IBM® WebSphere® Extended Deployment V6.1

WebSphere Virtual Enterprise

Formerly Operations Optimization

Routing policies

@business on demand.

This presentation will cover the advanced selective routing capabilities in WebSphere Extended Deployment V6.1.

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

- Overview

- Routing options

- Runtime resources
    - Work class
    - Routing rules

- Configuration

- Summary

First this presentation will provide an overview of the advanced selective routing capabilities in WebSphere Extended Deployment with operational scenarios illustrating the available options. The presentation will then cover the runtime resources required to accomplish advanced routing, and will conclude with an overview of the steps required to configure a cell for advanced routing.

Overview

- Intelligent routing
  - ‣ On-demand router matches requests to routing policy
  - ‣ Then matches request to service policy

- Selective request redirection

- Selective request rejection

- Edition manager
  - ‣ Concurrent activation
  - ‣ Validation mode

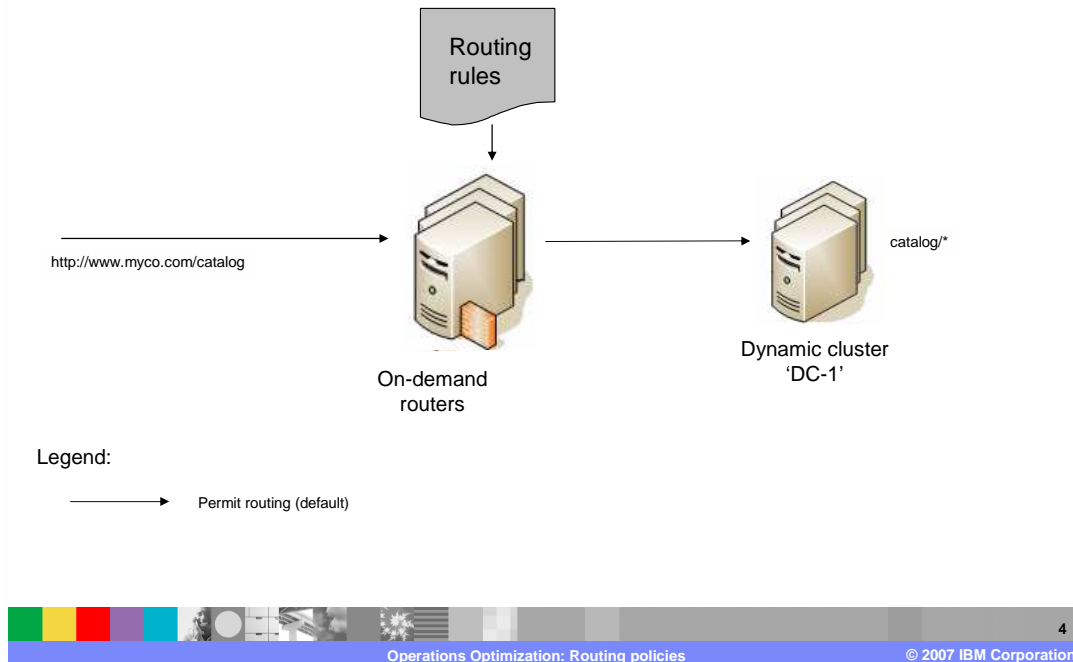Operations Optimization: Routing policies

3

© 2007 IBM Corporation

WebSphere Extended Deployment Operations Optimization is capable of advanced routing of HTTP and SOAP requests through the on demand router. When a request enters the on demand router the router will first determine which deployed application should handle the request. It will then search the target application's configuration to determine any special routing requirements for the request. If the matching routing policy is to reject or redirect the request, then the appropriate action is immediately performed by the on demand router and no additional classification is performed. If the resulting routing policy is to accept the request, then the on demand router will match the request to a service class which determines the response time requirements for the request.

WebSphere Extended Deployment routing policies are part of the application configuration. You can specify routing policies based on the URI of the incoming HTTP or SOAP request, and create specific routing rules based on information in the request.
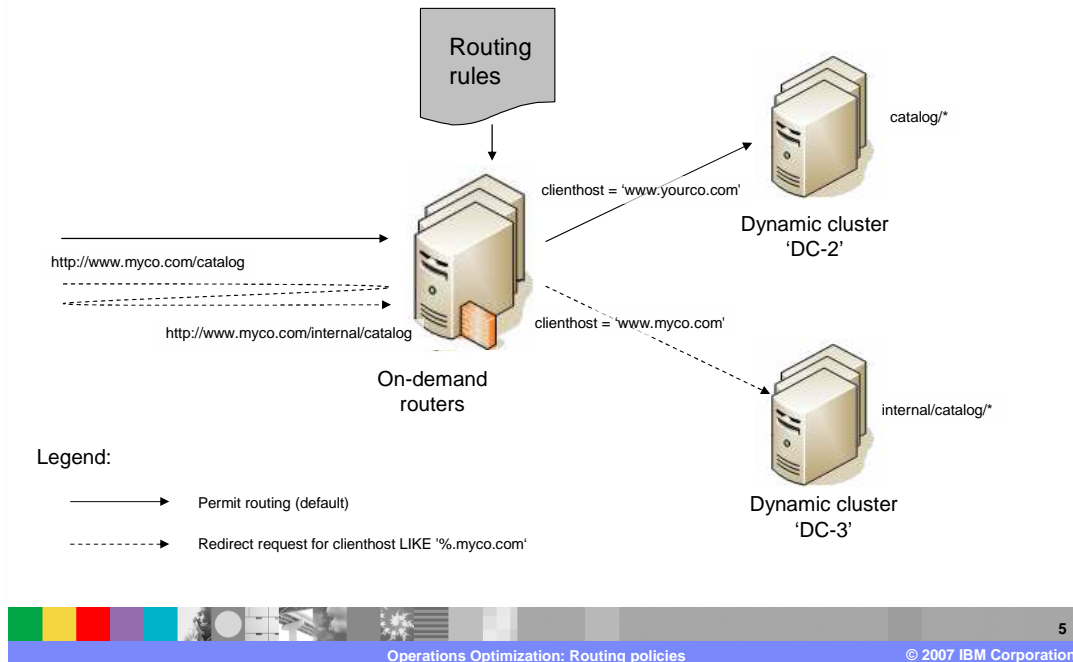
A specific request can be accepted with no special routing, redirected to a different URL, or rejected with a specified return code. The on demand router will also work with the application edition manager to selectively route requests between multiple active editions of an application.

Operational scenarios: Permit

Routing rules

http://www.myco.com/catalog

On-demand routers

Dynamic cluster 'DC-1'

catalog/*

Legend:

Permit routing (default)

The default routing policy is "permit routing to" and routes a request to the currently active edition of an application. This allows the request to continue as normal. The related policy "permit routing with affinity" is the same as the "permit routing to" policy, except that the on demand router also maintains client to server affinity for any future requests that come from the same client. In this case, the on demand router adds a SET-COOKIE header to the response before sending the response to the client.

Operational scenarios: Redirect

Routing rules

clienthost = 'www.yourco.com'

catalog/*

Dynamic cluster 'DC-2'

http://www.myco.com/catalog

http://www.myco.com/internal/catalog

clienthost = 'www.myco.com'

On-demand routers

internal/catalog/*

Dynamic cluster 'DC-3'

Legend:

→ Permit routing (default)

----→ Redirect request for clienthost LIKE '%.myco.com'

5

This diagram illustrates the ability to selectively redirect a request. With this routing policy, the on demand router redirects the request to the URL specified in the routing policy.

In the example shown, if the on demand router determines that a request is from an internal user, the request is redirected to a different location and will be served by a different application. Requests from external users are passed on to the requested application.

## Operational scenarios: Reject
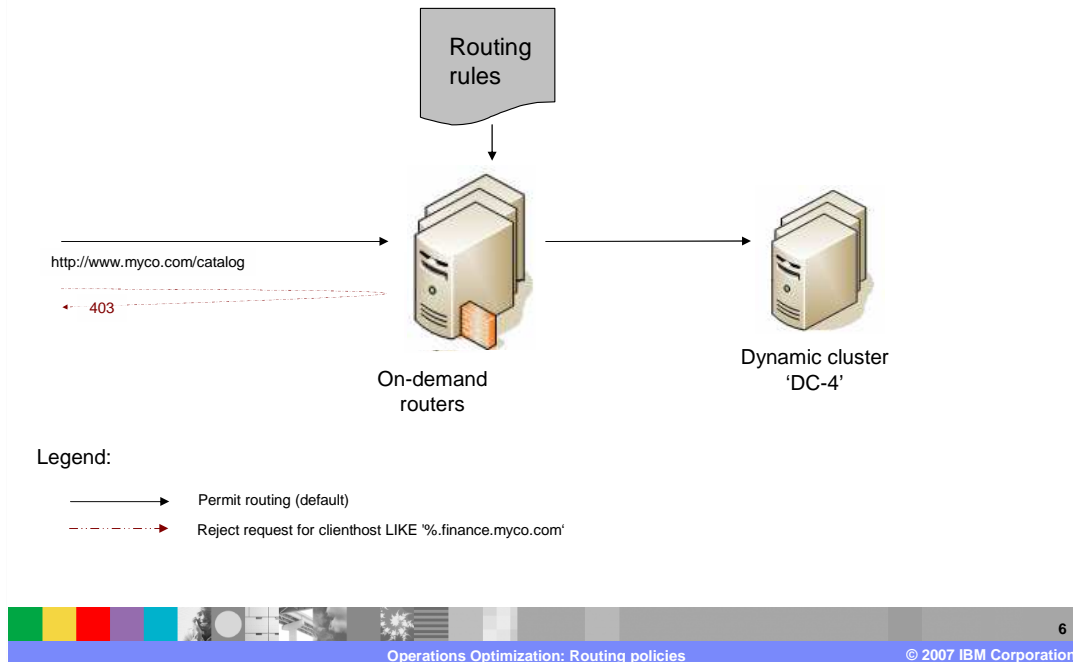
It is also possible to have the on demand router immediately reject requests that match a given rule, returning a specified error code to the client. In the example illustrated here, the on demand router will return an HTTP status code of "403" for any user from the "finance" domain. The status code "403" indicates the server understood the request, but is refusing to fulfill it. Authorization will not help and the request should not be repeated.

You can specify any number as the return code, but you should normally use a valid HTTP status code. To avoid confusion it would be best to avoid using common application server error codes such as 404 (not found) or 503 (service unavailable).

Note that the "Reject routing" policy is not intended to be used for security.
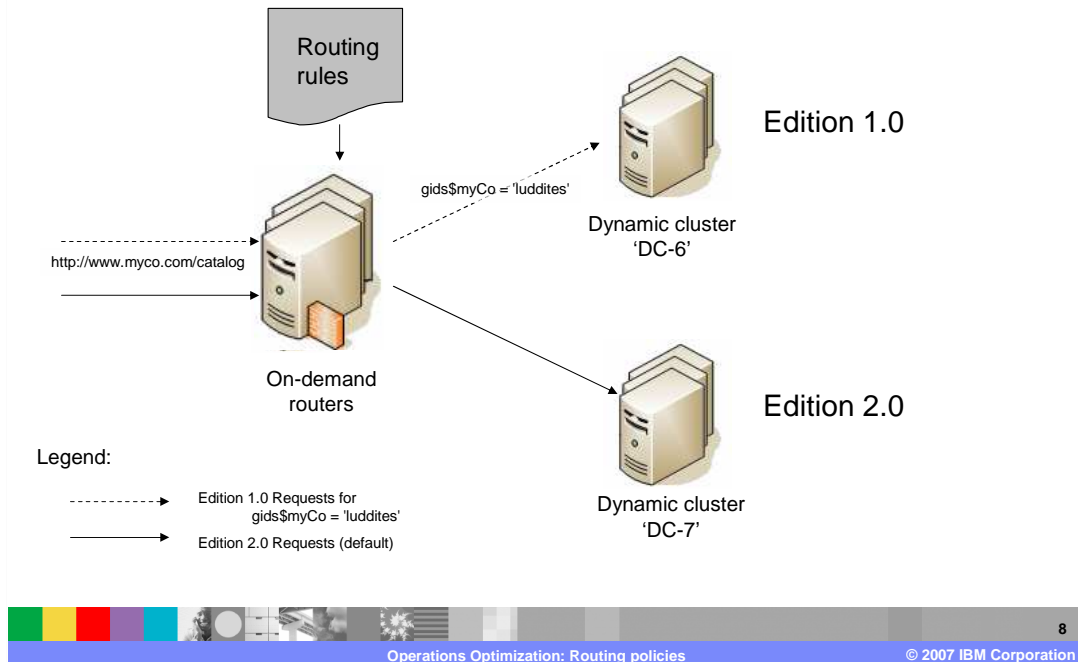
Operational scenarios: Validation mode
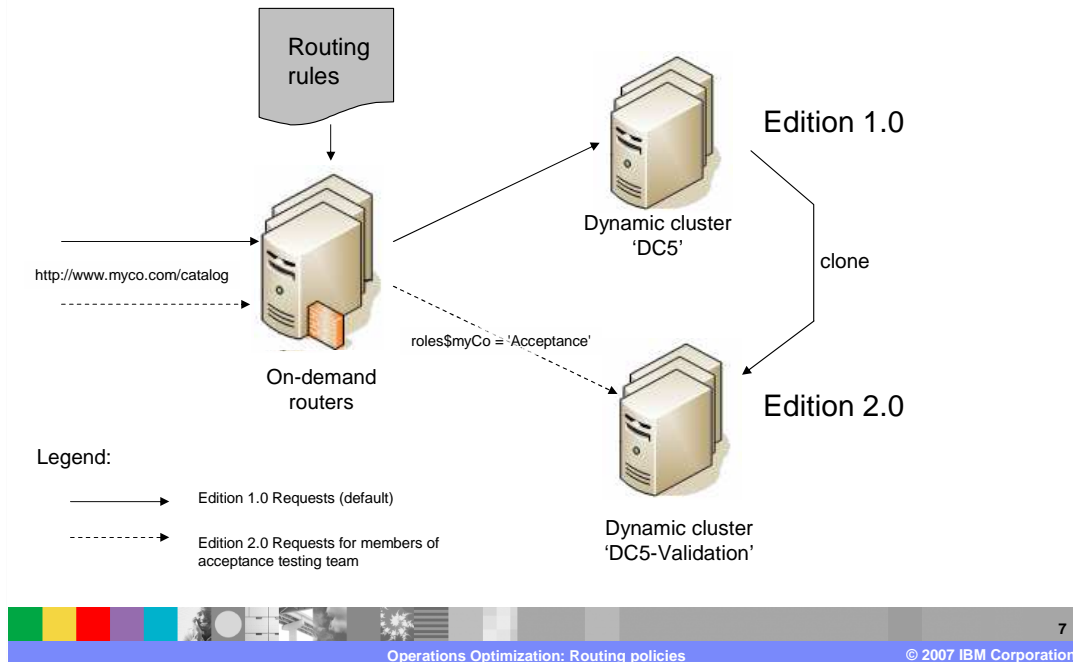
The application edition manager provides the ability to 'validate' a new edition of an application before rolling it out to the general audience. Validating an edition creates a clone in your production environment of the dynamic cluster where the application is deployed, and activates the new edition on that 'cloned' cluster. You then create routing policies to make the new edition available only to a limited subset of traffic. You can limit it to a certain group of users, a certain range of IP addresses, or many other limiting criteria. Validation mode gives you the ability to verify functionality of a new application edition in your production environment without exposing it to all users. You can later use the rollout feature to stop and remove the validation cluster and move the new edition into regular production mode.

This diagram illustrates the validation scenario. The on demand router controls the subset of requests that get routed to the validation cluster, or clone, based on routing policies that you have defined.

Operational scenarios: Concurrent activation

Routing rules

gids$myCo = 'luddites'

Edition 1.0

Dynamic cluster 'DC-6'

http://www.myco.com/catalog

On-demand routers

Edition 2.0

Dynamic cluster 'DC-7'

Legend:

Edition 1.0 Requests for gids$myCo = 'luddites'

Edition 2.0 Requests (default)

Finally, it is possible to create a similar situation without using validation mode to 'clone' a cluster. You can deploy a new edition of your application to a different deployment target. This is known as concurrent activation, where multiple versions of an application are active and serving requests at the same time. The on demand router uses routing policies defined in the application to determine how requests should be routed between the different editions.

In this diagram the default action is to "permit routing to" the newest edition of the application. However, a certain group of users do not want, or perhaps have not paid for, the enhancements in Edition 2.0. A routing rule ensures that users in this group continue to be routed to the older code.

Work classes

- Work classes define a certain type of work and associate it with a transaction class

- Requests are matched to work class by URI
  - Then to specific routing rules

- Each work class contains a default routing policy and an optional list of rules

- If no rules match, work will be routed to a default application edition

A routing policy is composed of one or more work classes. A work class is a set of rules that allow you to differentiate between incoming types of work. Each rule created within a work class has an associated routing policy. Work that matches a rule will be routed according to the associated policy. Each application contains default application routing policy work classes, which define how work that does not match any specific rules will be treated.

Runtime resources

- Work classes group work in (protocol, policy) pairs
  - ▶ HTTP or SOAP protocols
  - ▶ Routing policies or service policies
  - ▶ Further specialized by rules
    - Classification
    - Routing

- Rule builder panels provide an easy way to create rules
  - ▶ Rules can also be entered manually

Separate work classes are created for HTTP and SOAP. A work class is defined for HTTP requests by a set of HTTP URI patterns that the application can serve, or by the available Web service operations for SOAP requests. A work class defines a default routing policy, and can contain additional routing rules which specify how requests should be routed. The administrative console includes a rule builder to assist in building complex rules within a work class.

Routing rules

- Routing decision can be based on a number of variables:
    - ▶ Virtual host or Uniform Resource Identifier (URI)
    - ▶ HTTP headers, query parameters, and cookies
    - ▶ Web service and operation name
    - ▶ Client or server IP address, port, and host names
   - ▶ Time
    - ▶ User or group ID

Each work class contains an optional ordered list of rules that are evaluated for a particular request to determine the policy for that request. Each rule consists of a Boolean expression and a policy value. If the expression evaluates to "true" for a particular request, the policy associated with that rule is used. The Boolean expression for a rule is similar to the WHERE clause of an SQL expression.

Routing rule operands refer to various attributes that can be associated with a request. For example, a specific query parameter, cookie, or HTTP header. These operands can be specific to a protocol. For example, the SOAP service name is an operand that is valid only in a SOAP work class.

Creating work classes

- Work classes are defined under the *Routing Policies* tab of each application
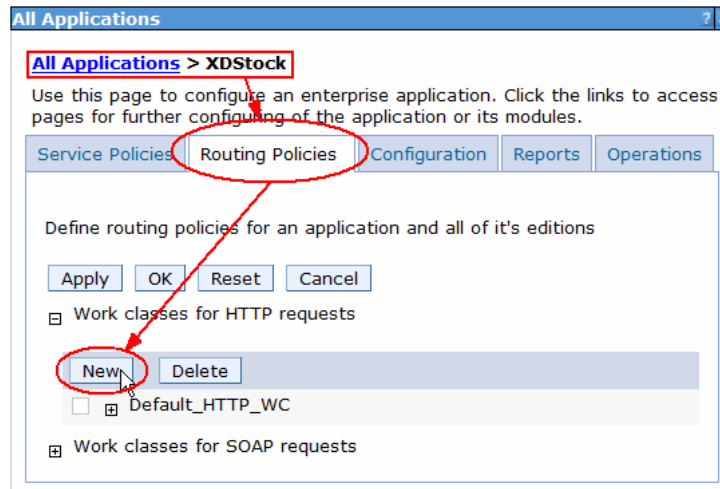
Routing policies can be accessed through the administrative console by expanding 'All applications' in the left-side menu and selecting the application.

From the application configuration panel, click on the 'Routing Policies' tab near the top of the frame.

There is also an MBean interface for creating or modifying routing policies from wsadmin scripts or custom Java Management Extensions (JMX) clients.

Creating work classes

Creating a work class

1. Name the work class
2. Select the war file and HTTP patterns
3. Review/Save
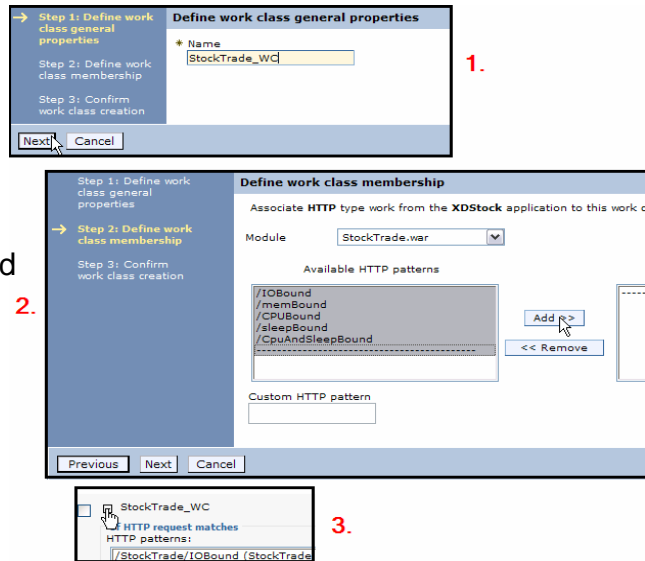
To define a work class you must first identify the Web modules and associated HTTP patterns or SOAP services to associate with the work class. A work class can include patterns from multiple Web modules in the enterprise application, and different patterns served by the application can be included in different work classes. Note that a request will only match a single routing policy work class.

Using the rule builder

Routing policies can be entered in the Administrative Console, as shown here. A routing rule is an IF-THEN style statement. The IF portion of the rule consists of a Boolean sub expression comparing a classification operand to a set of set values. You can create complex rules by combining sub expressions with the Boolean operations AND, OR, and NOT. The THEN portion of the rule specifies how to route the request if it matches.

If you do not know the syntax for the policy you would like to apply, the 'Build subexpression' link allows you to build policies using a menu system in the sub expression builder wizard. Once you have created an expression in the sub expression builder, you must use the browser's cut and past operations to move the expression to the routing rule. See the Rule builder demonstration to view how the rule builder can be used to create routing policies.

Sample routing rules

cookie$_MyCookieName_ IS NOT NULL

uid$ibm LIKE 'XDuser%' OR roles$ibm = 'administrator'

queryparm$timezone = 'EDT'

clientipv4 LIKE '10.12.18%.%' OR clientipv6 = '1:2:3:4:5:6:7:8'

HTTPMethod = 'POST' AND protocol = 'HTTPS'

time$11:40$13:30

clienthost LIKE '%.ibm.com' OR operation$calculate = 'finances'

This slide shows some example rule expressions. These rules can be used to classify a request, or to determine how a request should be routed.

Routing rule example

StockTradeRouting_WC

If HTTP request matches
HTTP patterns:

/StockTrade/IOBound (StockTrade.war)
/StockTrade/memBound (StockTrade.war)
/StockTrade/CPUBound (StockTrade.war)
/StockTrade/sleepBound (StockTrade.war)

[Edit HTTP Patterns]

Then apply the following routing rules

[Add Rule] [Delete Rule] [Move Up] [Move Down]

Select	Order	Routing rule
☐	1	**If** uid$ibm = 'joe' and protocol = 'HTTPS' **Then** Permit routing to XDStock-editionVersion 2.0 [Edit rule]
☐	2	**If** clienthost LIKE '%.ibm.com' OR queryparm$timezone = 'EDT' **Then** Reject routing with return code 403 [Edit rule]

If no routing rules apply
Select action
Permit routing to

Select edition name
XDStock

Operations Optimization: Routing policies
© 2007 IBM Corporation

This slide shows a work class with multiple rules. Note the IF-THEN structure. When a new request enters the on demand router it will first match the incoming request to the work class based on the request's HTTP pattern. The on demand router will then examine the routing rules until it finds one that matches the current request. If none of the routing rules matches, the default policy is applied.

Summary

- The on-demand router is an intelligent proxy server
  - Performs request classification, queuing, and dynamic workload management

- When a request enters the ODR, it is classified as follows:
  - A work class associates the request with a routing rule
  - If the routing policy is "permit", the on demand router associates the request with service policy
  - The service policy defines how the request should be treated

Operations Optimization: Routing policies © 2007 IBM Corporation

In summary, the on demand router is capable of advanced routing of HTTP and SOAP requests. When a request enters the on demand router it matches the incoming HTTP request to a routing policy to determine how to route the request. If the routing policy is to "permit" the request it will then match the request to a service class which determines the response time goal for the request.

Summary

- Routing policies
  - ▶ Rules used by the on-demand router to determine which edition should handle the request

- Routing policy can be
  - ▶ Permit
  - ▶ Redirect
  - ▶ Selective request rejection

- Edition manager
  - ▶ Concurrent activation
  - ▶ Validation mode

Routing policies are part the application configuration. A specific request can be accepted with no special routing, redirected to a different URL, or rejected with a specified return code. The on demand router will also work with the application edition manager to selectively route requests between multiple active editions of an application.

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_RoutingPolicies.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   ibm.com   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.