



IBM Software Group

IBM® WebSphere® Extended Deployment V6

The On-Demand Router



@business on demand.

© 2005 IBM Corporation
Updated August 16, 2005

This presentation will describe the On-Demand Router, a component provided by WebSphere Extended Deployment.

Agenda

- On-Demand Router overview
- Functional details



First, an an overview of the On-Demand Router will be provided, followed by a discussion of functional details.

Section

Overview

This section will give an overview of the On-Demand Router.

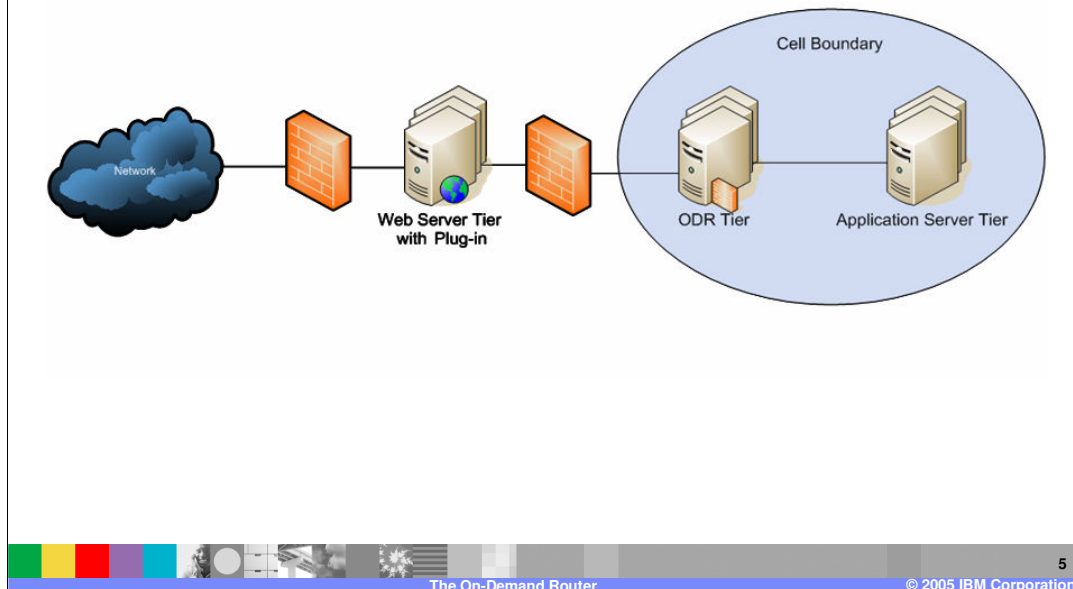
On-Demand Router Overview

- The On-Demand Router (ODR) is an intelligent HTTP proxy server
 - ▶ Enhanced version of the Proxy Server from ND 6.0.2
 - ▶ Can replace or complement the HTTP server plug-in
 - ▶ Prioritizes requests and controls traffic flow according to operational policy
 - Ensures that quality of service policies are observed
 - Flow control features ensure that application servers are not overloaded
 - ▶ Integrates with application placement to route requests to dynamic cluster members



The On-Demand Router is an intelligent HTTP proxy server that is provided with WebSphere XD. It is the point of entry into a WebSphere XD environment, and is responsible for request prioritization, flow control, and distributing requests to application servers. These advanced features distinguish the ODR from both the HTTP server plug-in and the Proxy Server provided with WebSphere Application Server Network Deployment, and give the ODR the ability to ensure a more consistent quality of service for your enterprise applications. It can be used in place of, or in concert with the HTTP server plug-in, depending on your needs. While it is most often associated with WebSphere XD's dynamic operations features, other features, such as interruption-free application upgrades and Deployment Manager high availability also rely on the On-Demand Router.

Topology Including the On Demand Router



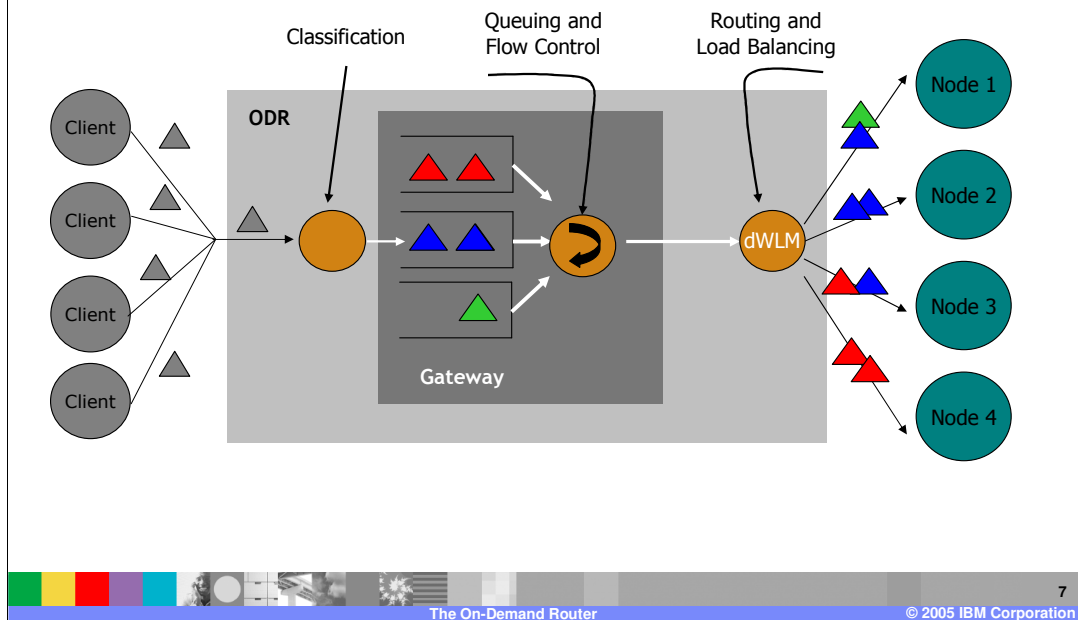
While the ODR can be used in place of the HTTP server plug-in, many WebSphere XD environments utilize both components. Often the HTTP server remains in the demilitarized zone, serving static content and acting as a point of entry into the private network. The plug-in uses a configuration file generated by WebSphere XD to route requests to one or more ODRs. The ODRs then dynamically route the HTTP requests to a Dynamic Cluster member in the application server tier.

Section

Functional Details

This section will discuss the functional details of the On-Demand Router.

ODR Functional Components Overview



This diagram illustrates the flow of an HTTP request through the ODR. When a request enters the ODR, it is classified according to rules that you have defined. These rules can be based on many criteria, including Uniform Resource Identifier, HTTP headers, cookies, or the client's IP address. When a rule is matched, the request is placed into a queue with all other requests that are in the same class of service. The ODR can drain these queues at different speeds, to ensure that requests in each class are able to meet their goals. This can mean holding less important requests in the queues for longer than more important requests, so that more resources are available to serve the important requests. As requests are released from the queues, they are routed to application servers by the dynamic workload management component, which is constantly aware of the location of each dynamic cluster member. Each member has a dynamically assigned weight, which can increase or decrease the number of requests it receives based on how quickly it is responding to requests.

Other XD 6.0 ODR Enhancements

- Classifies incoming requests based on custom service policies or routing policies
 - ▶ XD 5.1 ODR was limited to classifying by Uniform Resource Identifier (URI)
- Static and dynamic caching support
- SSL offload or onload is configured per-module
 - ▶ XD 5.1 had per-ODR setting
- Sticky routing to non-WebSphere servers
 - ▶ ODR sets a cookie in the response header



The On-Demand Router in WebSphere XD V6 has some new features that were not present in V5.1. Incoming requests are now classified based on customized service policies and routing policies. These policies allow classifying incoming work by a number of attributes, including user or group IDs, client IP address, or HTTP headers. WebSphere XD V5.1 was limited to classifying requests by Uniform Resource Identifier. The ODR can now also perform static and dynamic caching of Servlet and JSP data. SSL offloading can now be configured on a per-web-module basis, meaning that you can configure requests for one Web module to be terminated at the ODR and sent through to the application server unencrypted, and requests for a different web module to be forwarded all the way to the application server as HTTPS traffic. The ODR also has the capability to maintain server affinity when routing traffic to non-WebSphere servers by placing a cookie in the response header that uniquely identifies the server, so that subsequent requests can be routed to the same server.

Autonomic Request Flow Manager (ARFM)

- Dispatches work out of the queues based on class weights
- Dynamically adjusts weights
 - ▶ Achieves response time goals for different classes of requests
 - ▶ Responds to varying load conditions and request surges
 - Limits the number of requests on a given server
 - Overload protection



The ARFM is a component that runs inside of the On Demand Router. It is responsible for queuing work by service policy as requests come into the ODR, as well as for controlling the flow of requests exiting the queues. It dynamically adjusts the rates at which requests flow through to application servers to ensure that response time goals are met and that no application servers are sent more requests than can be handled.

ARFM Components

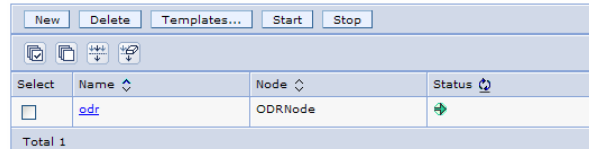
- ARFM Gateway queues and dispatches requests
 - ▶ Manages logical queues
 - ▶ Queues are dynamically created as service classes appear in the request flow
 - ▶ Acts on every request
- ARFM Controller sets gateway dispatch weights
 - ▶ Monitors request flow, execution times and application placements
 - ▶ Sets service class dispatch weights
 - ▶ Operates at scheduled intervals



The ARFM has two logical components, the gateway and the controller. The gateway is responsible for putting each incoming request into a queue that represents its service class, and later dispatching requests from the queues. The controller operates periodically, monitoring request flow, response times, and application placements, and dynamically sets the dispatch weights for each logical queue, dictating how quickly requests are released from each of them.

Configuring On-Demand Routers

- Create, modify, and manage ODRs under Servers > On Demand Routers
- Click on the name of the ODR to configure it just like a server
 - ▶ Manage server attributes like ports, thread pools, cache rules here
 - ▶ Service policies, work classes, and routing policies are not configured here



Select	Name	Node	Status
<input type="checkbox"/>	odr	ODRNode	OK

Total 1

On Demand Routers can be created and configured in the Administrative Console under the 'Servers' menu item. ODRs can be placed on any managed node; no special installation is required. While properties of the ODR itself, such as thread pool sizes or cache rules, can be configured in this panel, service policies, work classes, and routing policies are not configured here. Service policies are created under the 'operational policies' menu item, while work classes and routing policies are attributes of individual applications. A wsadmin script, 'createODR.jacl' is also provided for creating an On-Demand Router from the command line.

Summary

- The On-Demand Router is a new component provided by WebSphere XD
- Entry point for requests into a dynamic-operations based environment
- Intelligent HTTP proxy server
 - ▶ Performs request classification, queuing, and dynamic workload management



In summary, the On-Demand Router is an intelligent HTTP proxy server that is the point of entry into a dynamic-operations based environment. It provides more advanced features than the Proxy Server component that is provided with WebSphere Application Server Network Deployment. Such features include request classification, flow control, and dynamic workload management. In addition to dynamic operations, many other WebSphere XD features, including interruption-free rollout of application upgrades and Deployment Manager high availability, rely on the capabilities of the On-Demand Router.

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, 2005. 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.