



IBM Software Group

# IBM® WebSphere® Extended Deployment V6.1 for z/OS®

## *WebSphere Virtual Enterprise*

*Formerly Operations Optimization*

### **Overview**



@business on demand.

© 2007 IBM Corporation  
Updated June 19, 2008

This presentation will give an overview of WebSphere Extended Deployment Operations Optimization 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

- Introduction to WebSphere Extended Deployment
- Dynamic operations
- Advanced administration features



This presentation will briefly introduce WebSphere Extended Deployment Operations Optimization, and describe what is meant by dynamic operations and how WebSphere Extended Deployment leverages this capability. The presentation will then provide a high level overview of the numerous advanced administration enhancements that are part of the operations optimization package.

## Section

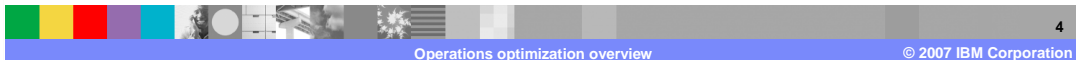
# *Introduction*



This section will introduce the WebSphere Extended Deployment operations optimization package.

## Operations optimization

- 'Add-on' product for WebSphere Application Server Network Deployment
  - ▶ Requires WebSphere Application Server V6.1
  - ▶ Fundamentally alters WebSphere Network Deployment operations
- Extends your existing middleware systems
  - ▶ Quality of service
  - ▶ Performance
  - ▶ Resiliency
  - ▶ Manageability
- Works with a variety of middleware systems



WebSphere Extended Deployment operations optimization is a set of extensions for your existing middleware systems focused around quality of service, performance, resiliency, and manageability. It includes dynamic operations features which enable a virtualized, dynamic environment for goal-oriented workload management, to take better advantage of hardware resources in a distributed environment. This package provides several advanced features for enhanced administration. These features include the ability to manage and seamlessly roll out multiple editions of an application, advanced visualization tools to help you better understand what is happening in your environment, and health monitoring agents that can automatically take action when certain software conditions are detected.

Each of these topics will be discussed briefly in this presentation, and covered in more detail in other presentations.

## Section

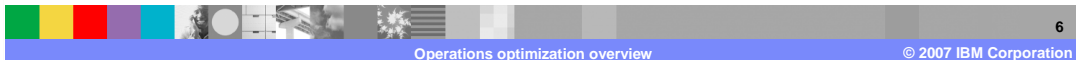
# *Dynamic operations*



This section will describe the dynamic operations features of WebSphere Extended Deployment.

## Dynamic operations overview

- Dynamic workload management
  - ▶ Virtualized
  - ▶ Goal-oriented
  - ▶ Policy-based
- More consistent quality of service for critical applications
- More efficient utilization of hardware resources



WebSphere Extended Deployment provides a virtualized environment that allows hardware resources to be shared and allocated dynamically based on performance goals that you have defined for your enterprise applications. This capability can help you provide a more consistent quality of service for your critical applications in times of excessive load, and more efficiently utilize the hardware resources that you already have. The following example will highlight the benefits of a dynamic, virtualized WebSphere Extended Deployment environment.

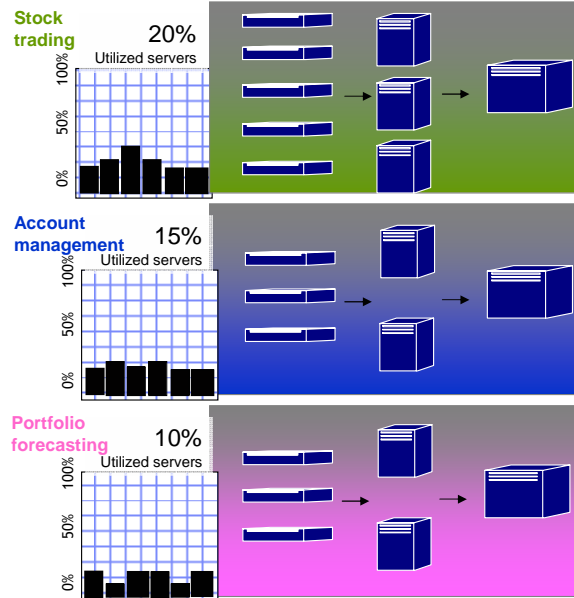
## Example: Typical distributed environment

### Environment

- ▶ Multiple business critical applications
- ▶ Hundreds of application servers

### Challenges

- ▶ Underutilized servers
- ▶ Inability to share resources across clusters
- ▶ Inconsistent quality of service
- ▶ Human-intensive monitoring and management environment



An example of a typical environment without WebSphere Extended Deployment is shown here, with three applications: stock trading, account management, and portfolio forecasting. A real environment would likely be an order of magnitude larger than this example, in terms of both applications and hardware. Most of the servers are underutilized because they have all been provisioned for peak load. You get inconsistent qualities of service in this scenario, because if one application gets more load than it can handle, it cannot take advantage of other resources that may be sitting idle.

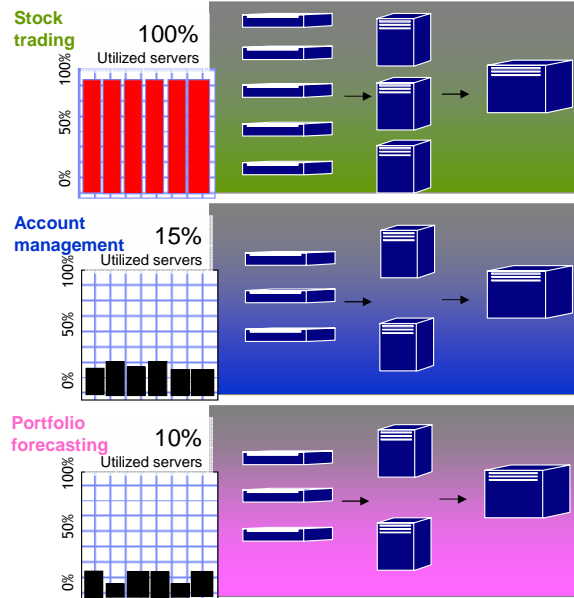
## Example: Typical distributed environment

### Environment

- ▶ Multiple business critical applications
- ▶ Hundreds of application servers

### Challenges

- ▶ Underutilized servers
- ▶ Inability to share resources across clusters
- ▶ Inconsistent quality of service
- ▶ Human-intensive monitoring and management environment



As you can see in this example, each of the three applications has its own cluster, running at low utilization. If for some reason a large amount of traffic hits the stock trading application, its cluster could become 100% utilized. At this point the application's performance will begin to degrade, despite the fact that you have other resources sitting underutilized.



## Example: Dynamic operations environment

- **Virtualized**
  - ▶ Pooled resources
  - ▶ Virtualized applications
- **Goals based**
  - ▶ Operational policies are attached to applications to reflect operational goals and importance of application
  - ▶ Autonomic managers monitor environment for maximum utilization using business goals
- **Results**
  - ▶ Reduce total cost of ownership (doing more with same/less)
  - ▶ Increase stability and repeatability of environment

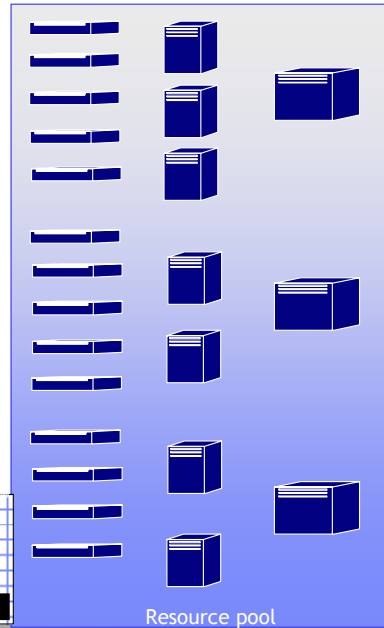
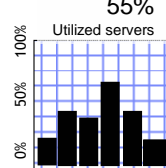
Stock trading

Customer support

Account management

Risk management

Portfolio forecasting



Resource pool

9

Operations optimization overview

© 2007 IBM Corporation

What WebSphere Extended Deployment does, instead, is provide a virtualized environment. The key concept is dynamic clusters (which are similar to regular 'clusters' but can dynamically grow and shrink), to which your applications will be deployed. You then define policies that dictate performance goals that will govern the management of workload for that application. WebSphere Extended Deployment will manage work against that goal, and as you see in this example, resources are dynamically allocated within the resource pool as work comes into the system to balance resources and maintain acceptable performance. You may even find that you can afford to deploy more applications onto the same amount of hardware, since it is now utilized in a more efficient manner.

## Key features

- On-demand router
  - ▶ New process built as a set of filters to the proxy server
  - ▶ Provides routing of HTTP traffic into a collection of cells
  - ▶ Provides three key on-demand features:
    - Flow control and queuing
    - Prioritization
    - Dynamic workload management
- Dynamic application placement
  - ▶ Autonomic placement of instances of application servers on nodes
  - ▶ Starts or stops preconfigured server instances to meet defined performance goals



WebSphere Extended Deployment includes a new process, the on-demand router that is an enhanced version of the proxy server included in WebSphere Application Server version 6.1. The on-demand router is an intelligent HTTP proxy that acts as the entry point for traffic coming into a WebSphere Extended Deployment cell, performing request prioritization, flow control, and dynamic workload management.

Another key feature is known as dynamic application placement, which uses dynamically sized server clusters to increase or decrease the number of servers on which a particular application is running to help meet your performance goals. Each node within a dynamic cluster has an instance of an application server running that cluster's applications that can be started and stopped dynamically as traffic for that application increases or decreases.

## On-demand router

- Important participant in many Extended Deployment functions
  - ▶ Dynamic operations
  - ▶ Edition management
  - ▶ Scale out management
  - ▶ Object grid (partitioning)
  - ▶ Not directly part of business grid
- World class edge of sysplex distribution
  - ▶ Compliments sysplex distributor
  - ▶ Complements z/OS workload management

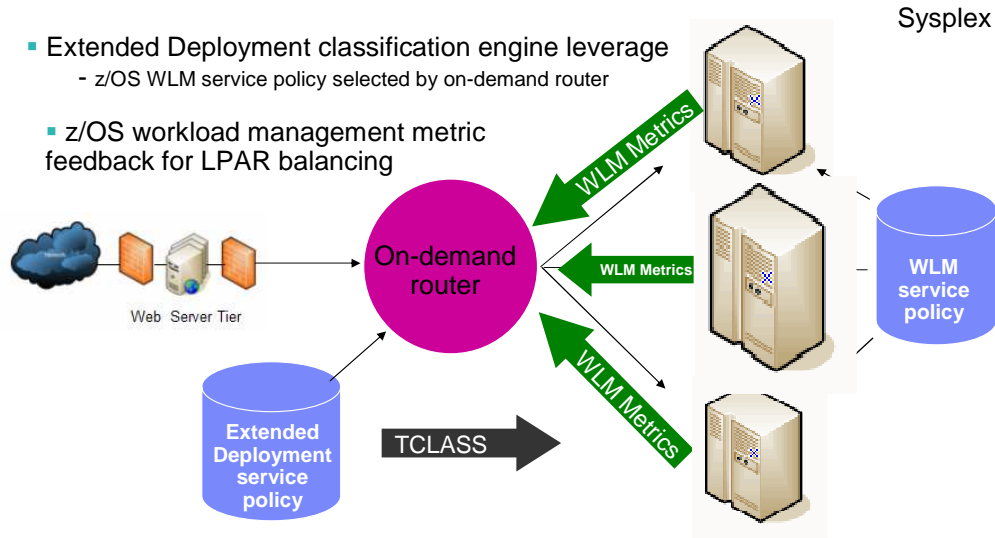


The on-demand router is one of two primary WebSphere Extended Deployment workload schedulers. The other scheduler is the Long Running Scheduler discussed earlier. As such the on-demand router is a critical piece of WebSphere Extended Deployment and is required for most of the new functionality, except for WebSphere business grid applications. For example, the on-demand router maintains both an application view of offered load and a two-way coordination with z/OS WLM. The on-demand router is key to application roll out for edition management by routing request to the correct cluster or temporarily caching request as necessary.

The on-demand router also provides world class edge of Sysplex distribution. When placed as end points for Sysplex distributor, the on-demand router uses workload management information to enhance the Sysplex Distributor. The on-demand router is integrated with z/OS workload management and the on-demand router rules builder actually enhances service policy selection. The on-demand router can be placed on a non z/OS based platform and still maintain integration with z/OS WLM.

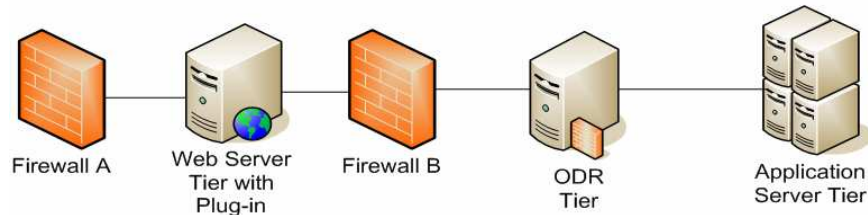
## z/OS workload management integration

- Extended Deployment classification engine leverage
  - z/OS WLM service policy selected by on-demand router
- z/OS workload management metric feedback for LPAR balancing



z/OS WLM integration is a unique value-add that Extended Deployment brings to the z/OS platform. It provides two-way integration; first the service unit metric data is retrieved from the platform to enable the on-demand router to make informed decisions on workload placement. Then the classification engine in Extended Deployment is leveraged to assign and propagate the transaction class (TCLASS). This class will be used downstream by a WebSphere cluster to register the work request with z/OS and to select the z/OS service policy that governs how it runs.

## Common topology



- Typical topology has multiple on-demand routers for scalability and availability
- on-demand router
  - ▶ Generates the plugin-cfg.xml file for Web server plug-ins
  - ▶ Dynamically learns configuration of backend cells
  - ▶ Can be connected to multiple cells simultaneously
  - ▶ Supports routing to any HTTP endpoint
  - ▶ Can be placed in front of the Web server tier, for example



The most common topology places the on-demand router (or more commonly multiple on-demand routers) inside the private network between the Web server and the application servers. This leaves the Web server to continue serving static content, but the plug-in running on the Web server is now configured to route traffic to the on-demand router, rather than to the application servers. The on-demand router is constantly updated with the location of dynamic cluster instances within the cell (or cells), and routes traffic accordingly.

Other topology options are also possible, including placing the on-demand router tier in the de-militarized zone, where it could be used in front of, or instead of the Web server tier.

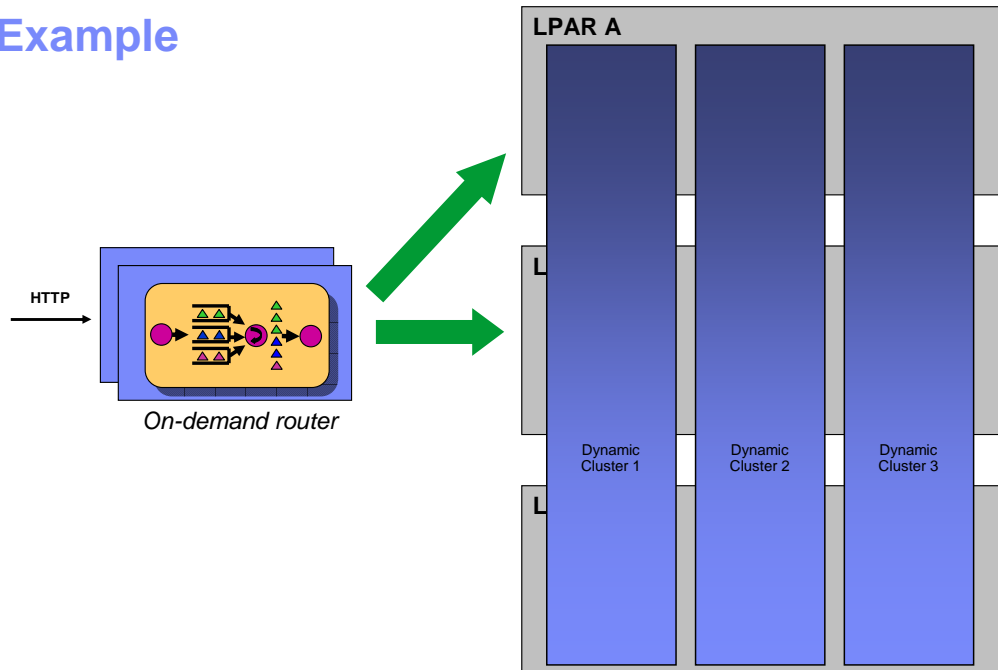
## Scenario: Overflow LPAR

- Three LPAR's
  - ▶ Two Primary LPAR's
    - Satisfy majority of requests
    - Provide redundancy
  - ▶ One Overflow LPAR
    - Limited memory available
    - Can only have a few address spaces running



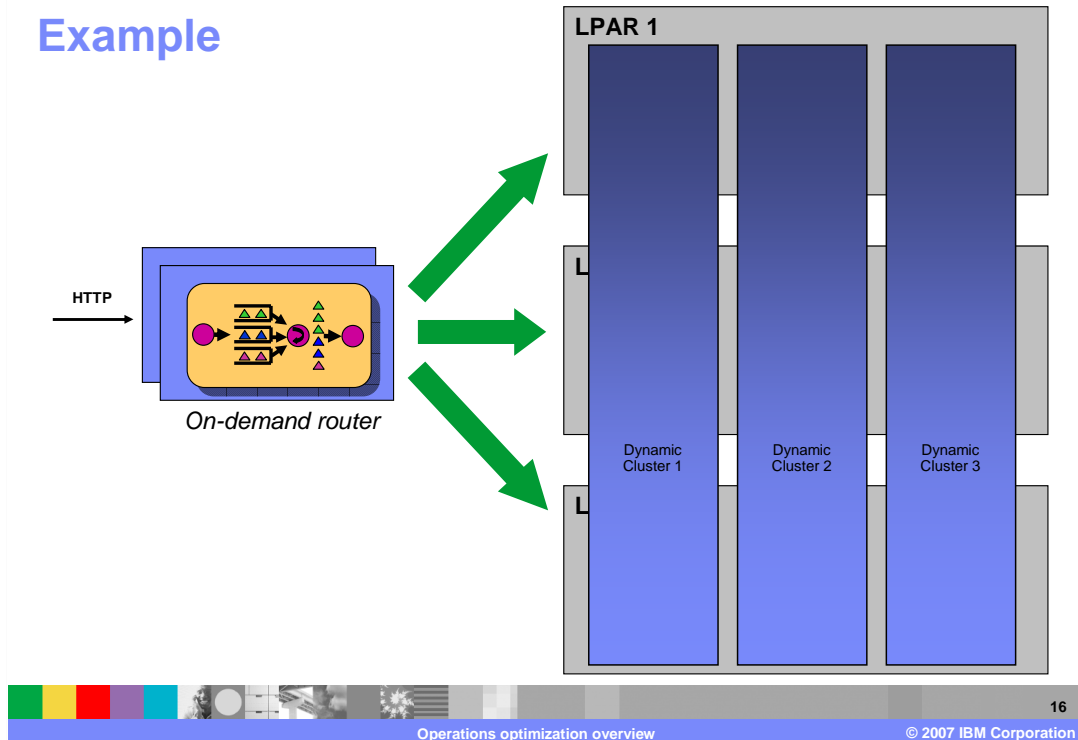
In this example, your organization has many departments and they each have several applications that run on department level WebSphere servers in a central complex. For administrative purposes these departments don't share servers but rather maintain their own through a central IT group. As the overall administrator, you have three LPARs configured. "LPAR A", "LPAR B" and "LPAR C". A and B are the primary LPAR's for satisfying requests for all departments. On occasion, one or more servers exceed the capacity available on LPAR's A and B. To handle this overflow you have configured "LPAR C" as a spare. Because "LPAR C" resides on a machine that is primarily used for other purposes, it has a limited amount of memory and can only have a few servers running at any one time.

## Example



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 allocated to it as do the other two LPAR's. Here LPAR A and LPAR B have servers running for all 3 departments and these servers are handling the entire offered workload. As the workload for Department 2 begins to increase and the running servers can no longer satisfy the preconfigured goals, the on-demand router determines that LPAR C has spare bandwidth available.

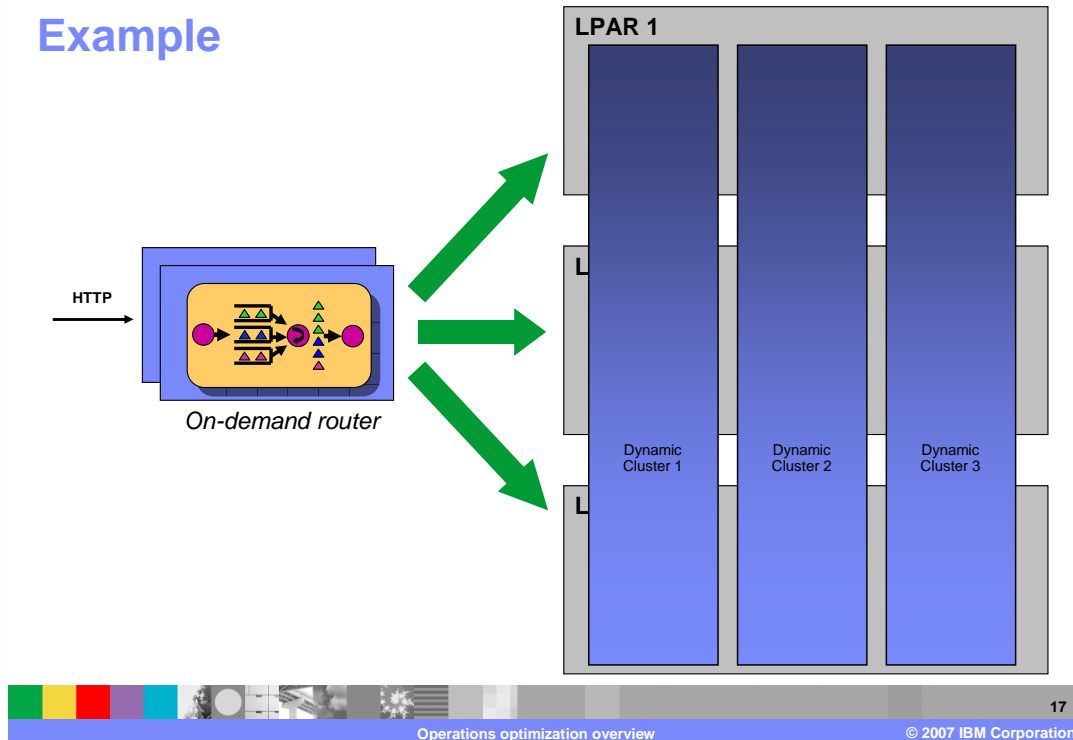
## Example



Here the Automatic Placement Facility in the on-demand router has started Department 2's server on LPAR C and is routing work to the new server in addition to the two existing servers. Later on the offered load for Department 2's server decrease and the offered loads for Department 1 increases. The on-demand router will detect the change and reconfigure the dynamic clusters.



## Example



Here, the third server for Department 2 has been stopped and a new server for Department 1 has been started on LPAR C. This dynamically accommodates the decrease in Department 2's offered load and the increase in Department 1's offered load. Now assume Department 3's offered load increases and the two Department 3 servers can not satisfy its goals. Now the on-demand router 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 the on-demand router is able to maximize the resources available.

## Benefits

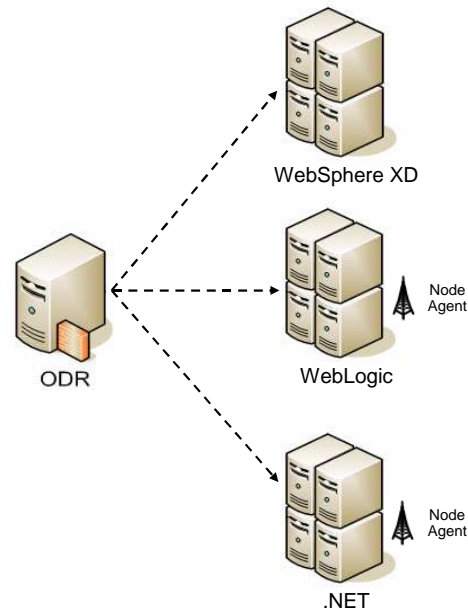
- Enables more efficient hardware utilization
  - ▶ Dynamic allocation of resources to handle variations in traffic
    - Takes advantage of differing peak times
  - ▶ Server consolidation reduces total cost of ownership
- Helps ensure a consistent level of service for critical applications
  - ▶ Decisions 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 response time goals defined by the administrator, WebSphere Extended Deployment can allocate hardware resources to help ensure that applications meet their defined goals. This allows the most important requests to perform better than less important requests when there is contention for resources.

## Integration with other middleware servers

- Full Operations Optimization support for non WebSphere Extended Deployment servers
  - ▶ Any HTTP endpoint, including WebSphere Community Edition, Tomcat, or .NET
- Complete Extended Deployment Dynamic Operations Capabilities
  - ▶ Health Management and Monitoring
  - ▶ Application Placement for Resource Control
- Provided through WebSphere Extended Deployment augmented Node Agent on distributed platforms



WebSphere Extended Deployment operations optimization version 6.1 extends dynamic operations capabilities beyond the traditional WebSphere administrative domain, allowing other middleware server types to be integrated into a WebSphere Extended Deployment environment. The on-demand router can route traffic to any HTTP endpoint. This means that you can use it as a front end to applications running on other vendor's servers, including Tomcat, or .NET; or other versions of WebSphere such as WebSphere Application Server Community Edition. This provides you with all of the benefits of the on-demand router's request prioritization and flow control features. WebSphere Extended Deployment Version 6.1 additionally provides full dynamic operations support for these other middleware servers, including dynamic application placement, through the WebSphere Extended Deployment node agent running on a distributed platform.

## Section

# ***Advanced administration features***

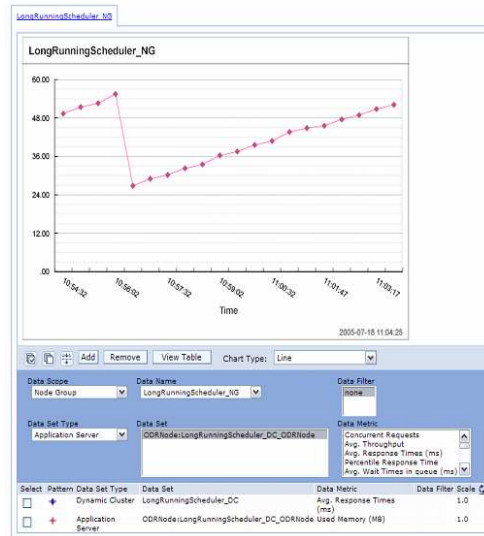


WebSphere Extended Deployment operations optimization includes many advanced features that make life easier for the administrators of a WebSphere environment.

This section will provide a high level overview of these features.

## Runtime visualization

- Custom charting of performance and goal data
- Data can be logged for reference or chargeback
- System health summary panels



With the new complexities of dynamic operations, the need arises for tools that extend monitoring and manageability capabilities. The visualization components of WebSphere Extended Deployment enhance the administrative console to provide live data on the performance and health characteristics of the entire cell.

Real-time reporting shows you alerts indicating anomalies with the runtime environment. You can view the status of the cell based on on-demand routers, core groups, autonomic managers, and nodes. You also have the ability to build and save customized reports for dynamic viewing. These reports can display charts representing numerous statistics about components of your cell, from an application's average response time to the processor utilization of an individual server.

## Application edition management

- WebSphere Extended Deployment supports managing multiple editions of an application in a WebSphere cell
  - ▶ Interruption-free rollout of application updates
  - ▶ Ability to “roll back” to a previous application version
  - ▶ “Validation mode” to verify functionality using a subset of users
  - ▶ Routing policies can be used to define which particular requests are handled by which application edition
- An *edition* is a distinct instance of a J2EE application (similar to *version*)
  - ▶ May be a distinct build version, or the same build with different deployment bindings



WebSphere Extended Deployment allows you to deploy and manage multiple editions of the same application within a cell. You can use the Edition Control Center to manage the rollout of new editions to some or all of your servers without service interruption. You also have the ability to roll back to previous editions if you find a problem with the new edition after deployment. Validation mode deploys the new edition to a clone of the deployment cluster where the previous edition is deployed. This allows you to verify the functionality of the new edition in your production environment. If you have multiple editions of an application deployed concurrently, you can create routing policies that dictate which edition should serve each incoming request. For example, all requests from users in a certain group, or from a certain IP address range, might be served by a different edition than all other users, for testing purposes.

## Health monitoring

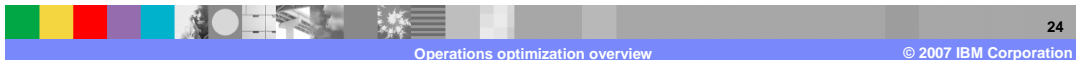
- A WebSphere Extended Deployment environment can be monitored for various software health conditions
  - ▶ Age, work performed, memory usage, response time
  - ▶ Stuck requests, storm drain detection
- Health conditions and reactions are defined by health policies
  - ▶ Servers can be restarted as a corrective or preventive action
  - ▶ Java™ heap can be dumped for memory problems



Health monitoring is another useful administration feature. WebSphere Extended Deployment allows you to configure 'health policies'. The 'health policies' define software conditions that may indicate a problem including excessive memory usage, stuck requests, or uptime and automatically takes a user-defined action if those conditions are matched. For example, if your application has a known bug that causes degraded performance after a week of uptime, you could choose to automatically restart the server weekly. Similarly, you could also choose to generate a dump of the Java heap every time memory consumption exceeded a certain level, and restart the server as well. Note that heap dumps can only be generated automatically on platforms that utilize the IBM Java Runtime Environment.

## Repository checkpoints

- Repository checkpoints give you advanced control over the configuration repository
- A full checkpoint is a snapshot of the repository
  - ▶ Restoring to a full checkpoint returns to the configuration state from the time the snapshot was taken
- A delta checkpoint records the most recent change to your configuration
  - ▶ Delta checkpoints can be rolled back individually, in order
  - ▶ Can be created automatically when changes are made



Repository checkpoints make it easier to manage your WebSphere Extended Deployment configuration. A 'full checkpoint' is a snapshot of your entire configuration repository that you can create at any time, and use in the future to restore your configuration back to that state. You can also configure WebSphere Extended Deployment to create 'delta checkpoints' each time you save changes to your configuration repository. Delta checkpoints contain a record of configuration objects that were changed during the last save action, and you can roll back one or more delta checkpoints to undo a series of changes in reverse order.



## Standby deployment managers

- WebSphere Extended Deployment supports running multiple deployment managers for high availability
- One deployment manager is active; others run in standby mode until a failure is detected
  - ▶ A single master repository is used, stored on a shared file system
- Administrator accesses the deployment manager through the on-demand router
  - ▶ On-demand router routes requests to the current 'active' deployment manager



Although it is not required to have a deployment manager running at all times, you may be interested in running more than one deployment manager concurrently if you require highly available administrative capability. WebSphere Extended Deployment supports redundant deployment managers by allowing multiple deployment manager processes to be concurrently active in a cell in an active - standby configuration. All of the deployment managers work from the same master repository, stored on a shared file system. The on-demand router routes requests to the currently active deployment manager, and to one of the other deployment managers if the originally active one fails.

## Centralized installation manager

- Single installation to the deployment manager
  - ▶ Includes all code for all of the scenarios described for all platforms
- “Push” of installation from deployment manager to endpoints
  - ▶ Select a set of nodes and push extended deployment to those endpoints
  - ▶ Installs appropriate endpoint code based on type of endpoint
- Agent-less endpoint for installation
- Centralization of patch management



The centralized installation manager simplifies the tasks of deploying product packages and maintenance to your WebSphere Application Server Extended Deployment Version 6.1 cell. After you first install WebSphere Extended Deployment Operations Optimization to the deployment manager, you can use the centralized installation manager to install WebSphere Extended Deployment components to other nodes in your cell. You can also use the centralized installation manager to install the middleware agent for operations optimization on a node that does not currently have WebSphere installed.

In addition to product installation, the centralized installation manager provides the ability to download WebSphere maintenance. The maintenance includes interim fixes and fix packs, to the deployment manager and to install those maintenance packages to nodes within your cell.

## Summary

- WebSphere Extended Deployment 6.1 is an 'add-on' product for WebSphere Application Server that provides many advanced features
- Dynamic operations features create a virtualized, goal-based environment for workload management



In summary, WebSphere Extended Deployment is an extension to WebSphere Application Server that provides several advanced features to improve performance and manageability. The dynamic operations features enable a virtualized, dynamic environment for better utilizing shareable hardware resources based on user-defined performance goals.

## Summary

- **Advanced administrative capabilities**
  - ▶ System operations views provides new ways to look at performance data about your environment
  - ▶ Application edition management features enable seamless rollout of multiple editions of an application
  - ▶ Health monitoring can trigger alerts and take action when software health problems are detected
  - ▶ Extended Repository Service provides ability to create cell configuration checkpoints
  - ▶ High availability deployment manager removes single point of failure for administrative activities
  - ▶ Centralized Installation Manager allows you to install products and maintenance from the deployment manager.
  - ▶ Administration and monitoring for non-WebSphere environments



The advanced administration features provided by WebSphere Extended Deployment include many tools to ease the administrative burden for large systems.

## 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 OPS 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      WebSphere      z/OS

J2EE, Java, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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 NON-INFRINGEMENT. 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.