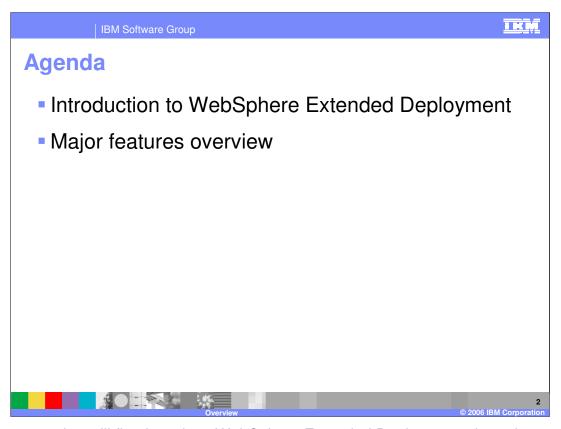
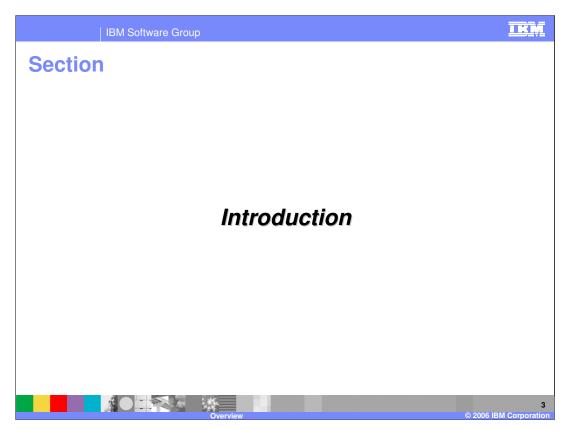


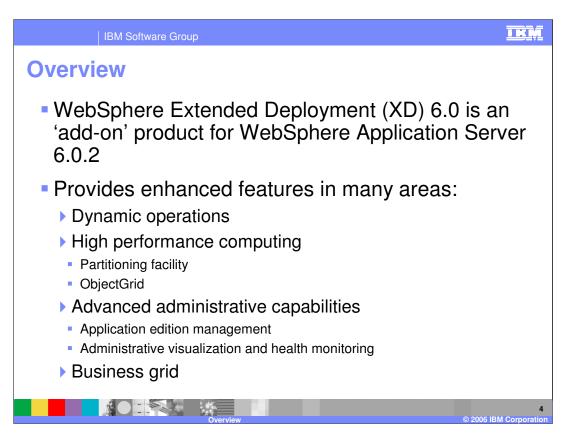
This presentation will give an overview of WebSphere Extended Deployment V6.



This presentation will first introduce WebSphere Extended Deployment, then give an overview of the major features.



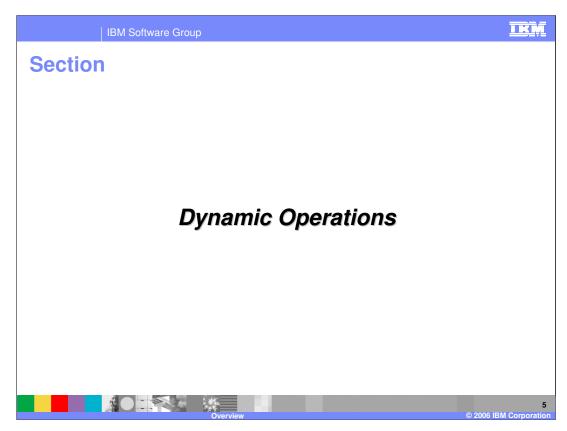
This section will introduce the WebSphere Extended Deployment product.



WebSphere Extended Deployment, also called WebSphere XD, is an add-on product for WebSphere Application Server that provides advanced features for your environment. Version 6 of WebSphere XD requires version 6.0.2 of WebSphere Application Server. WebSphere XD adds features in the following areas:

Dynamic operations features enable a virtualized, dynamic environment for goal-oriented workload management, to take better advantage of hardware resources in a distributed environment. The partitioning facility is a programming framework and runtime environment for implementing high-volume transactional applications that scale linearly with additional hardware, by dividing a large environment into smaller, uniquely addressable 'partitions' to increase efficiency and reduce contention. ObjectGrid is a new high-performance, transactional, and customizable cache framework for Java™ objects. WebSphere XD also provides several advanced features for enhanced administration, including the ability to manage and seamlessly roll out multiple editions of an application; 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. Lastly, Business grid functionality enables you to run and manage long-running and compute-intensive J2EE applications in a WebSphere XD environment designed for those types of work.

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

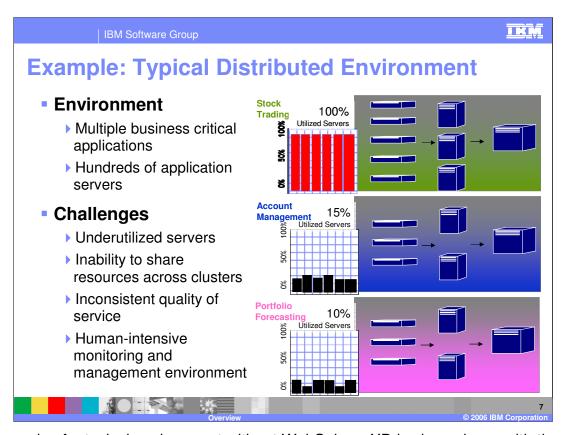


This section will cover the dynamic operations features of WebSphere XD.

Dynamic Operations Overview

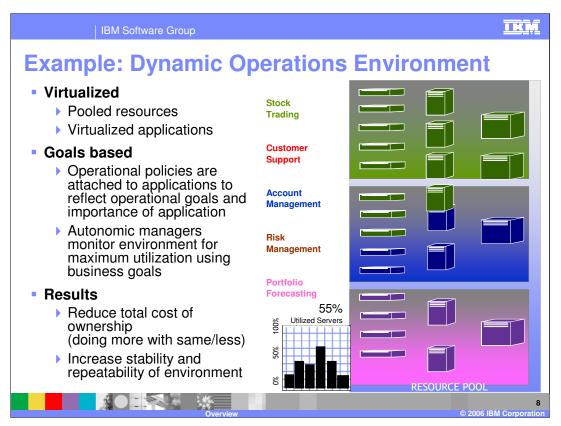
- Virtualized, goal-oriented, policy-based, dynamic workload management
- More consistent quality of service for critical applications
- More efficient utilization of hardware resources

WebSphere XD 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 XD environment.



An example of a typical environment without WebSphere XD 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.

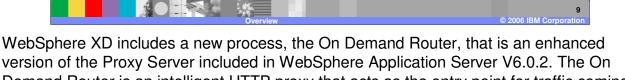
As you can see in this example, each of the three applications has its own cluster, running at low utilization. If for some reason a large amount of traffic hits the stock trading application, its cluster could become 100% utilized, at which point the application's performance will begin to degrade,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



What WebSphere XD does, instead, is provide a virtualized environment. The key concept is called a node group, which is essentially a shareable pool of machines. Within a node group, you define 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 XD 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.

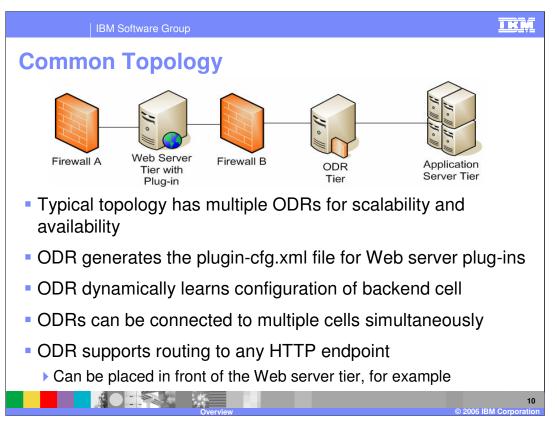
Key Features

- On Demand Router (ODR)
 - 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/Queuing
 - Prioritization
 - Dynamic Workload Management
- Dynamic application placement
 - Autonomic placement of instances of applications on servers
 - Starts/stops preconfigured application instances to meet defined performance goals



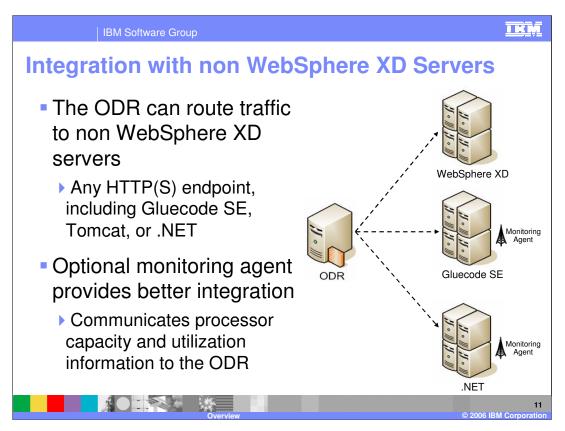
Demand Router is an intelligent HTTP proxy that acts as the entry point for traffic coming into a WebSphere XD cell, performing request prioritization, flow control, and dynamic workload management.

Another key feature is known as dynamic application placement, which uses dynamic 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.

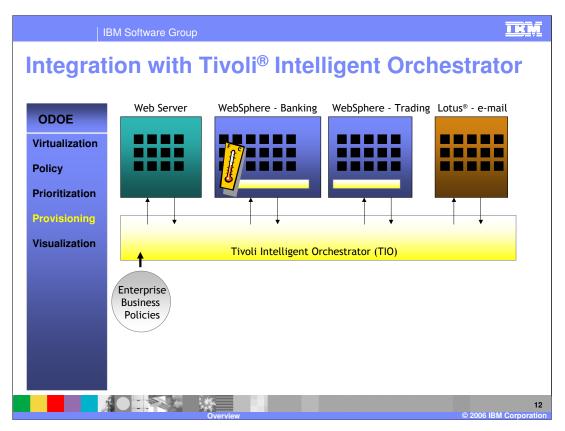


The most common topology places the On Demand Router (or more commonly multiple On Demand Routers) inside the private network between the Web server and the application servers. This leaves the Web server to continue serving static content, but the plug-in running on the Web server is now configured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

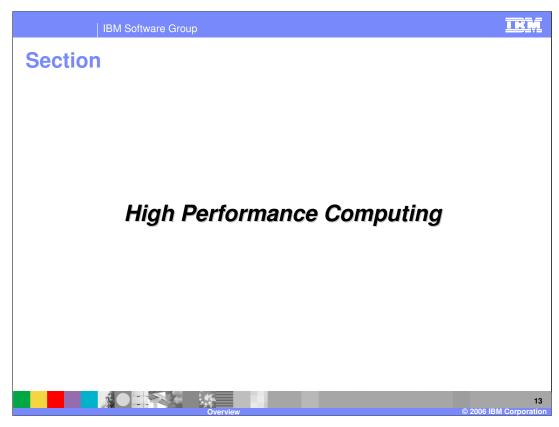
Some less common topology options are also possible, including placing the ODR tier in the de-militarized zone, where it could be used in front of, or instead of the Web server tier.



Non WebSphere XD servers can also be integrated into a WebSphere XD environment. The On-Demand Router can route traffic to any HTTP endpoint, meaning that you can use it as a front end to applications running on third-party servers, including Gluecode SE, Tomcat, or .NET. While dynamic application placement cannot be performed on these non WebSphere XD servers, you can benefit from the On-Demand Router's request prioritization and flow control features. For enhanced integration, you can also install a lightweight monitoring agent on each of the non WebSphere XD servers that will communicate processor capacity and utilization information back to the ODR. This enables the ODR to make better decisions as to how much traffic should be routed to each server.



WebSphere XD also provides hooks for integrating with Tivoli Intelligent Orchestrator, allowing it to participate in larger-scale macroprovisioning with servers of disparate types. For example, if a WebSphere XD node group is fully utilized, TIO could take some servers from an underutilized Lotus Domino[®] environment, apply the appropriate WebSphere XD software and configuration, and add the servers to the node group in question, helping to handle the excess load.



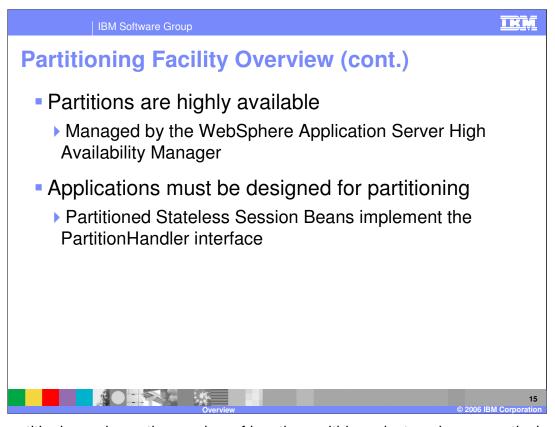
The next section will cover the high performance computing features of WebSphere XD.

Partitioning Facility Overview

- Creates a highly scalable environment for high volume transaction processing
 - Near-linear scalability on distributed hardware
- Application is 'partitioned' across multiple servers
 - A partition is a unique endpoint within a cluster
 - Requests for certain data or certain Enterprise Java™ Beans (EJBs) are always routed to the same partition
- Avoids traditional challenges of scaling
 - Large scale data replication and caching, for example

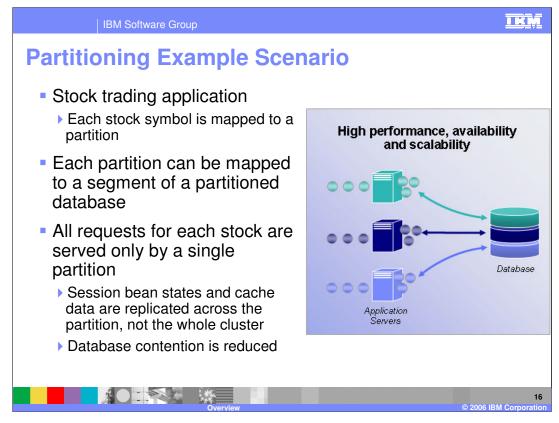


The partitioning facility is a programming framework and runtime environment that makes it possible for high transaction-volume applications to scale linearly by adding hardware capacity. To accomplish this, an application is partitioned across multiple servers in a cluster. Each partition is a uniquely addressable endpoint within the cluster, to which requests for certain data sets are always routed. Partitioning solves some of the traditional challenges of very large clustering, because it can reduce data contention and reduce the overhead of replicating shared data, like caches or state information.



While partitioning reduces the number of locations within a cluster where a particular piece of logic may be running, your entire application remains highly available, because partitions are managed by WebSphere Application Server's High Availability Manager. The High Availability Manager ensures that every partition is running at all times, even in the case of a server failure.

While the standard J2EE programming model is available to partitioned applications, you must also create a special kind of bean, known as a Partitioned Stateless Session Bean, to instruct the partitioning facility about how your application should be partitioned, and how to associate requests with the correct partition, so they can be routed properly.



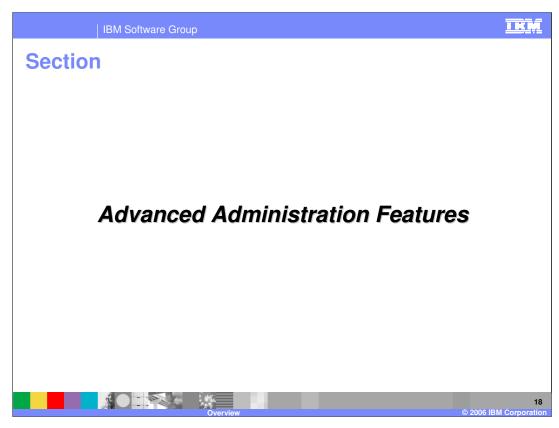
In this example scenario, consider a high-volume stock-trading application that has very high year-to-year traffic growth. The application is partitioned across a cluster of servers, such that buy or sell requests for each stock symbol are routed to a partition associated with that symbol. There may be more partitions than servers in the cluster, meaning each server is running multiple partitions. The work is effectively divided into unique data sets across the cluster, simplifying data replication and reducing overhead. Further, if the underlying database is designed such that separate database instances correspond to the data that would be accessed by individual partitions, then contention on the database will be reduced as well. This type of an environment scales extremely well. Adding additional hardware will simply reduce the number of partitions that must be hosted by each individual server, giving the cluster added capacity without additional overhead normally associated with large-scale clustering.

ObjectGrid Overview

- ObjectGrid is a high-performance, transactional cache framework for storing Java objects
 - Scalable from a local JVM cache to a distributed, partitioned cluster of 100 JVMs
- Can be backed by a hardened data store
- Customizable cache lifecycle features
 - Declaration, configuration, invalidation, size management, cache loading

ObjectGrid is a new WebSphere XD technology that provides a high-performance,

transactional cache framework for storing Java objects. An ObjectGrid can be used as a generic object cache, and optionally persisted to a hardened store. It can in the same fashion also be used as a local cache for objects stored in a database. ObjectGrid is a highly customizable feature, with interfaces provided for custom data loaders, invalidation and size management schemes, and more.



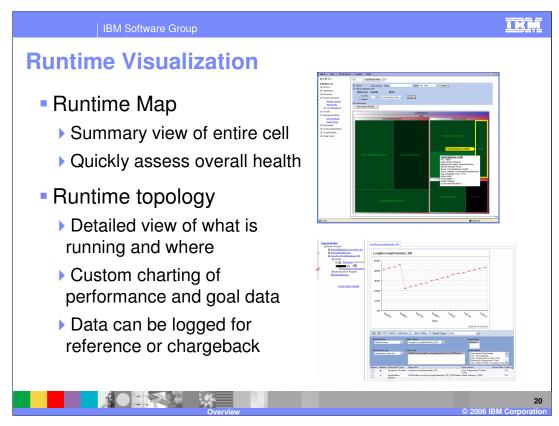
This section will cover the advanced administrative features of WebSphere XD.

Advanced Administration Features

- Visualization tools simplify health and performance assessments
- Edition Control Center supports multiple application editions and seamless rollout
- Health monitoring tools provide means for remediation of software health problems
- Repository checkpoints enable returning to a previously saved 'snapshot' of the configuration repository
- Standby Deployment Managers can be created for administrative high availability



WebSphere XD includes many features that make life easier for the administrator of a WebSphere environment. The visualization components enhance the Administrative Console to provide live data on the performance and health characteristics of the entire cell. The Administrative Console also features the Edition Control Center, which enables management of multiple editions of an application, including interruption-free rollout. WebSphere XD also has health monitoring capability, and can take remediation steps when software health problems are detected. The extended repository service is a new feature that lets you create saved checkpoints of your configuration repository. These checkpoints allow you to later return to a saved configuration, or they can be used as an 'undo' feature for configuration changes. WebSphere XD also lets you create multiple Deployment Managers in a single cell. One of these standby Deployment Managers can take over if the active Deployment Manager fails, so that a Deployment Manager is always available for managing your cell.



The visualization features can be accessed from the Administrative Console. The first view, the Runtime Map, displays an overview of the entire cell for quick assessment of the overall performance relative to defined goals. Color is used to indicate goal-relative performance, and size is used to indicate the amount of work performed, relative to the other servers. You can also drill down or search within this view to get information about a specific server.

The runtime topology view displays a simple overview of your entire topology, and shows where individual components (such as dynamic cluster instances or highly available runtime components) are currently running. You can also display custom charts that can represent 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

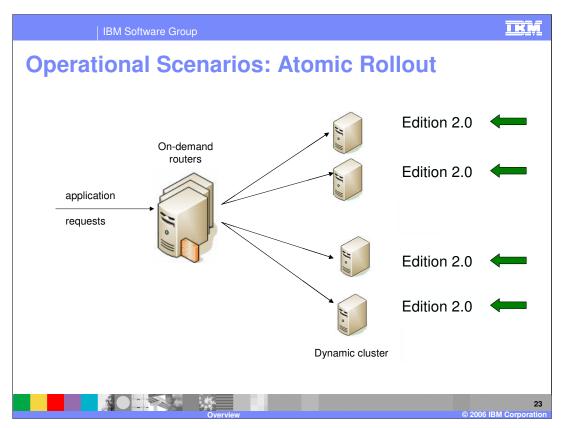
- WebSphere XD supports managing multiple editions of an application in a WebSphere cell
 - Interruption-free rollout of application updates
 - ▶ Ability to "roll back" to a previous application version
 - "Validation mode" to verify functionality using a subset of users
 - Routing policies can be used 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 XD 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, and you also have the ability to roll back to previous editions if you find a problem with the new edition after deployment. 'Validation' of an edition rolls it out to a new cluster that is a clone of the cluster to which a previous edition is deployed, so that you can verify the functionality of the 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.



The Edition Control Center is a new section of the Administrative Console that enables you to manage the various editions of your installed applications. It allows you to see which editions are currently active in your environment, and manage how those editions are rolled out. From here you can choose to roll a new edition out to your cluster one server at a time, or use a 'divide and switch' approach, or to put an application edition into the previously mentioned 'validation mode'.



This example illustrates one of the available rollout methods, called 'atomic rollout'. It uses a 'divide-and-switch' technique to rollout a new edition across a cluster while ensuring that only one edition is actively serving traffic at any given time. The On Demand Router will first quiesce traffic to half of the servers in the cluster, and when existing requests complete, the new edition is installed to those servers, and they are restarted. Traffic will then be quiesced on the other half of the servers. While waiting for those requests to finish, there may be a brief build-up of requests in the On Demand Router's dispatch queue, since no servers are receiving traffic. Once the existing requests have been served, requests (including those that built up in the queue) will start being routed to the first set of servers, which is now running the new edition of the application. The new edition will then be installed on the second set of servers. Once they have been restarted, the entire cluster will be running the new application edition.

Note that since your cluster will effectively be running at half capacity during this process, it is wise to perform this type of an edition upgrade during a low-traffic time period.

Health Monitoring

- A WebSphere XD environment can be monitored for various software health conditions
 - ▶ Age, work performed, memory usage, response time
 - ▶ Stuck requests, storm drain detection (new in XD 6.0)
- 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

platforms that utilize the IBM Java Runtime Environment.

Health monitoring is another useful administration feature. WebSphere XD allows you to configure 'health policies', which define software conditions that may indicate a problem including excessive memory usage, stuck requests, or simply uptime, and take 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

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 are a new feature that make it easier to manage your WebSphere XD 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 XD 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.

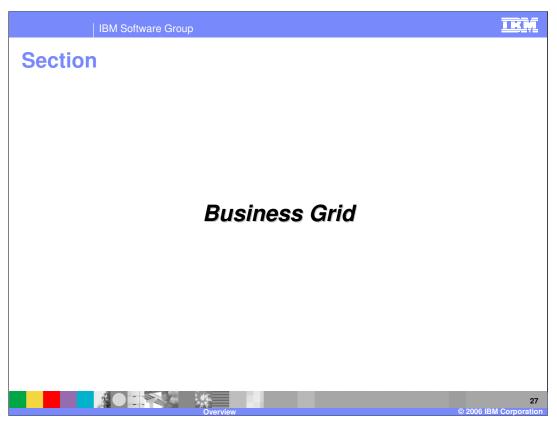
Standby Deployment Managers

Managers if the originally active one fails.

- WebSphere XD 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 (ODR)
 - ODR 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 XD 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



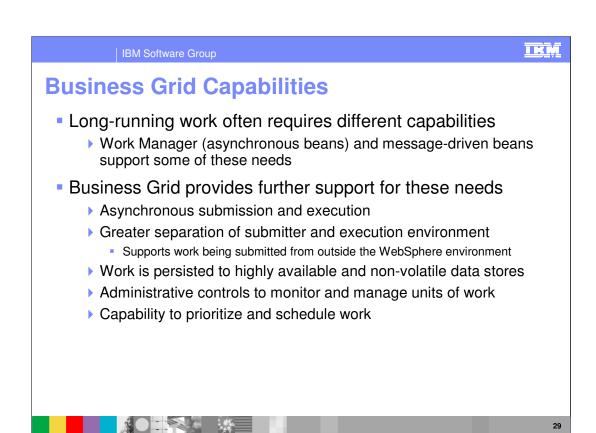
This section will cover the Business Grid feature.

Business Grid Overview

- WebSphere Application Server has traditionally focused on transactional applications
 - Applications typically are designed to handle large volumes of relatively small tasks
 - Not all applications fit this type of design
- Business Grid enhances WebSphere Application Server to support long-running applications
 - Provides capabilities to deploy long-running applications along with transactional applications
- It is usually preferable to run long-running applications and transactional work on different servers
 - Business Grid balances the work across different nodes within a cell



Business Grid is a new feature that expands the types of work that WebSphere XD is designed to handle. Applications designed to run on WebSphere Application Server tend to be transaction-oriented applications, made up of several relatively short-lived tasks. Business Grid adds to WebSphere XD the capability to run and manage long-running applications in the same cell as your transactional applications. WebSphere XD dynamically manages both workload types and allocates server resources as necessary, while keeping the two types of applications separate, since they tend to have different runtime needs.



WebSphere Application Server has some features that can be beneficial to long-running applications, such as the work manager and message-driven beans for running asynchronous tasks in the EJB container, but WebSphere XD expands on these features by providing a full set of capabilities for running and managing long-running work. These features include a special execution environment for long-running work, a separate job submission and scheduling environment, and the administrative tools to monitor and manage long-running work.

Types of Long Running Work

- WebSphere Extended Deployment V6.0 supports two styles of long-running work
- Batch applications focus on doing large amounts of work based on a specific task, for example record processing
 - Application will provide logic for a single unit of work (process one record)
 - Container provides support for transactions and mechanisms checkpoint and restart work
- Compute-intensive applications focus on large amounts of processor-bound work
 - Container provides thread of execution and has limited contact with the work after it is started
 - Application provides all other logic



Long-running work can generally be divided into two types, both of which are supported by WebSphere XD. A batch application is designed to repeatedly perform some specific unit of work, such as processing a database record, over a large set of data. If an application provides the logic for performing the work, WebSphere XD can manage the transaction support and lifecycle aspects, such as restarting the work at the point it was stopped in the case of an interruption. Compute-intensive applications are also long-running tasks with little user interaction, but focus on performing a large amount of work that tends to be processor-bound, such as large-scale simulation work. In this case WebSphere XD provides a runtime environment and a thread of execution, and all other logic is implemented by the application.

Summary

- WebSphere XD 6.0 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
- High performance computing

- ▶ The partitioning facility provides a means to create a highly scalable environment for high-volume transaction processing
- ObjectGrid is a high-performance, transactional, and extensible cache framework

In summary, WebSphere Extended Deployment is an add-on product that provides several advanced features for WebSphere Application Server. The dynamic operations features enable a virtualized, dynamic environment for better utilizing shareable hardware resources based on user-defined performance goals. The partitioning facility provides a programming model and runtime environment for implementing highly scalable solutions for transaction processing. ObjectGrid is a new high-performance, transactional, and extensible object cache for Java applications.

Summary (cont.)

- Advanced administrative capabilities
 - Administrative visualization 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
- Business grid
 - An environment for running and managing computeintensive and batch styleJ2EE applications



The advanced administration features provided by WebSphere XD include advanced visualization tools for monitoring the performance of your environment, the Edition Control Center for the management and seamless deployment of multiple editions of an application, and health monitoring features that can detect software health problems and take actions on your behalf. Lastly, Business Grid provides an environment for running and managing compute intensive and batch style J2EE applications within a WebSphere XD cell.



Template Revision: 3/09/2005 9:40 AM

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:

CICS Cloudscape DB2 DB2 Universal Database IBM IBM(logo) e(logo)business AIX

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.

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 and objectives only. References in this document in this document in this document is not intended to make such products, programs, or services available in all countries or 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 program product 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 OF IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY. FITHES 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, 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 2004, 2005. All rights reserved.