



IBM Software Group

IBM® WebSphere® Extended Deployment V6.1

WebSphere Virtual Enterprise

Formerly Operations Optimization

On-demand router



@business on demand.

© 2007 IBM Corporation
Updated June 17, 2008

This presentation will cover the on-demand router component of WebSphere Extended Deployment version 6.1.

This module was originally recorded for WebSphere Extended Deployment Operations Optimization, which is now called WebSphere Virtual Enterprise. Though the module uses the previous names, the technical material covered is still accurate.

Agenda

- On-demand router in WebSphere Extended Deployment
- Functional details
- Topologies
- Administration



First, an overview of the on-demand router will be provided, followed by a discussion of functional details and how to create and administer an on-demand router .

Section

On-demand router in WebSphere Extended Deployment



This section will discuss how the on-demand router interoperates with other WebSphere Extended Deployment features.

On-demand router overview

- The on-demand router is an intelligent HTTP proxy server
 - ▶ Enhanced version of the proxy server from WebSphere Network Deployment 6.1
 - ▶ 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 Extended Deployment. It is the point of entry into a WebSphere Extended Deployment environment, and is responsible for request prioritization, flow control, and distributing requests to application servers. It can momentarily queue requests for less important applications in order to allow requests for more important applications to be handled more quickly. It is aware of the current location of a dynamic cluster's server instances, so that requests can be routed to the correct endpoint. The on-demand router can also dynamically adjust the amount of traffic sent to each individual server instance based on processor utilization and response times. The on-demand router contributes to the health of a WebSphere cell, for example it can detect and react to a "storm drain condition". A storm drain condition can occur when a backend subsystem or resource terminates unexpectedly. Until the resource becomes available again, the server can return error codes quite quickly. A simple router can falsely assume that the server is processing requests very efficiently and start sending more requests to the affected server. The server can become swamped with work when it is able to satisfy requests again. The on-demand router detects this condition and deals with the condition in cooperation with the health monitor. These and other advanced features distinguish the on-demand router from both the HTTP server plug-in and the proxy server provided with WebSphere Application Server Network Deployment, and give the on-demand router the ability to ensure a more consistent quality of service for your enterprise applications. It can be used in place of, or in conjunction with the HTTP server plug-in, depending on your needs.

On-demand router

- Required for:
 - ▶ Edition management
 - ▶ Health monitoring
 - ▶ Visualization
 - ▶ Scale out
 - ▶ Dynamic operations



While it is most often associated with WebSphere Extended Deployment's dynamic operations features, other features, such as interruption-free application upgrades and deployment manager high availability also rely on the on-demand router.

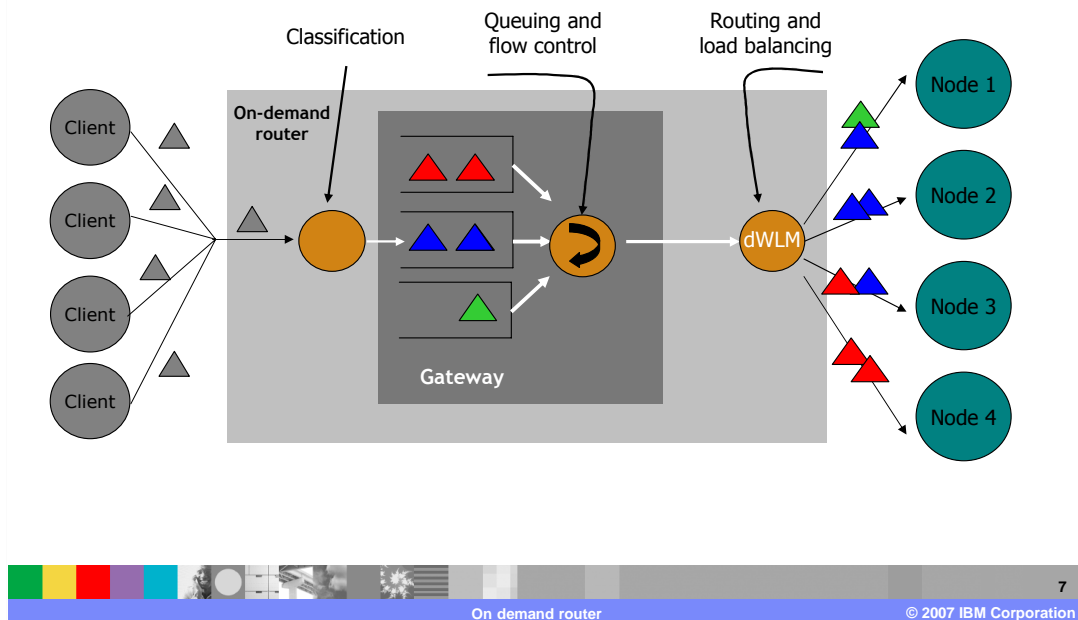
Section

Functional details



This section will discuss the functional details of the on-demand router.

Functional components overview



This diagram illustrates the flow of an HTTP request through the on-demand router. When a request enters the on-demand router, 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 on-demand router 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.

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 autonomic request flow manager is a component that runs inside of the on-demand router for managing HTTP requests. It is responsible for queuing work by service policy as requests come into the on-demand router, and 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.

Autonomic request flow manager 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 autonomic request flow manager 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.

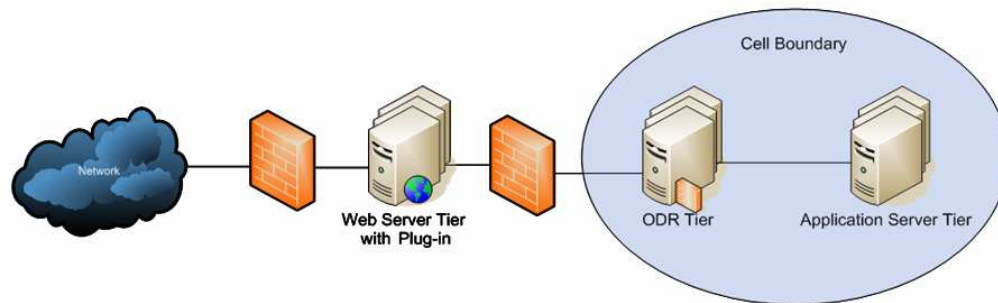
Section

On-demand router topologies



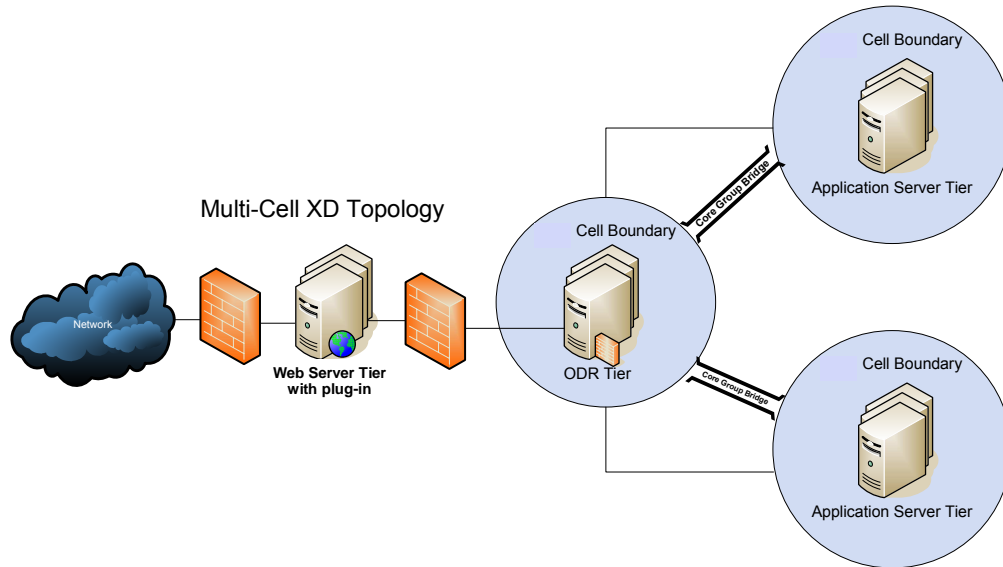
This section will present several possible topologies to use an on-demand router.

Topology including the on-demand router



While the on-demand router can be used in place of the HTTP server plug-in, many WebSphere Extended Deployment environments utilize both components. Often the HTTP server remains in the de-militarized 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 Extended Deployment to route requests to one or more on-demand routers. The on-demand routers then dynamically route the HTTP requests to dynamic cluster members in the application server tier.

Extended Deployment outboard OLTP topology



12

On demand router

© 2007 IBM Corporation

This slide shows the on-demand router isolated from other WebSphere Extended Deployment cells, communicating with them through core group bridges. You can use this multi-cluster routing capability to separate administrative domains. You can also use it for failover and load balancing between identically configured cells, which could be hosted by geographically separate data centers. With failover, you can specify a cluster to take over the workload when the primary cluster fails. With load balancing routing, you can balance the request loads between multiple clusters.

The on-demand router automatically gathers information about all the WebSphere Extended Deployment application servers and applications that are deployed in the cells and uses this information to dynamically configure itself at runtime so it can accurately route HTTP requests to those application servers. The on-demand router is able to route HTTP requests to any HTTP endpoint, not just WebSphere Extended Deployment nodes.

Section

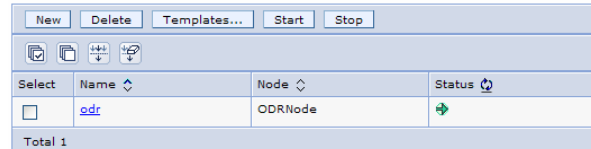
On-demand router administration



This section will introduce the creation and administration of an on-demand router. You can also view the demonstration “Administering an on-demand router”.

Configuring on-demand routers

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



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

Total 1



You can create a new on-demand router from the wsadmin.sh script “createodr.jacl” or you can use the administrative console. The administration page for on-demand routers can be found under “Servers” in the left pane of the administrative console. To create a new on-demand router, click on new and follow the wizard instructions. While properties of the on-demand router 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. On-demand routers can be placed on any managed node that has been augmented for WebSphere Extended Deployment Operations Optimization.

Summary

- on-demand router is critical to many WebSphere Extended Deployment features
- 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 Extended Deployment features, including interruption-free rollout of application upgrades and deployment manager high availability, rely on the capabilities of the on-demand router.

Feedback

Your feedback is valuable

You can help improve the quality of IBM Education Assistant content to better meet your needs by providing feedback.

- Did you find this module useful?
- Did it help you solve a problem or answer a question?
- Do you have suggestions for improvements?

Click to send e-mail feedback:

mailto:iea@us.ibm.com?subject= Feedback about XD61_On_Demand_Router.ppt



You can help improve the quality of IBM Education Assistant content by providing feedback.

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 WebSphere

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 or changes in the products or programs 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 (for example, 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 2007. 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.

