



IBM Software Group

IBM® WebSphere® Application Server V6.0.2

Load Balancer



@business on demand.

© 2004 - 2006 IBM Corporation
Updated October 22, 2006

This presentation will cover the changes to the Load Balancer Edge Component in the WebSphere Application Server V6.0.2 release.

Goals

- The goal is to understand the WebSphere Application Server V6.0.2 Load Balancer changes

- Prerequisite:
 - ▶ An understanding of the Load Balancer Edge Component



The goal of this presentation is to explain the changes to the Load Balancer in the V6.0.2 maintenance pack.

Agenda

- Load Balancer
 - ▶ Topology
 - ▶ Benefits
- New Features in 6.0.2
 - ▶ Deprecated functions



This presentation will first explain the concepts behind the Load Balancer component, then explain its new features and deprecated components.

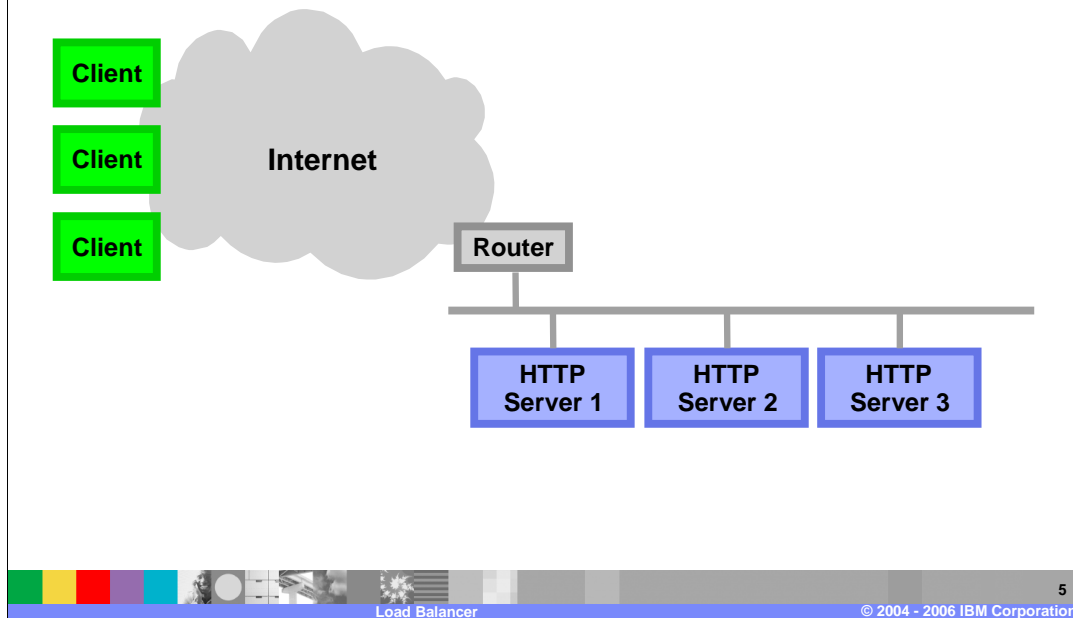
Section

Load Balancer



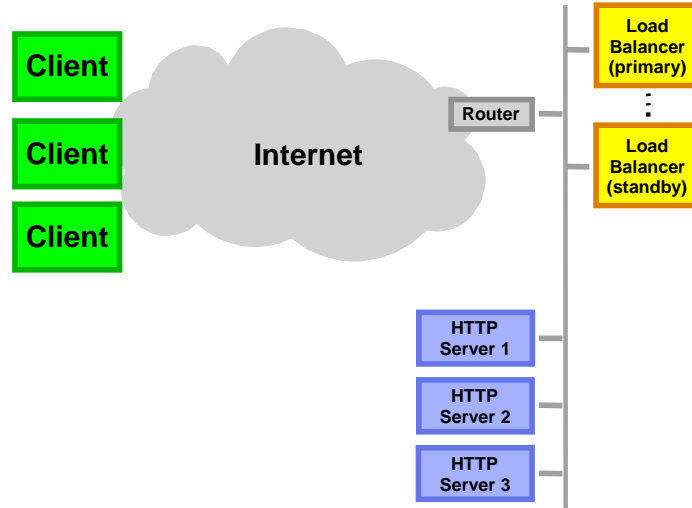
This section will explain the load balancer component.

The big picture: Without Load Balancer



Here you see a simple scenario of a Web Server topology without Edge Components. There are several problems with this model -- most importantly is the single point of failure at the HTTP server. If a Web Server is unavailable, requests may be lost.

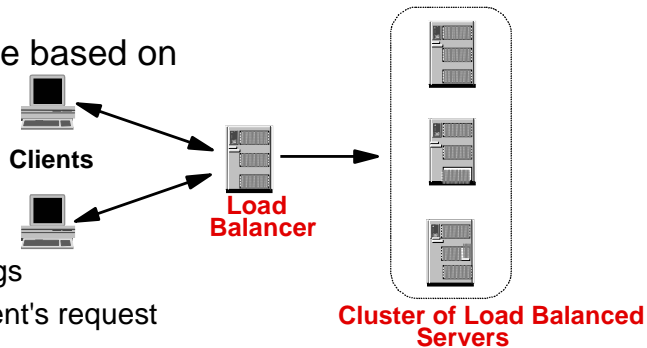
The big picture: With Load Balancer



Here you see a scenario which illustrates how content hosting sites, shown on the right, and access providers, shown on the left, can benefit from the Load Balancer. By introducing the Load Balancer it has removed the single point of failure at the Web Servers. The Load Balancer will route requests to the load balancers based on availability. There is also a second load balancer configured in case the primary load balancer should fail.

Load Balancer: Overview

- Responsible for balancing the load across multiple servers that can be within
 - ▶ A Local Area Network
 - ▶ Wide Area Network
- Load balancing can be based on
 - ▶ Dynamically computed weights
 - ▶ Domain name based round-robin
 - ▶ User-specified settings
 - ▶ Information in the client's request
- Can be configured for high availability



The Load Balancer is responsible for load balancing traffic to servers that are in a LAN or WAN. You must configure the Load Balancer to know about a number of configurations, such as:

- 1) The cluster which represents the load-balanced servers.
- 2) The port you are load balancing traffic for.
- 3) The servers that make up the load-balanced cluster, and
- 4) The load-balanced servers are configured to be part of the cluster.

Having one Load Balancer is a single point of failure, so you can configure two Load Balancers for high availability. You can also configure the load balancer to specify what the load balancing decision is based upon.

Why use the Load Balancer?

Availability: Users must reach the application regardless of failed server sites or links

- ▶ Requests are not routed to down servers
- ▶ Intelligent weighted round robin load balancing
- ▶ Servers can be removed for scheduled maintenance
- ▶ High availability feature to ensure load balancer health

Scalability: Applications need to scale for increasing numbers of simultaneous users on a wide range of web access devices

- ▶ Provides close to linear scaling
- ▶ Supports collocation of servers



If you configure the Load Balancer to monitor the load-balanced servers, requests will not be routed to unavailable servers. Keep in mind that there are application-specific advisors that monitor more than just the state of the server. Load balancing can be based on information gathered from the load balanced servers; for example, processor load, free memory, response time, and so on. System administrators can add or remove servers to a cluster without disrupting user requests. High availability provides failover support for the Load Balancer.

You can scale up the number of requests and the number of servers in the cluster.

Why use the Load Balancer?

Performance: End users demand quick response times, regardless of their geographic location or the type of content being delivered

- ▶ No packet modification performed by Dispatcher
- ▶ Allows split inbound/outbound traffic flow for network scalability

Server Awareness

- ▶ Connection, application and system monitoring

Flexibility

- ▶ Protocol independent IP load balancing
- ▶ Not limited to load balancing web servers



The Dispatcher does not modify the source or destination in the packet, so the response from the load-balanced server is sent back to the client without passing through the Dispatcher.

You can have separate networks to handle inbound and outbound traffic. This can result in improved performance for an environment. As described on the previous slide, you can monitor your servers and detect when a load-balanced server becomes unavailable. You can also load balance many different types of traffic.

Section

New features in V6.0.2



The next section details the new capabilities in the V6.0.2 maintenance pack and shows the deprecated features from previous releases.

V6.0.2 features

- IPv6/IPv4 Load Balancer
 - ▶ TCP connection routing using MAC packet forwarding
 - ▶ High Availability for no single point of failure
 - ▶ Server Health Monitoring / failure detection
- Local command line interface



The primary update to the Load Balancer in the V6.0.2 release is to support the Internet Protocol version 6 standard. This standard specifies larger 128 bit addresses for the Internet. The load balancer supports TCP connection routing and server health monitoring to detect failure. The load balancer is supported with a command line interface to configure and monitor behavior.

V6.0.2 deprecated components

- IPv4 only Load Balancer
- SiteSelector – DNS load balancing not strategic to WebSphere
- Proxy Content Based Routing – replaced by WebSphere Secure Caching Proxy
- Cisco/Nortel controllers – out of date
- See also:

<http://www-306.ibm.com/software/webservers/appserv/doc/v602/ec/infocenter/edge/LBguide.htm>



This slide details the functionality from past versions of the Load Balancer that is now deprecated.

Section

Summary

And, in summary...

Summary

- Discussed the uses for the Load Balancer
- Explained the new capabilities of the V6.0.2 Load Balancer



This presentation has discussed the benefits of using a Load Balancer, and outlined some of the capabilities that it provides. It also explained the changes to the Load Balancer component in V6.0.2.

Trademarks, copyrights, and disclaimers

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

IBM	CICS	IMS	MQSeries	Tivoli
IBM (logo)	Cloudscape	Informix	OS/390	WebSphere
e(logo)/business	DB2	iSeries	OS/400	xSeries
AIX	DB2 Universal Database	Lotus	pSeries	zSeries

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a registered trademark of Linus Torvalds.

Other company, product and service names may be trademarks or service marks of others.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

Information is provided "AS IS" without warranty of any kind. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

© Copyright International Business Machines Corporation 2004 - 2006. All rights reserved.

Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.