This presentation will give an overview of the dynamic operations features in WebSphere Extended Deployment for z/OS 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

- Dynamic operations overview
  - Example scenario
  - Benefits
- Key concepts

This presentation will begin by illustrating the main ideas behind dynamic operations using a simple example scenario that highlights the benefits of a dynamic WebSphere Extended Deployment environment. It will then introduce some of the key concepts involved in creating a dynamic operations-based WebSphere Extended Deployment environment.

Section

Overview

This section will give an overview of dynamic operations.

Dynamic operations overview

- Virtualized, policy-based, dynamic workload management
- On-demand router
  - Enhanced version of the proxy server
  - Controls request prioritization, flow, and routing in an Extended Deployment environment
- Dynamic application placement
  - Enables starting and stopping server instances based on application load and user-defined goals

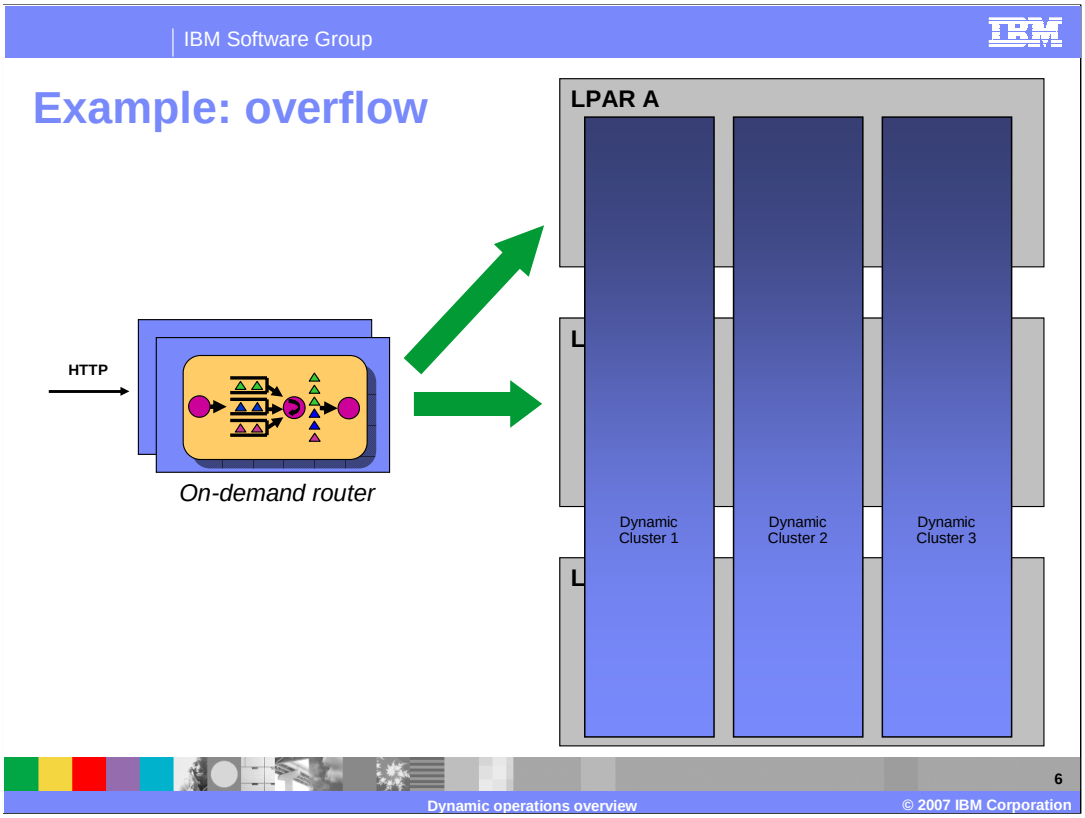
The dynamic operations features of WebSphere Extended Deployment gives you the capability to build a dynamic, virtualized, goal-oriented environment for workload management. The two major features that enable these capabilities are the On-Demand Router and dynamic application placement. The On-Demand Router is an intelligent HTTP proxy server that manages request prioritization, flow control and dynamic routing of requests to your application servers.

Dynamic application placement enables starting and stopping additional server instances to accommodate changes in load, balancing processing power among your applications to best meet your defined performance goals.

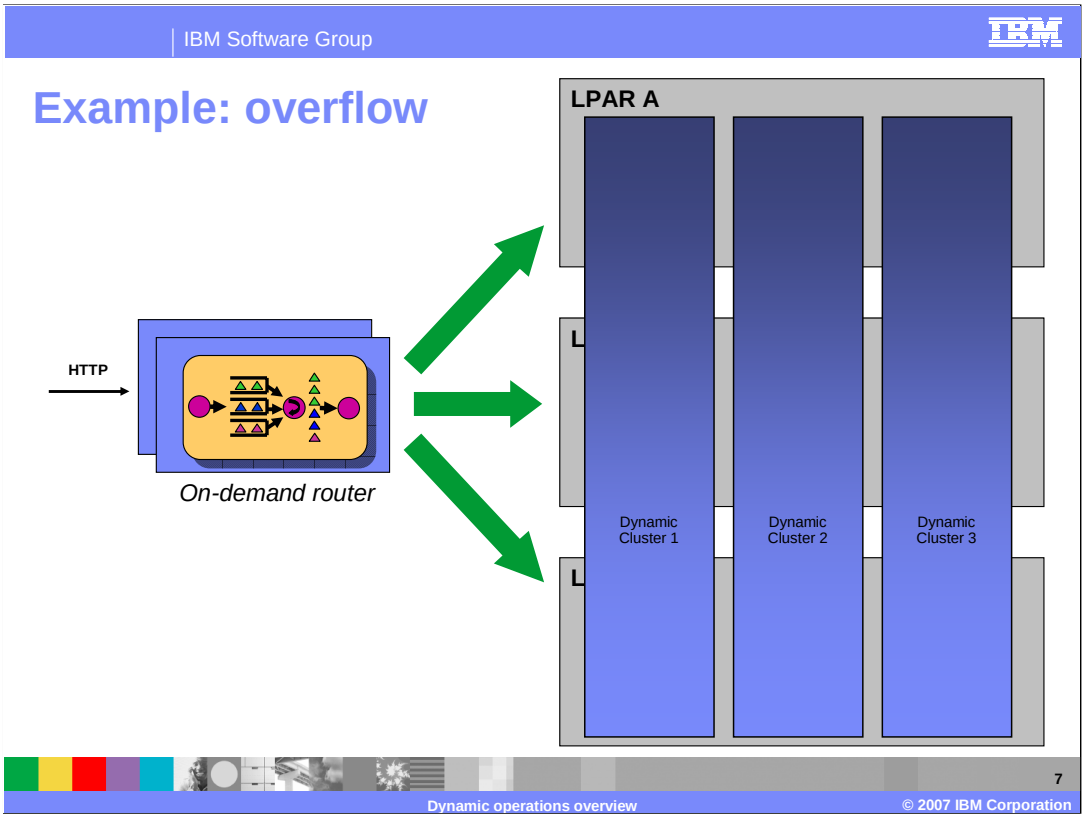
Example: overflow LPAR

- Three LPARs
  - Two primary LPARs
    - Satisfy majority of requests
    - Provide redundancy
  - One overflow LPAR
    - Limited memory available
    - Can only have a few address spaces running

In this example, your organization has many departments and they each have several applications that they run on their own department's WebSphere server in a central complex. For administrative purposes these departments do not share servers but rather maintain their own through a central IT group. As the overall administrator, you have three LPARs configured. "LPAR A", "LPAR B" and "LPAR C". A and B are the primary LPARs for satisfying requests for all departments. On occasion, one or more servers exceed the capacity available on LPARs A and B. To handle this overflow you have configured "LPAR C" as a spare. Therefore "LPAR C" can only have a few servers (controller/servant region pairs) running at any one time.

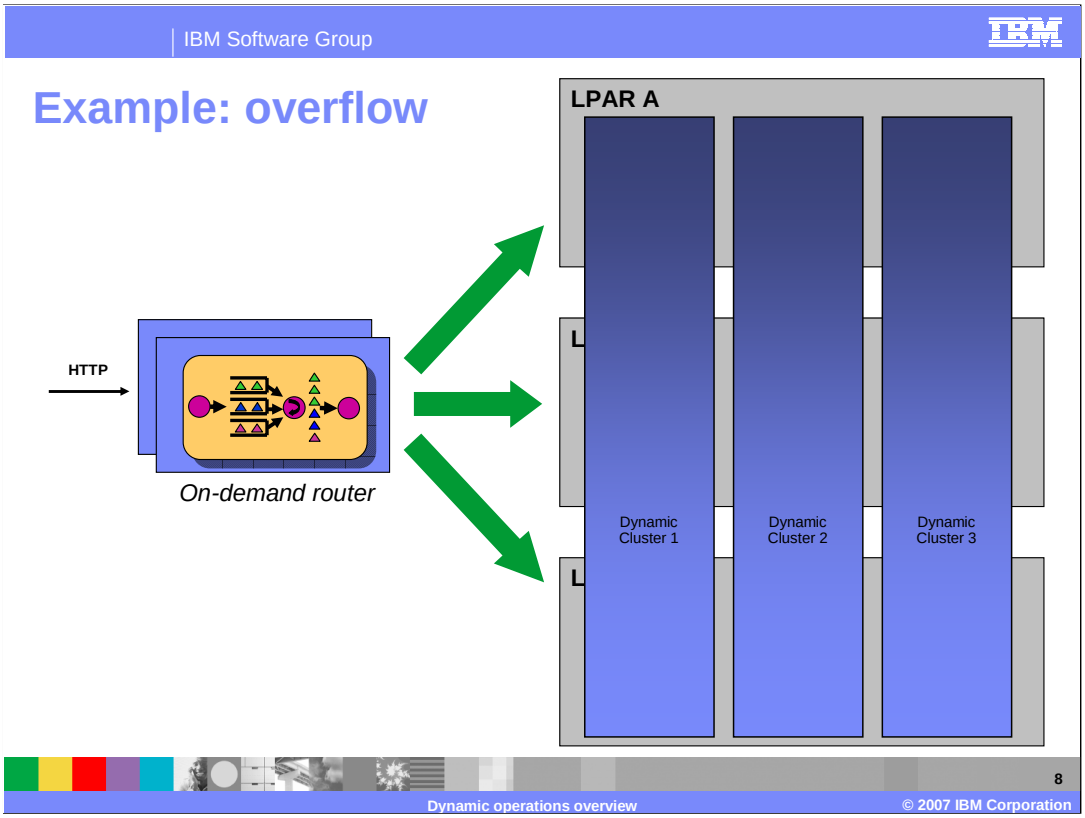


This picture represents nominal configuration. When the on demand router starts up, it detects that LPAR C does not have as much memory or processor available as the other two LPARs. Here LPAR A and LPAR B have servers running for all three departments and these servers are handling the entire offered workload. As the workload for Department Two begins to increase and the running servers can no longer satisfy the pre-configured goals, the on demand router determines that LPAR C has spare bandwidth available.



Here the application placement controller has started a server for Department Two on LPAR C. The on demand router is routing work to that new server, in addition to the two existing servers, in order to better balance how Department Two's workload is processed.

Later, the offered load for the Department Two server decreases and the offered load for Department One increases. The application placement controller will detect the change and reconfigure the dynamic clusters.



Here, the third server for Department Two has been stopped and a new server for Department One has been started on LPAR C. This dynamically accommodates the decrease in the offered load for Department Two and the increase in the offered load for Department One.

Now assume that the offered load for Department Three increases and the two Department Three servers can not satisfy the configured goals. Now the application placement controller examines the LPARs and determines that LPAR C does not have enough memory resource to start another server, so it does not start another server. The dynamic behavior of autonomic managers, like the on demand router and the application placement controller, enable WebSphere Extended Deployment to effectively maximize the resources available.

Benefits

- Enables more efficient hardware utilization
  - Dynamic allocation of resources to handle variations in traffic
    - Takes advantage of differing peak times
  - Server consolidation reduces total cost of ownership
- Helps ensure a consistent level of service for critical applications
  - Decisions are based on user-defined policies
  - In times of contention, more important requests will perform better than less important requests

By taking advantage of differing peak times in application load, hardware can be utilized much more efficiently in a dynamic operations environment, resulting in lower overall hardware costs. A dynamic operations environment also helps ensure a consistent quality of service for your applications with less administrative monitoring. WebSphere Extended Deployment can allocate hardware resources to help ensure that applications meet their defined goals and route work accordingly. This allows the most important requests to perform better than less important requests when there is a contention for resources.

Section

Key concepts

This section will cover the key concepts and components of a dynamic operations environment.

Dynamic clusters

- A dynamic group of servers to which applications can be deployed
- Similar to a static cluster, but the number of active servers can be resized dynamically at run-time
- Each dynamic cluster has a template that defines settings for member servers
  - Modifying this template affects all servers in the dynamic cluster
  - Static clusters require changing each individual server to achieve the same effect

A Dynamic Cluster is similar to the familiar concept of a ‘cluster’ from WebSphere Application Server, but can be resized dynamically at run-time. As demand for applications running on a Dynamic Cluster increases or decreases, instances of that Dynamic Cluster can be started or stopped on nodes within the cluster to accommodate the changes in load. Each node in the set of defined cluster members has a configured instance of the Dynamic Cluster that is ready to be started dynamically when needed. These server instances are configured based on a server template that defines the configuration for all of the cluster members. This template is used as a single point of configuration for all members of the Dynamic Cluster. No template is used if a Dynamic Cluster is created from existing servers.

Dynamic cluster membership methods

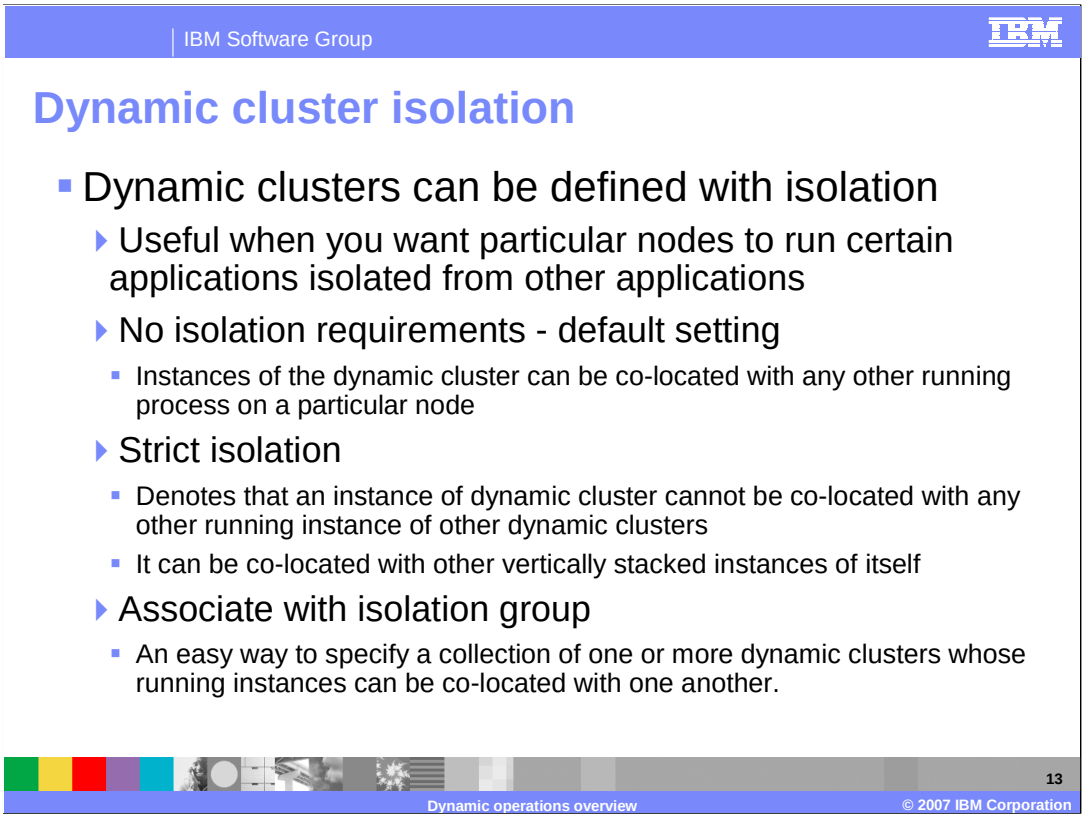
- A membership method determines which servers can join the dynamic cluster
- You can choose to manually define dynamic cluster members from existing servers
- You can select to automatically define dynamic cluster members with rules
  - Create a membership expression in the administrative console (sub-expression builder tools are available)
  - Each existing node is evaluated against the expression. If it is a match, the node will be considered a candidate to host a cluster member.

When you create a dynamic cluster, you must specify which nodes can host members of the dynamic cluster. WebSphere Extended Deployment version 6.1 allows two methods to determine which nodes can join the dynamic cluster. You can identify pre-existing servers as members of your dynamic cluster or you can specify a “membership policy”, which is a rule that specifies the set of nodes that can host dynamic cluster members.

Prior to WebSphere Extended Deployment version 6.1, members of a dynamic cluster were scoped explicitly by node group. In version 6.1 you have more options available to delimit your dynamic cluster. In addition to node group, you can define membership rules based on node name, node host name, and node property values. You can create complex rules using Boolean operators and, or, and not.

The membership policy is evaluated against the nodes in your cell when the dynamic cluster is created. WebSphere Extended Deployment will create servers for the dynamic cluster using nodes that match the membership policy that you define. If new nodes are added to your environment, they will automatically be added to the dynamic cluster if they match the defined membership policy. Similarly, if you change a member ship policy, it is reevaluated and new server instances are created or removed based on the new policy definition. If you change a node’s properties such that it should be added to or removed from an existing dynamic cluster the corresponding server instances are added or removed on that node when you save your changes.

The default membership policy includes all nodes in the cell that have been augmented for WebSphere Extended Deployment version 6.1 operations optimization.



At times you may have applications that should not run on the same node as other applications. Dynamic cluster isolation allows you to specify to the application placement controller which dynamic cluster instances can coexist on the same node. WebSphere Extended Deployment version 6.1 provides three options for configuring the dynamic cluster isolation requirements: no isolation requirements, strict isolation, and associate with isolation group.

“No isolation requirements” is the default isolation policy. It specifies that instances of the dynamic cluster can be co-located with any other running process when placed on a node. This is the same behavior as previous versions of WebSphere Extended Deployment.

“Strict isolation” specifies that when an instance of the dynamic cluster is placed on a node, it must not be co-located with running instances of any other dynamic clusters. In other words, it can only be co-located with other vertically stacked instances of itself.

“Associate with isolation group” is a convenient way to specify a collection of one or more dynamic clusters whose running instances can be co-located with each other. The application placement controller can place running instances of a dynamic cluster with an isolation policy of “associate with isolation group” on the same node as other running dynamic cluster instances, so long as their dynamic clusters are members of the same isolation group.

Note that when a dynamic cluster is in manual mode it is up to the administrator of that

On-demand router

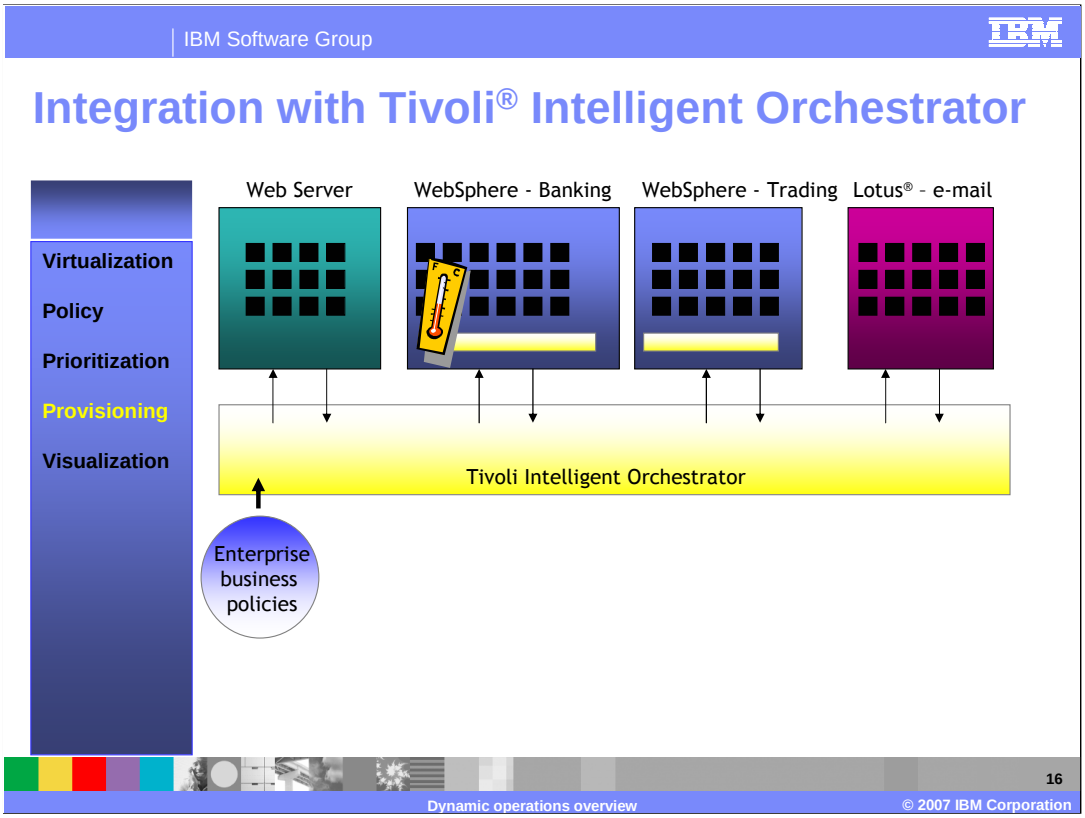
- Intelligent HTTP proxy server
- Enhanced version of the proxy server from Network Deployment 6.1
- Can replace or complement the HTTP server plug-in
- Prioritizes requests and controls traffic flow according to operational policy
  - Ensures consistent quality of service
  - Enables more elegant degradation of performance when all resources are consumed
- 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 request distribution 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 dynamic cluster 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. These and other advanced features distinguish the on demand router from the HTTP server plug-in, 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 together with the HTTP server plug-in, depending on your needs.

Application placement controller

- Decides how many instances of each dynamic cluster should run, and where they should run
- Determines the available capacity (memory, processor) of each node
  - Aware of the capacity in use by other processes and subtracts from available capacity
  - Isolation policy is taken into account
- One application placement controller per cell
  - This change accommodates overlapping node groups

The Application Placement Controller is the component that decides how many instances of each Dynamic Cluster should be running to most effectively handle the current amount of traffic. It determines the available processor and memory capacity of each node, including resources that are in use by other processes. It uses this information to determine the optimal placement of each application to best meet your defined performance goals. The Dynamic Cluster's isolation policy requirements are also considered when the Application Placement Controller determines where an instance of a Dynamic Cluster may be placed. Each cell has one Application Placement Controller, which is a highly available singleton service that runs inside one of the Node Agents within the cell.



WebSphere Extended Deployment can be integrated with Tivoli Intelligent Orchestrator for server provisioning in a larger, heterogeneous environment. Tivoli Intelligent Orchestrator is a product that provides the capability to dynamically allocate hardware resources across products within an enterprise. Business policies dictate the allocation of enterprise-wide server resources, which can be reallocated based on need.

For example, if a WebSphere Extended Deployment cell has exhausted all of the resources available to it, servers that were previously part of an underutilized Lotus® e-mail environment can be reprovisioned as WebSphere Extended Deployment servers, and be added into the WebSphere Extended Deployment cell to begin hosting Dynamic Cluster instances. WebSphere Extended Deployment provides the required hooks for operating within a Tivoli Intelligent Orchestrator environment.

Summary

- WebSphere Extended Deployment creates a dynamic, virtualized, goal-based environment for application hosting
  - The environment can adapt to varying traffic levels and allocate server resources as necessary
- Dynamic clusters are similar to clusters, but the number of active servers can be resized dynamically at run-time
  - The application placement controller decides when this should take place

In summary, WebSphere Extended Deployment enables you to create a dynamic, virtualized, goal-based environment for hosting your enterprise applications. This environment can adapt to varying traffic levels and allocate server resources as necessary to help meet the performance goals of your applications. Applications are installed to Dynamic Clusters, which can be dynamically resized within the virtual pool of resources. The Application Placement Controller is the component that is responsible for making placement decisions based on current load levels and user-defined performance goals.

Summary

- The on-demand router is a new component provided by WebSphere extended deployment
  - On-demand router is an intelligent HTTP proxy server
  - Performs request classification, flow control, and dynamic workload management

Finally, the on demand router is the point of entry for HTTP requests into a dynamic operations environment. It performs request classification based on user-defined rules, ensures that more important requests flow through to the back end more quickly than less important requests, and dynamically routes requests to application servers and dynamic cluster members running in the application server tier.

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 XD61z_Dynamic_Operations_Overview.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 Lotus Tivoli WebSphere z/OS

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.

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.