must accommodate peak utilization, so that when sites are most popular – and as they grow in popularity over time – they continue to provide adequate service to users. This was a core design tenet of the BAA website planning because it is a core design tenet of the HP DISA architecture. Scalability must be built into a DISA site to accommodate the level of peak usage so that when website success (high user counts) is achieved, levels of service do not drop off.

The cache hit rates declined steadily throughout the day. This provides good insight to the utilization patterns on the BAA website as race day progressed. Initially, there were very few requests per second. The initial cache hit rates were high, around 83%. This indicates that in the early hours of use, most users were asking for the same data. Especially during the first few hours of a marathon, this is entirely to be expected; the leaders complete the marathon in less than three hours, and early users of the site may have been extremely interested in seeing the results of the leaders as they progressed toward the finish line.

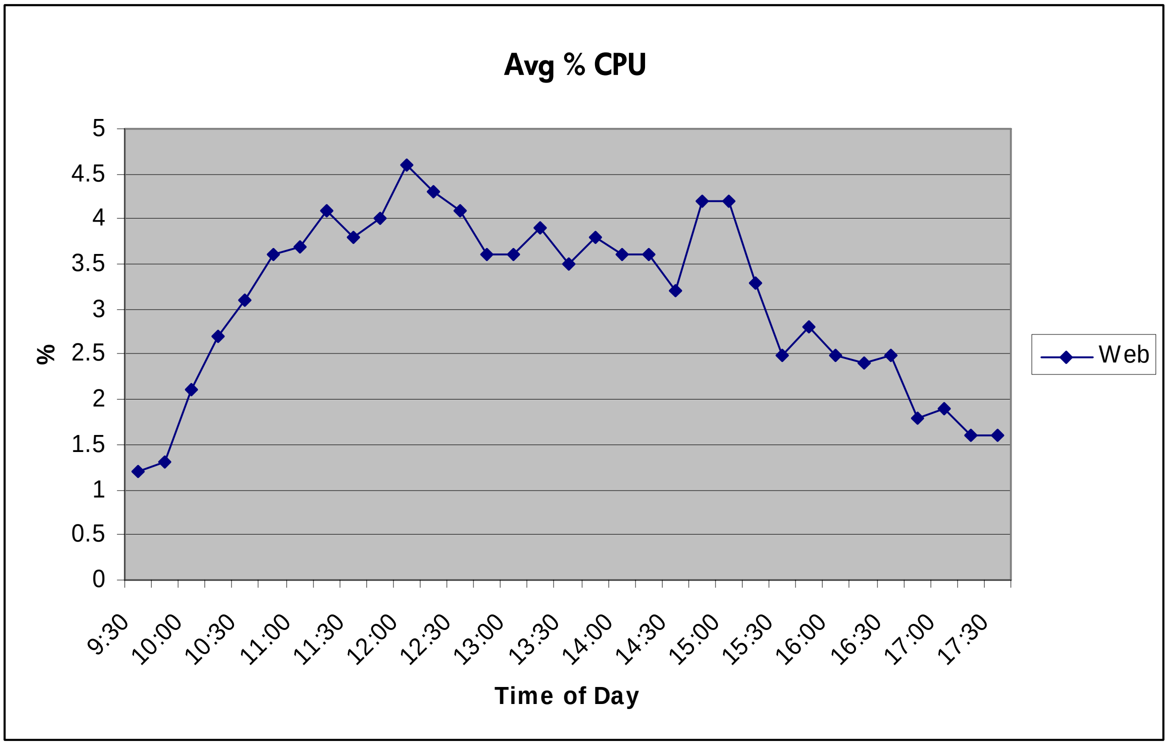
These kinds of insights can be readily derived from adequate performance and utilization monitoring of a website. Store operators can apply their deep knowledge of their businesses to the performance statistics collected from their e-commerce sites to benefit performance in terms of their businesses, allowing effective decisions to be made for both technical and business realms.

Web/application layer

The web/application layer utilized both Windows Server IIS 6.0 and ColdFusion MX 7.0. ColdFusion MX 7.0 was selected for the 2006/2007 marathon because it yielded a 57-83% increase in performance over ColdFusion 4.5 and at a lower CPU utilization. The average CPU utilization throughout the majority of race day was between 2-4% and peaked at 5%. One of the largest factors for lower CPU utilization was the 83% cache hit ratio of the BIG-IP.

The Avg % CPU is the average of the seven web/application servers. Windows performance monitor was used to gather the Avg % CPU at 15 minute intervals for each of the servers.

Figure 8. Avg % CPU – average of seven web/application servers

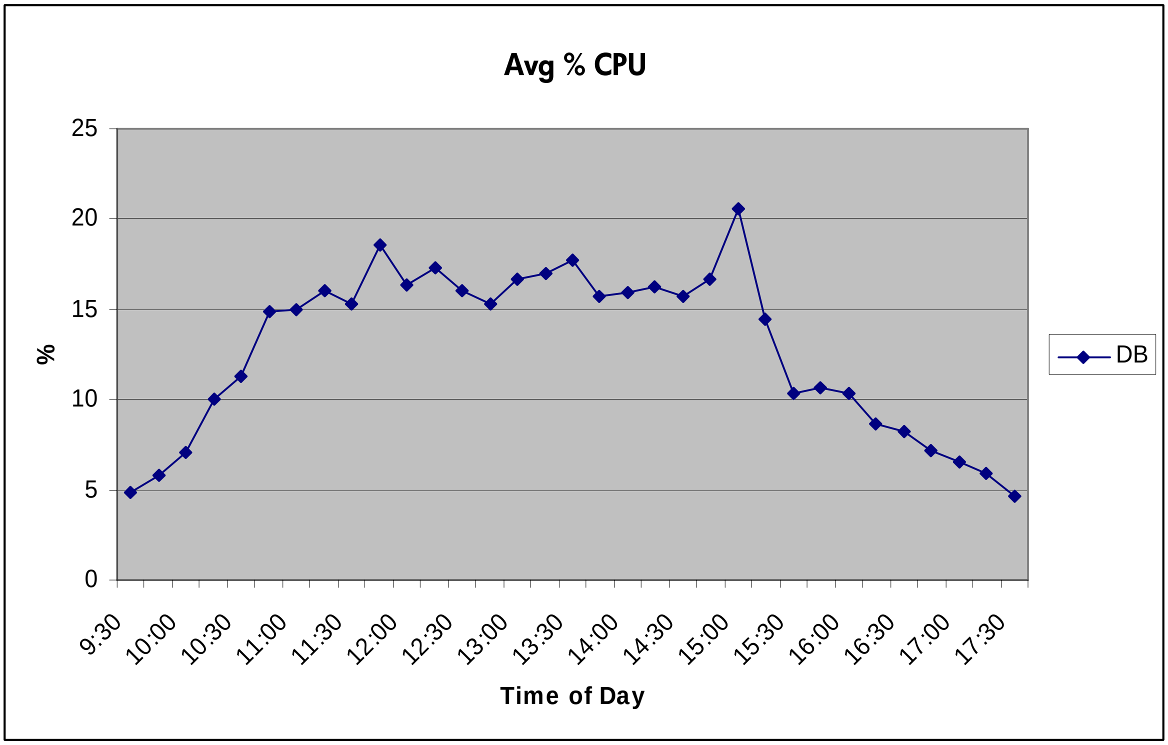


Database layer

Windows Server 2003 R2 and SQL Server 2000 were used to build an active/active SQL cluster for the back-end database on two ProLiant DL585 G2 servers. CPU utilization increased throughout the day until after 3:30 PM for two reasons (3,170,000 database queries): 1) athletes' tracking data increased for each athlete as each crossed the 13 checkpoints throughout the course thus taking more CPU cycles for queries, 2) Visitors to the site performed more searches as the day went on, tracking friends and family members and top known athletes.

The Avg % CPU is the average of the two database servers (as graphed in Figure 9). Windows performance monitor was used to gather the Avg % CPU at 15 minute intervals for each of the servers.

Figure 9. Avg % CPU – database servers



Summary

In conclusion, it is evident through this case study that HP provides customers with the technology and best practices to expertly architect very robust websites to meet the needs of a globally accessed system. The architecture designed for the Boston Marathon provided access to hundreds of thousands of global users on an annual basis. Because the system was designed with availability, reliability, scalability, and manageability in mind, it provided the type of application access that is required in many scenarios. HP, with the BAA's approval, publishes this case study as proof of its technology leadership and expertise to exceed the requirements of the BAA for a successful Boston Marathon experience.

For more information

- The Boston Athletic Association (BAA), http://www.baa.org/: Overall responsibility for the management of all Boston Marathon activities including the Boston Marathon website.
- Information-Overload, http://www.info-overload.com/: Contracted by the BAA and responsible for the Boston Marathon website development, day to day project management and the coordination of computer related race week activities.
- Versatile Communication Inc., http://www.vteg.com/: Responsible for architecting and implementing the computer network backbone that supports the Boston Marathon website.
- F5 Networks, http://www.f5.com/: Responsible for architecting the load balancing, firewall and VPN support for the website.
- HP, http://www.hp.com/: Responsible for the design of the multi-tiered hardware and networking architecture used by the 2004-2007 Boston Marathon website. HP was also responsible for proof testing the Boston Marathon website in its Houston labs prior to the actual race to ensure the website had sufficient scaling and failover capacity. This fully pre-tested site was migrated to the production site once it had been thoroughly tested by HP.
- HP.com - ProLiant servers, http://www.hp.com/go/proliant/
- HP Dynamic Internet Solutions Architecture (DISA), http://www.hp.com/solutions/disa/
- HP StorageWorks Modular Smart Array: modular array systems - overview & features, http://h18006.www1.hp.com/storage/arraysystems.html
- HP.com - ProLiant remote management with lights-out technologies (ilo), http://www.hp.com/go/ilo
- Microsoft – http://www.microsoft.com
- HP Management Solutions, http://h71028.www7.hp.com/enterprise/cache/4213-0-0-0-121.aspx

To help us improve our documents, please provide feedback at www.hp.com/solutions/feedback

4AA0-9567ENW, Revision 3, July, 2007