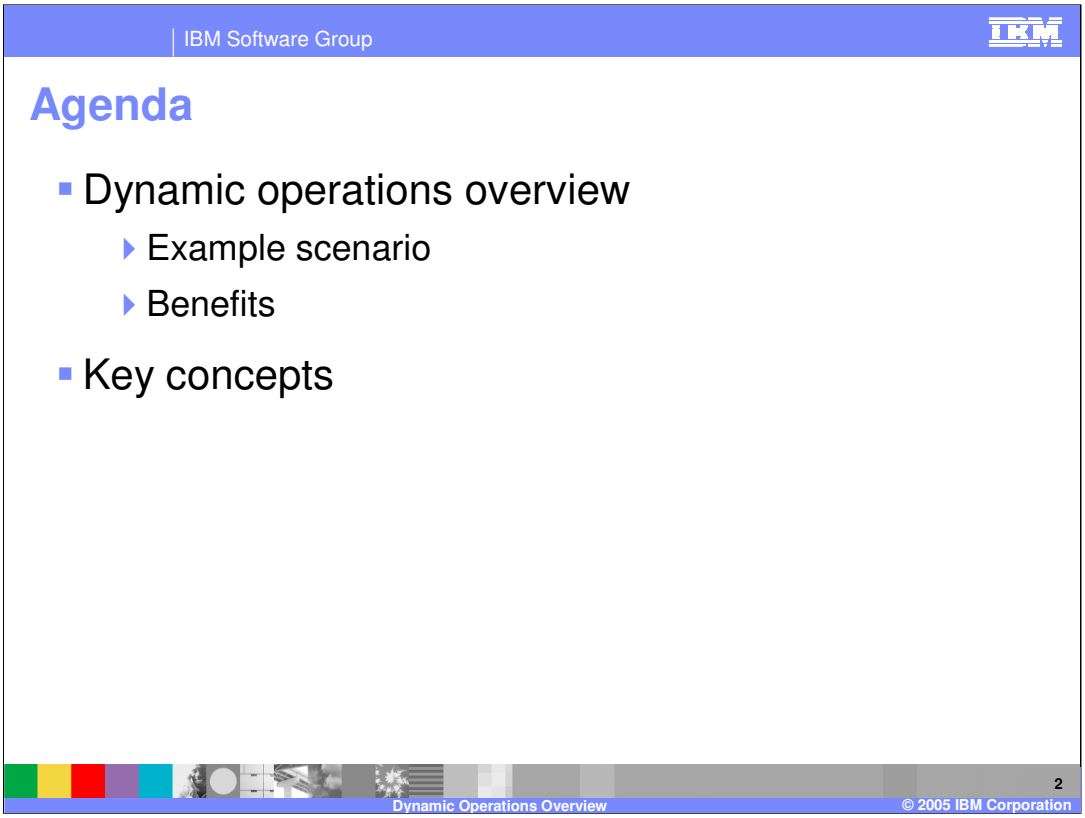
This presentation will give an overview of the dynamic operations features in WebSphere XD version 6.

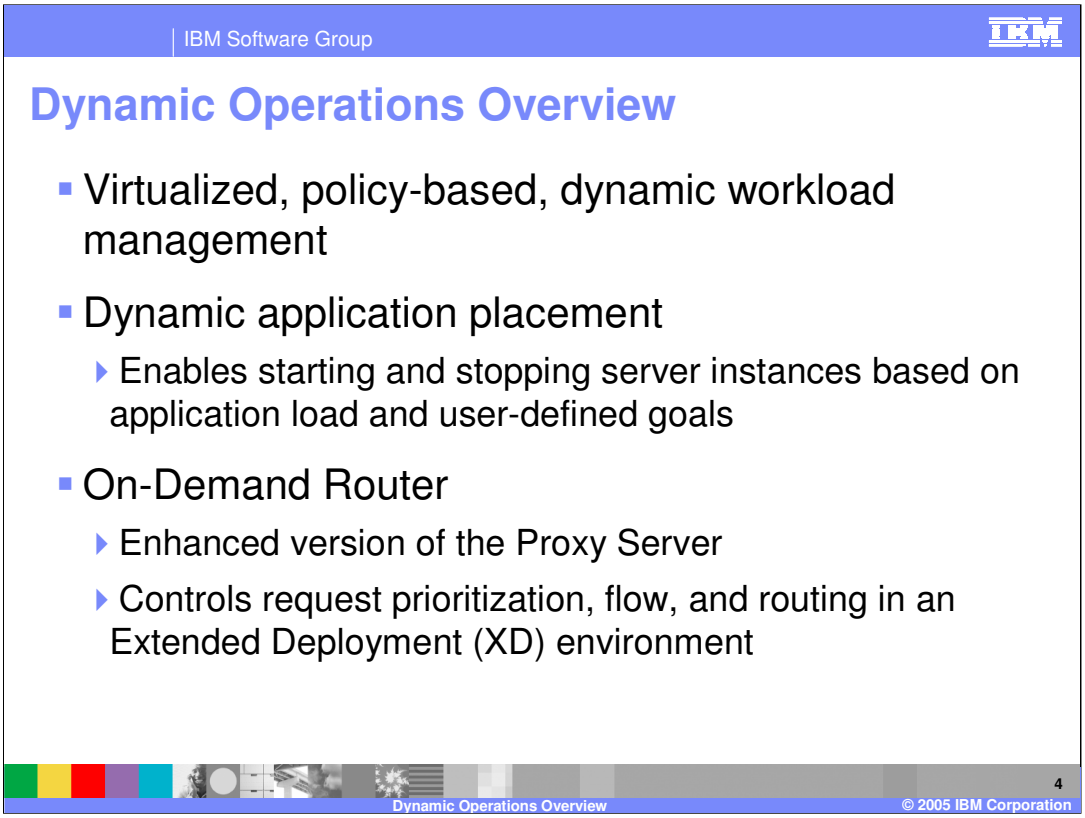


This presentation will begin by illustrating the main ideas behind dynamic operations using a simple example scenario that highlights the benefits of a dynamic WebSphere XD environment. and introduce some of the key concepts involved in creating a dynamic operations-based WebSphere XD environment.

Section

Overview

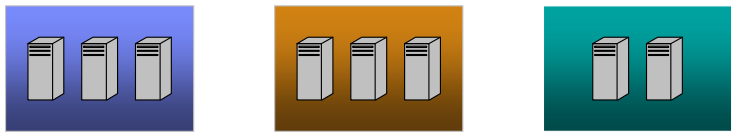
This section will give an overview of dynamic operations.



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

Example Scenario

- Multiple critical applications, each with dedicated resources
  - Supply chain management
  - Inventory reporting
  - Retail sales (browsing, purchasing, etc.)

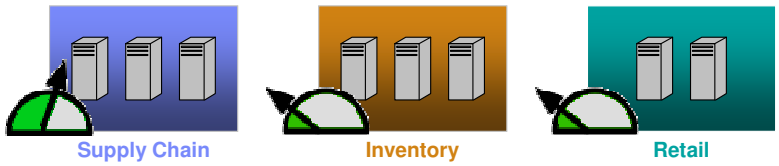


Picture an example scenario in which there are three critical applications deployed to a typical WebSphere Application Server cell. Each of these applications is deployed to its own cluster.

Challenges

- Each application must be sized to support peak load
  - Idle resources during non-peak times
- Inability to use idle resources to handle extra load
  - Quality of service may suffer during periods of exceptional load

6:00 AM

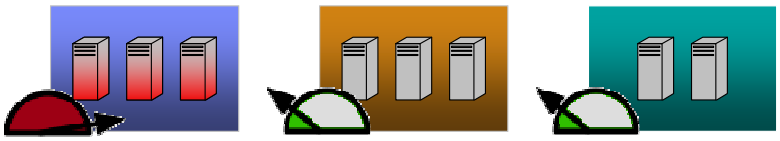


Each of these clusters must be individually sized to support its expected peak load, meaning the hardware resources allocated to each application sits idle during non peak times, and these idle resources cannot be used to handle extra load if other applications are overloaded. Not only does this situation leave you unable to take full advantage of your hardware investment, but the quality of service for each application may suffer during periods of exceptional load. In this example scenario, the supply chain application is heavily loaded early in the day, and completely utilizing its available processing power for a while, leading to lower response times. During this time, the other two applications are not heavily used, but the hardware dedicated to them cannot be used to improve the quality of service for the supply chain application. After low mid-day traffic, the same situation arises when the inventory reporting application becomes overloaded later.

Challenges

- Each application must be sized to support peak load
  - Idle resources during non-peak times
- Inability to use idle resources to handle extra load
  - Quality of service may suffer during periods of exceptional load

6:30 AM



Each of these clusters must be individually sized to support its expected peak load, meaning the hardware resources allocated to each application sits idle during non peak times, and these idle resources cannot be used to handle extra load if other applications are overloaded. Not only does this situation leave you unable to take full advantage of your hardware investment, but the quality of service for each application may suffer during periods of exceptional load. In this example scenario, the supply chain application is heavily loaded early in the day, and completely utilizing its available processing power for a while, leading to lower response times. During this time, the other two applications are not heavily used, but the hardware dedicated to them cannot be used to improve the quality of service for the supply chain application. After low mid-day traffic, the same situation arises when the inventory reporting application becomes overloaded later.

Challenges

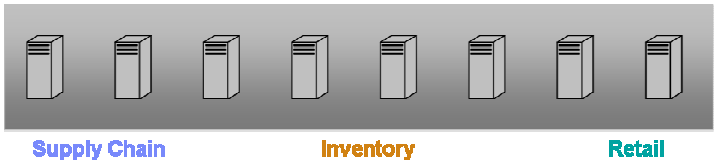
- Each application must be sized to support peak load
  - Idle resources during non-peak times
- Inability to use idle resources to handle extra load
  - Quality of service may suffer during periods of exceptional load

12:00 AM

Each of these clusters must be individually sized to support its expected peak load, meaning the hardware resources allocated to each application sits idle during non peak times, and these idle resources cannot be used to handle extra load if other applications are overloaded. Not only does this situation leave you unable to take full advantage of your hardware investment, but the quality of service for each application may suffer during periods of exceptional load. In this example scenario, the supply chain application is heavily loaded early in the day, and completely utilizing its available processing power for a while, leading to lower response times. During this time, the other two applications are not heavily used, but the hardware dedicated to them cannot be used to improve the quality of service for the supply chain application. After low mid-day traffic, the same situation arises when the inventory reporting application becomes overloaded later.

Dynamic Operations

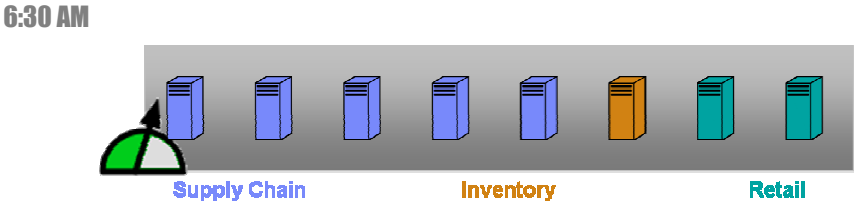
- Resources are treated as virtualized pools of resources (Node Groups)
- User-defined policies specify operational goals and relative importance of applications
- Pooled resources are dynamically allocated to meet the defined goals



In a WebSphere XD environment, rather than statically allocating hardware resources, hardware is treated as a single virtual pool of resources, known as a Node Group. Resources can be dynamically allocated within the Node Group to meet the operational goals that you have defined for your applications. If the example environment were turned into a virtualized WebSphere XD environment, the supply chain application could be dynamically started on more servers to maintain a consistent quality of service despite the increased load. The currently low mid-day retail load can has plenty of room for growth, and the inventory reporting application can take advantage of several servers later in the day to maintain its high quality of service. Each physical node can run multiple applications at the same time, so the division of workload can be more fine-grained than this example demonstrates, because multiple applications may be sharing the same hardware resource. WebSphere XD monitors hardware utilization to carefully balance application needs against the capabilities of each node.

Dynamic Operations

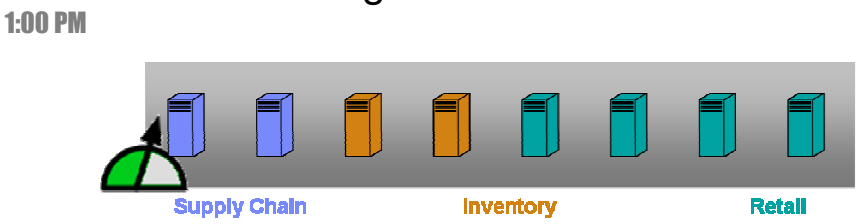
- Resources are treated as virtualized pools of resources (Node Groups)
- User-defined policies specify operational goals and relative importance of applications
- Pooled resources are dynamically allocated to meet the defined goals



In a WebSphere XD environment, rather than statically allocating hardware resources, hardware is treated as a single virtual pool of resources, known as a Node Group. Resources can be dynamically allocated within the Node Group to meet the operational goals that you have defined for your applications. If the example environment were turned into a virtualized WebSphere XD environment, the supply chain application could be dynamically started on more servers to maintain a consistent quality of service despite the increased load. The currently low mid-day retail load can has plenty of room for growth, and the inventory reporting application can take advantage of several servers later in the day to maintain its high quality of service. Each physical node can run multiple applications at the same time, so the division of workload can be more fine-grained than this example demonstrates, because multiple applications may be sharing the same hardware resource. WebSphere XD monitors hardware utilization to carefully balance application needs against the capabilities of each node.

Dynamic Operations

- Resources are treated as virtualized pools of resources (Node Groups)
- User-defined policies specify operational goals and relative importance of applications
- Pooled resources are dynamically allocated to meet the defined goals

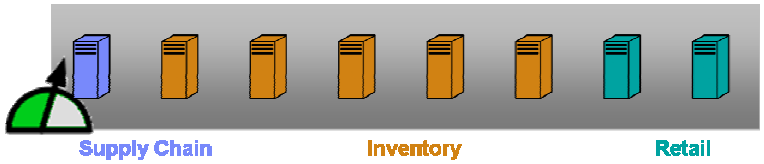


In a WebSphere XD environment, rather than statically allocating hardware resources, hardware is treated as a single virtual pool of resources, known as a Node Group. Resources can be dynamically allocated within the Node Group to meet the operational goals that you have defined for your applications. If the example environment were turned into a virtualized WebSphere XD environment, the supply chain application could be dynamically started on more servers to maintain a consistent quality of service despite the increased load. The currently low mid-day retail load can has plenty of room for growth, and the inventory reporting application can take advantage of several servers later in the day to maintain its high quality of service. Each physical node can run multiple applications at the same time, so the division of workload can be more fine-grained than this example demonstrates, because multiple applications may be sharing the same hardware resource. WebSphere XD monitors hardware utilization to carefully balance application needs against the capabilities of each node.

Dynamic Operations

- Resources are treated as virtualized pools of resources (Node Groups)
- User-defined policies specify operational goals and relative importance of applications
- Pooled resources are dynamically allocated to meet the defined goals

12:00 AM



In a WebSphere XD environment, rather than statically allocating hardware resources, hardware is treated as a single virtual pool of resources, known as a Node Group. Resources can be dynamically allocated within the Node Group to meet the operational goals that you have defined for your applications. If the example environment were turned into a virtualized WebSphere XD environment, the supply chain application could be dynamically started on more servers to maintain a consistent quality of service despite the increased load. The currently low mid-day retail load can has plenty of room for growth, and the inventory reporting application can take advantage of several servers later in the day to maintain its high quality of service. Each physical node can run multiple applications at the same time, so the division of workload can be more fine-grained than this example demonstrates, because multiple applications may be sharing the same hardware resource. WebSphere XD monitors hardware utilization to carefully balance application needs against the capabilities of each node.

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 user, WebSphere XD can allocate hardware resources to help ensure that applications meet their defined goals, allowing the most important requests to perform better than less important requests when there is contention for resources.

Section

Key Concepts

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

Node Groups

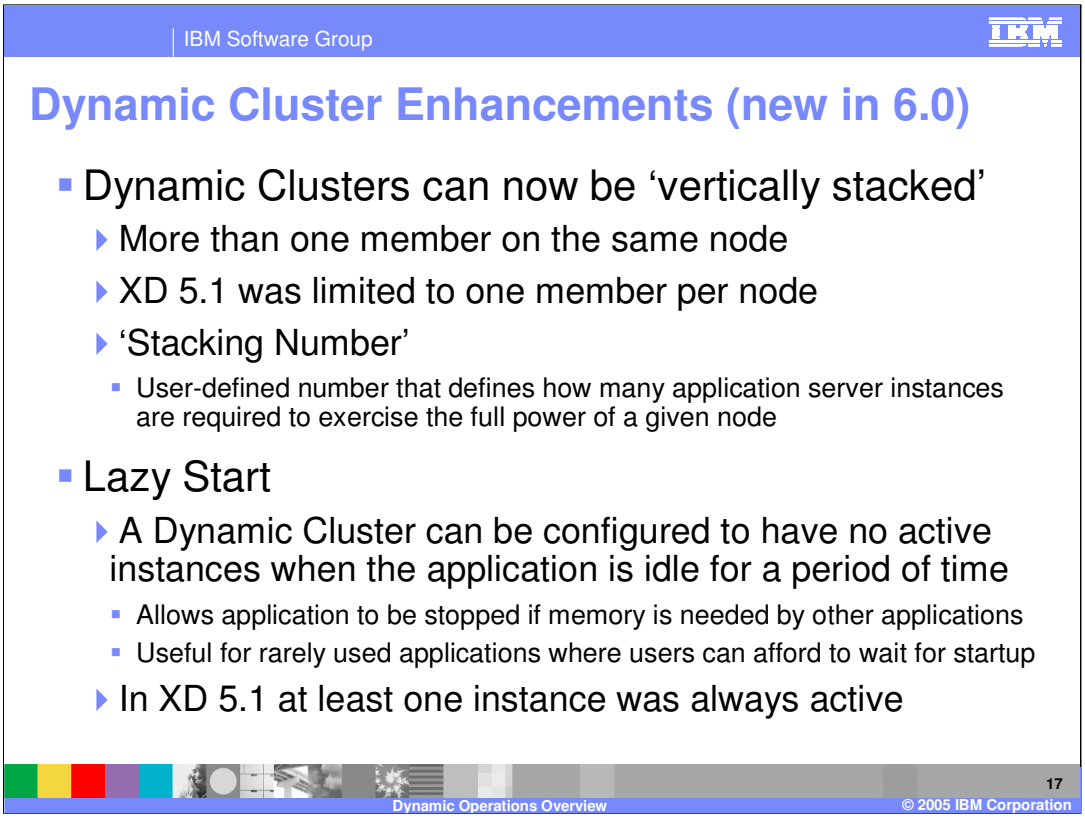
- A Node Group defines a shareable pool of hardware resources (nodes)
  - These nodes have the same capabilities
    - Network, database access, etc.
- Node Groups are a boundary for clusters
  - That is, all members of a cluster must be within the same Node Group
- Node Groups can overlap
  - New in XD 6.0
  - In XD 5.1, Node Group membership was exclusive

A Node Group is a shareable pool of hardware resources, or nodes. All members of a cluster must be contained in the same Node Group. It is important that all members of a Node Group have the same capabilities, such as database drivers or access to network resources, so that they have the ability to run the same set of applications. In WebSphere XD version 6, a node can be a member of more than one Node Group, unlike version5.1, in which node group membership was exclusive.

Dynamic Clusters

- A dynamic group of servers to which applications can be deployed
- Similar to a static Cluster, but can be resized dynamically within a bounding Node Group
- 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 within a Node Group. 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 bounding Node Group to accommodate the changes in load. Each node in the bounding Node Group 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.



In version 6, Dynamic Clusters have gained some new configuration options. 'Vertical stacking' allows more than one instance of a Dynamic Cluster to run on the same node if multiple Java™ Virtual Machines are required to fully utilize the processing power of your machine. This setting may come in handy when using hardware with many processors or when your application is heavily synchronized. Lazy start is a setting that enables you to configure a Dynamic Cluster that will have no active instances after a defined period of inactivity. This setting enables you to keep rarely-used applications installed and ready for use without consuming any server resources. If a request is made for an application with no running instances, the user will have to wait for the application server to start.

On Demand Router

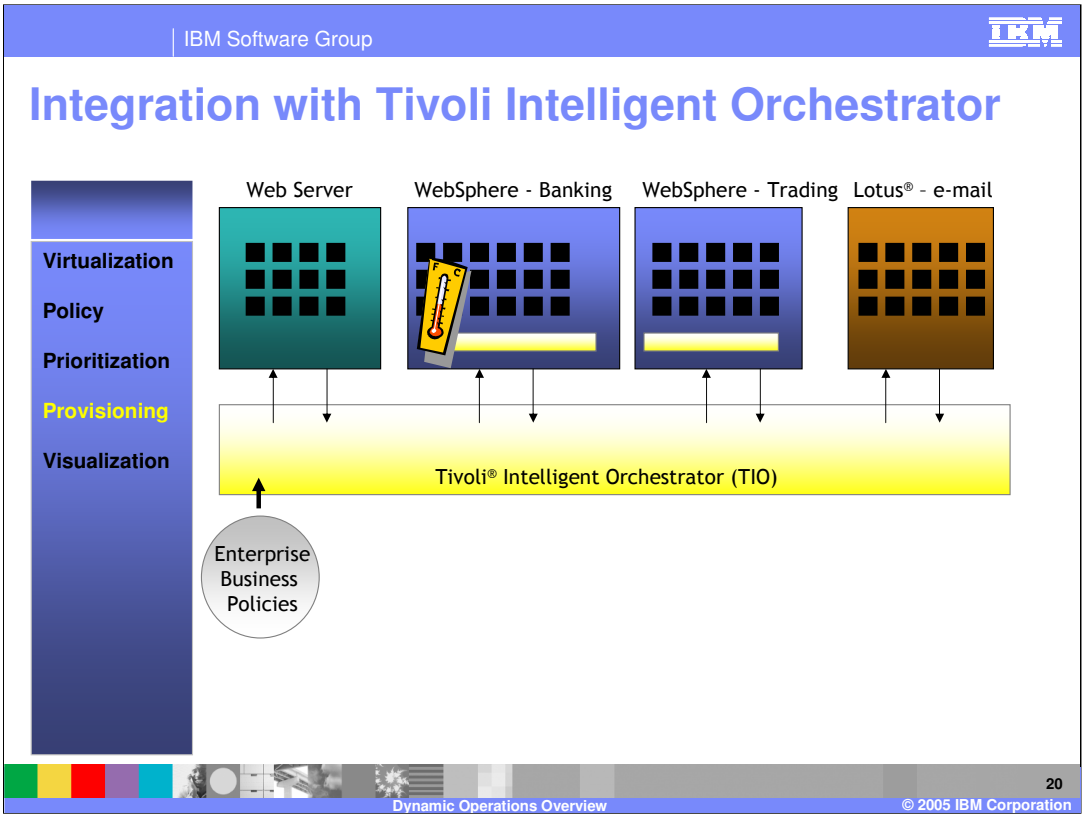
- The On Demand Router (ODR) is an intelligent HTTP proxy server
  - Enhanced version of the Proxy Server from ND 6.0.2
  - Can replace or complement the HTTP server plug-in
  - Prioritizes requests and controls traffic flow according to operational policy
    - Ensures consistent quality of service
    - 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 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. 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 ODR 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 ODR from the HTTP server plug-in, 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.

Application Placement Controller

- The APC 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 (new in XD 6.0)
- One APC per cell
  - XD 5.1 had one APC per node group
  - 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, and uses this information to determine the optimal placement of each application to best meet your defined performance goals. 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 XD can be integrated with Tivoli Intelligent Orchestrator (TIO) 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 XD 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 XD servers, and be added into the WebSphere XD cell to begin hosting Dynamic Cluster instances. WebSphere XD provides the required hooks for operating within a Tivoli Intelligent Orchestrator environment. This capability now works with overlapping node groups, which was not possible in WebSphere XD V5.1.

Summary

- WebSphere XD 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 can be dynamically resized within a Node Group
  - The Application Placement Controller decides when this should take place

In summary, WebSphere XD 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 a virtual pool of resources, called a Node Group. 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 (cont.)

- The On Demand Router is a new component provided by WebSphere XD
  - ODR is an intelligent HTTP proxy server
  - Performs request classification, flow control, and dynamic workload management

Lastly, 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 dynamic cluster members.

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.

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.