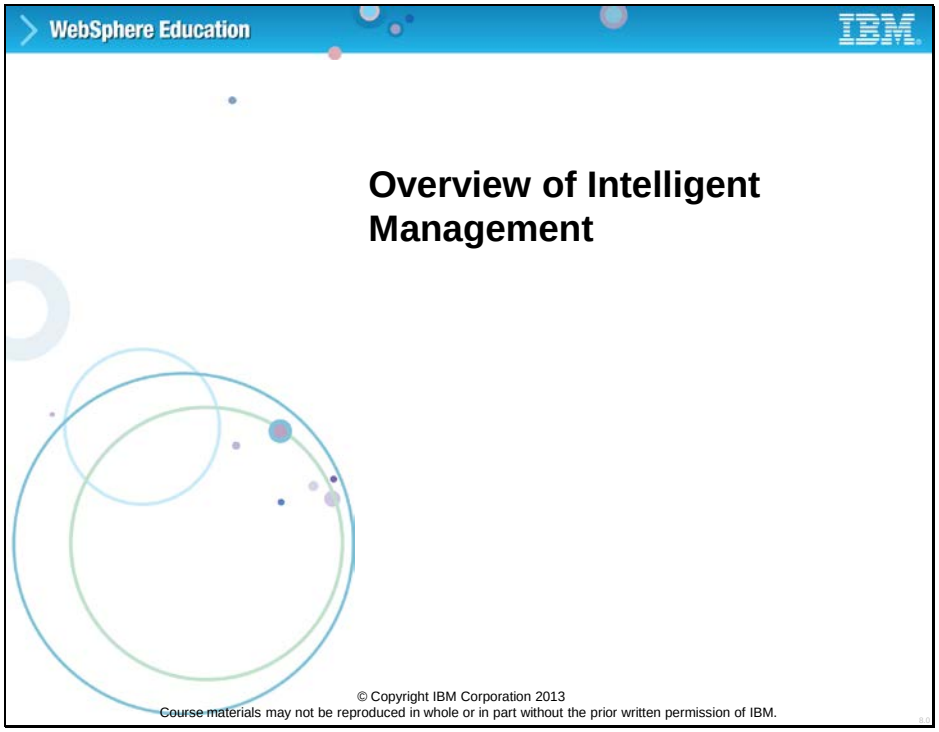
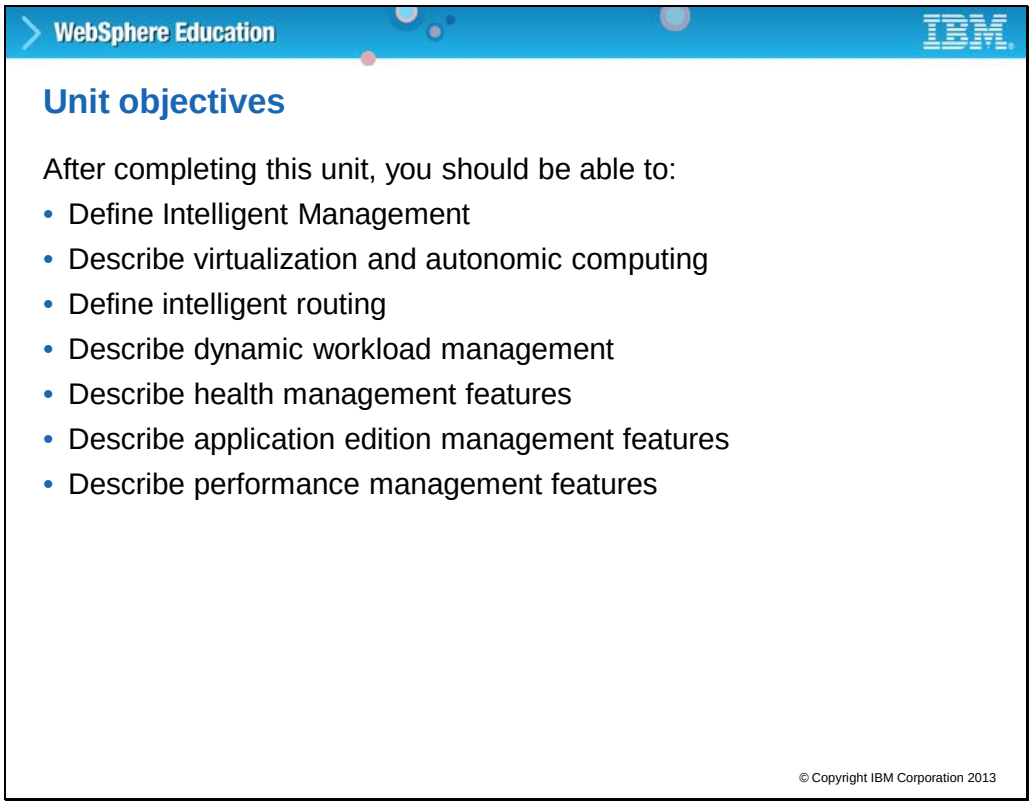
Slide 1



Overview of Intelligent Management. This unit introduces the concepts of the Intelligent Management feature of WebSphere Application Server.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 2

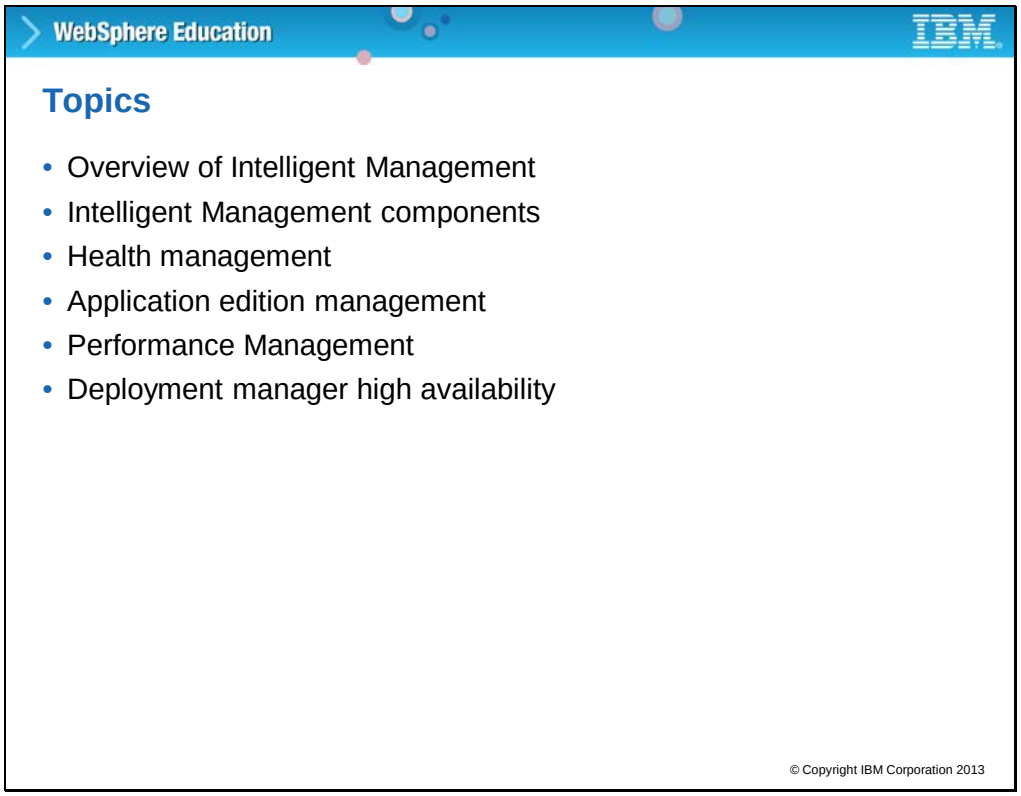


After completing this unit, you should be able to:

- Define Intelligent Management
- Describe virtualization and autonomic computing
- Define intelligent routing
- Describe dynamic workload management
- Describe health management features
- Describe application edition management features
- Describe Performance Management features

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

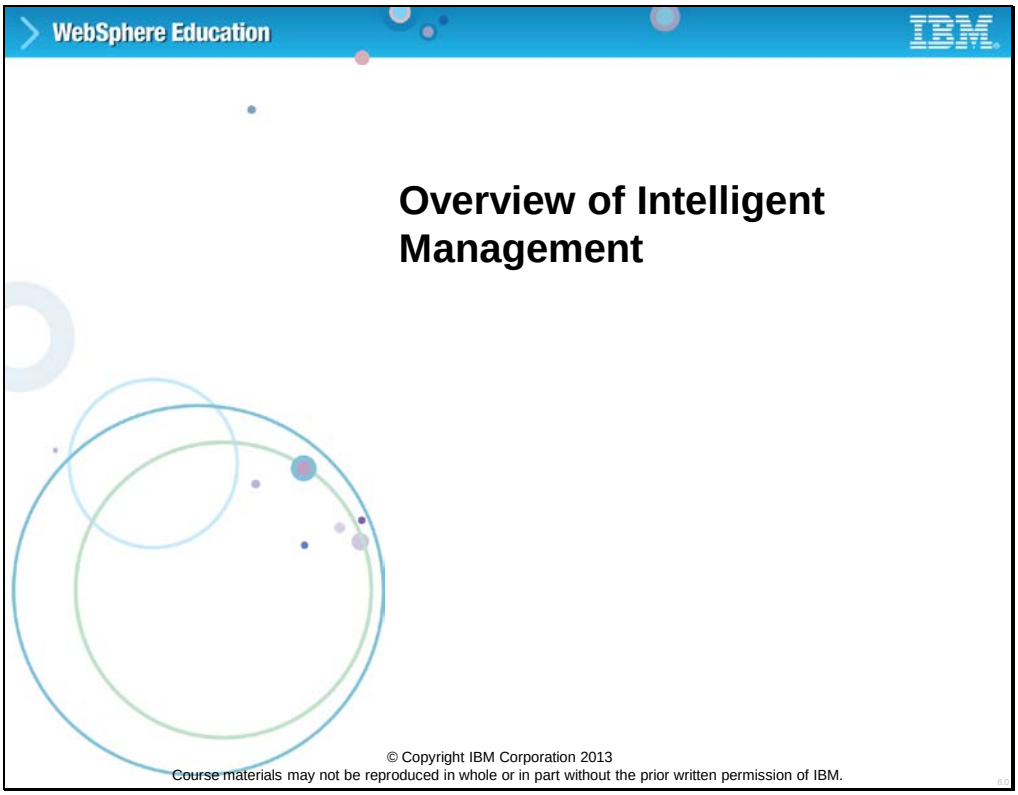
Slide 3



This unit is divided into six topics.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 4

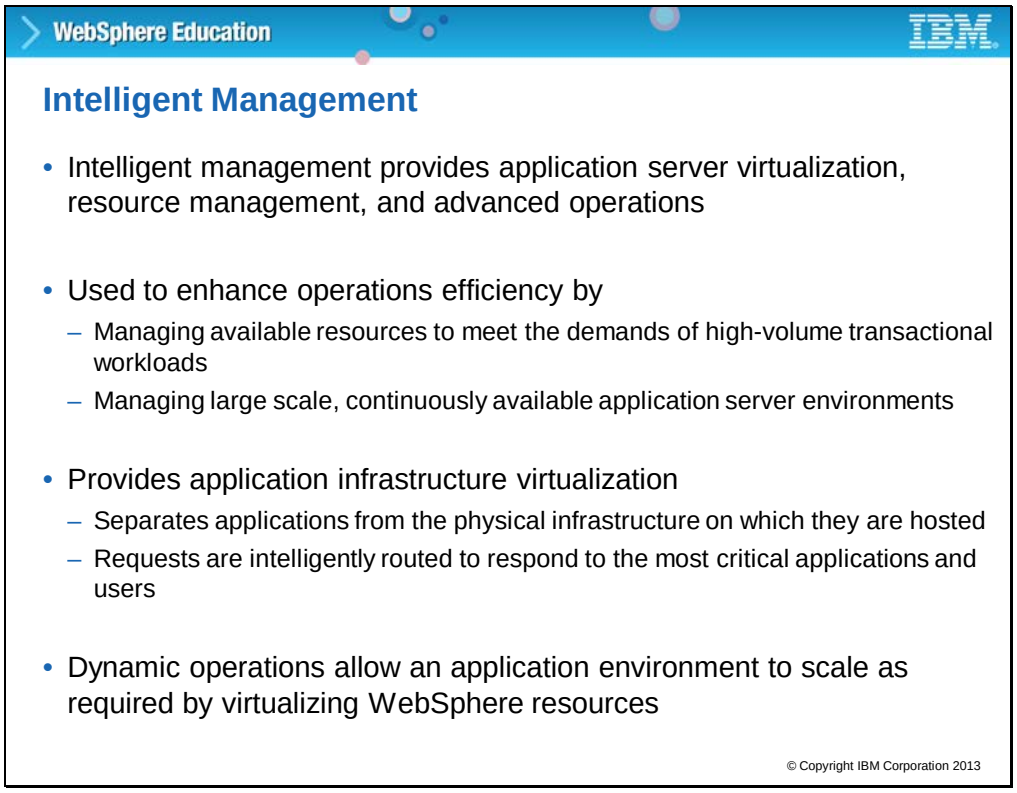


Topic: Overview of Intelligent Management. In this topic, you get an overview of Intelligent Management.

–

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 5



Intelligent Management provides a virtualized infrastructure that redefines the traditional concepts of Java Platform, Enterprise Edition (Java EE) resources and applications, and their relationships with one another. This application infrastructure virtualization facilitates the ability of the product to automate operations in an optimal manner, increasing the quality of service. By introducing an automated operating environment with workload management, you can reduce total cost of ownership by completing more work by using less hardware.

Intelligent Management capabilities are available with WebSphere Application Server Network Deployment V8.5. It provides a virtualized infrastructure that redefines the traditional concepts of Java EE resources and applications and their relationships with one another.

The dynamic operations environment consists of autonomic managers whose purpose is to maximize utilization by using defined business goals. Dynamic operations allow an application environment to scale as required by the virtualization of WebSphere resources and the use of a goals-directed infrastructure. Therefore, you can increase the speed at which your environment adapts to the business requirements. Using the dynamic operations features of WebSphere Application Server, you can change the way a typical WebSphere environment is configured to one that has the following features:

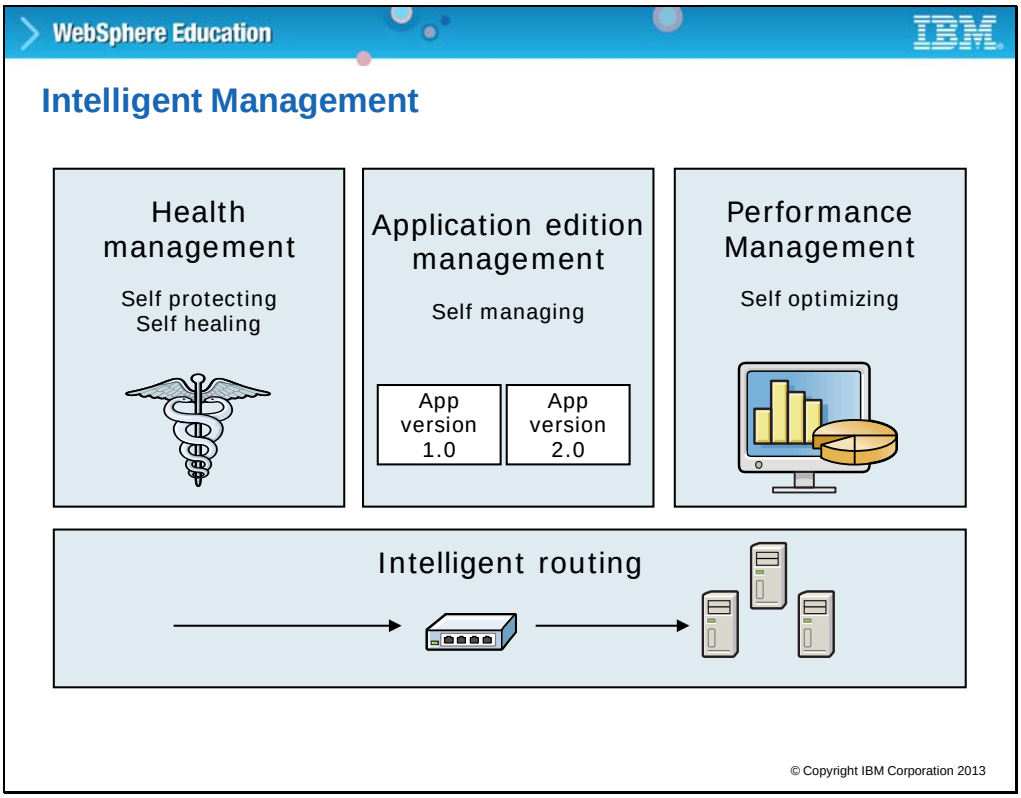
- Improves the utilization of available resources such as processor and memory
- Classifies and monitors the workload

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

- Provides a business-centric view of the workload and how it functions
- Responds in real time to changes in the workload mix (without human intervention if so wanted), by using business guidelines that the organization specified

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 6



The Intelligent Management features extend the quality of service that the middleware environment provides. Configurable operational policies govern the performance and health of your applications.

Intelligent routing improves business results by ensuring that priority is given to business critical applications. Requests to applications are prioritized and routed based on rules that are defined.

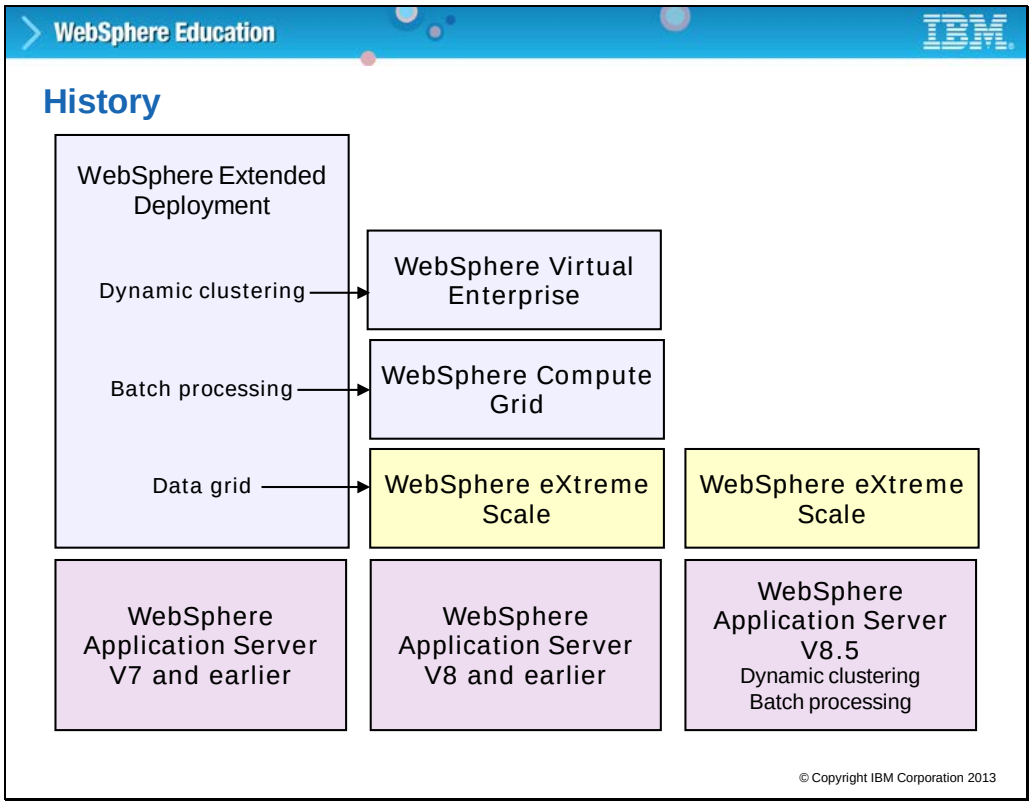
You can use health management to monitor conditions automatically and take corrective actions when the conditions are observed. You can monitor the status of application servers, sense problem areas, and then respond to these problem areas before an outage occurs.

You can use application edition management to roll out new versions of applications without experiencing downtime for a maintenance window. You can use this feature to manage interruption-free production application deployments.

Performance Management provides a self-optimizing middleware infrastructure. You can use dynamic clusters to scale up and out the number of running cluster members automatically as needed to meet response time goals for users.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

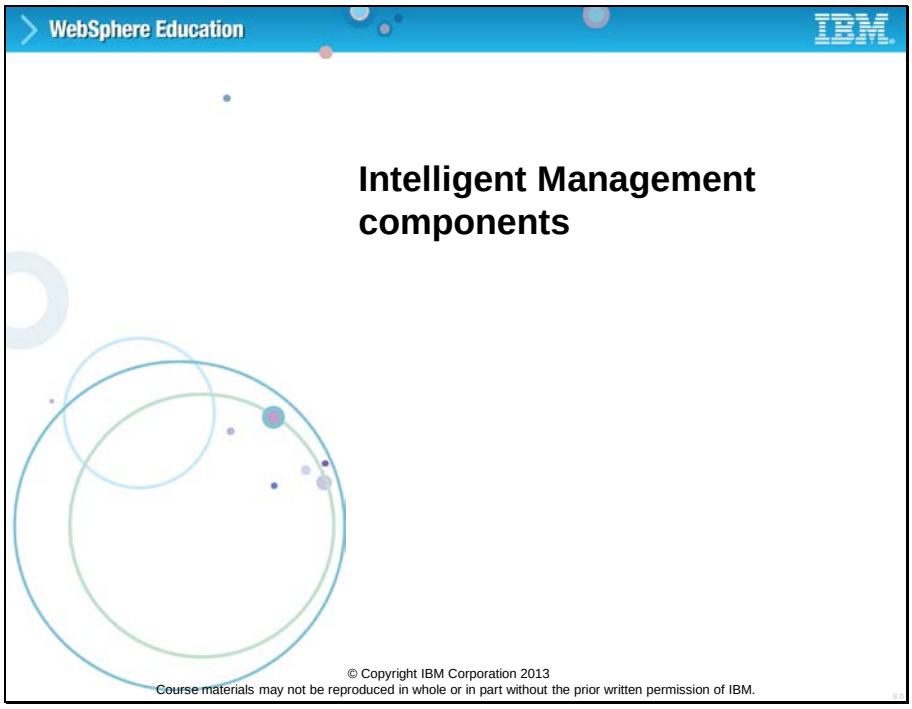
Slide 7



The biggest news under the theme of application resiliency is the integration of the features from WebSphere Virtual Enterprise into WebSphere Application Server Network Deployment V8.5. This merge allows for a single WebSphere Application Server Network Deployment installation to deliver the traditional Network Deployment functions, and WebSphere Virtual Enterprise functions. The former WebSphere Virtual Enterprise functions now in Network Deployment V8.5, are characterized as Intelligent Management features and encompass intelligent routing, application edition management, dynamic clustering, and health management.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

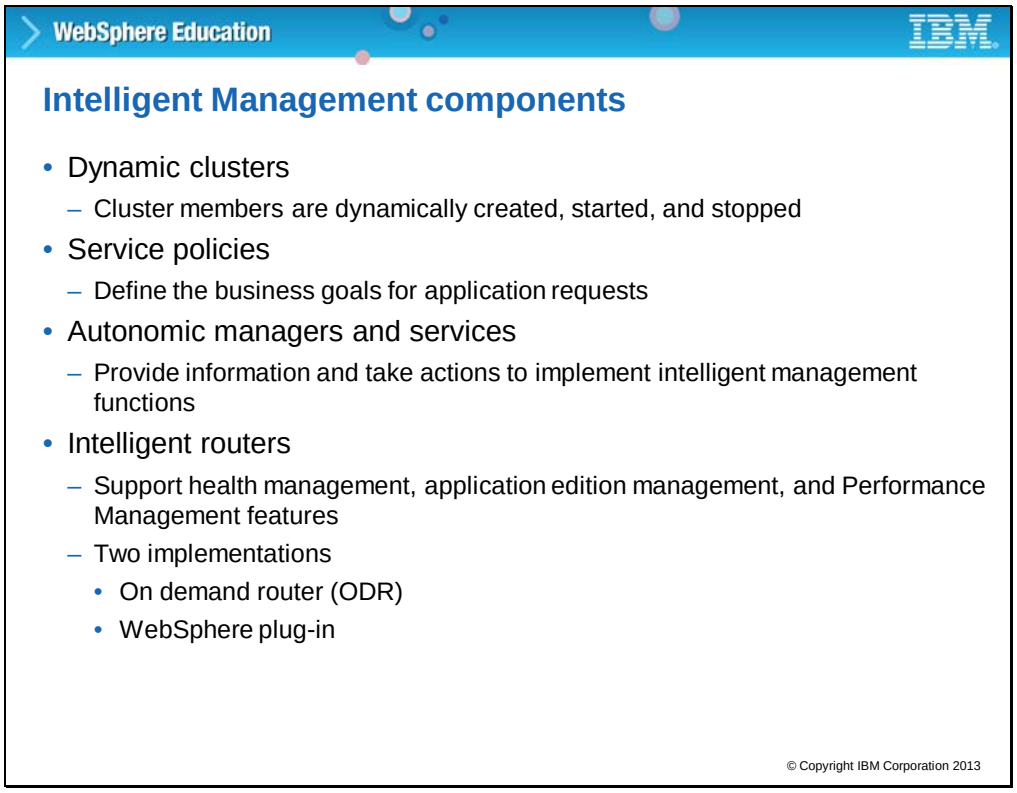
Slide 8



Topic: Intelligent Management components. In this topic, you learn about the components that work together to provide Intelligent Management functions.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

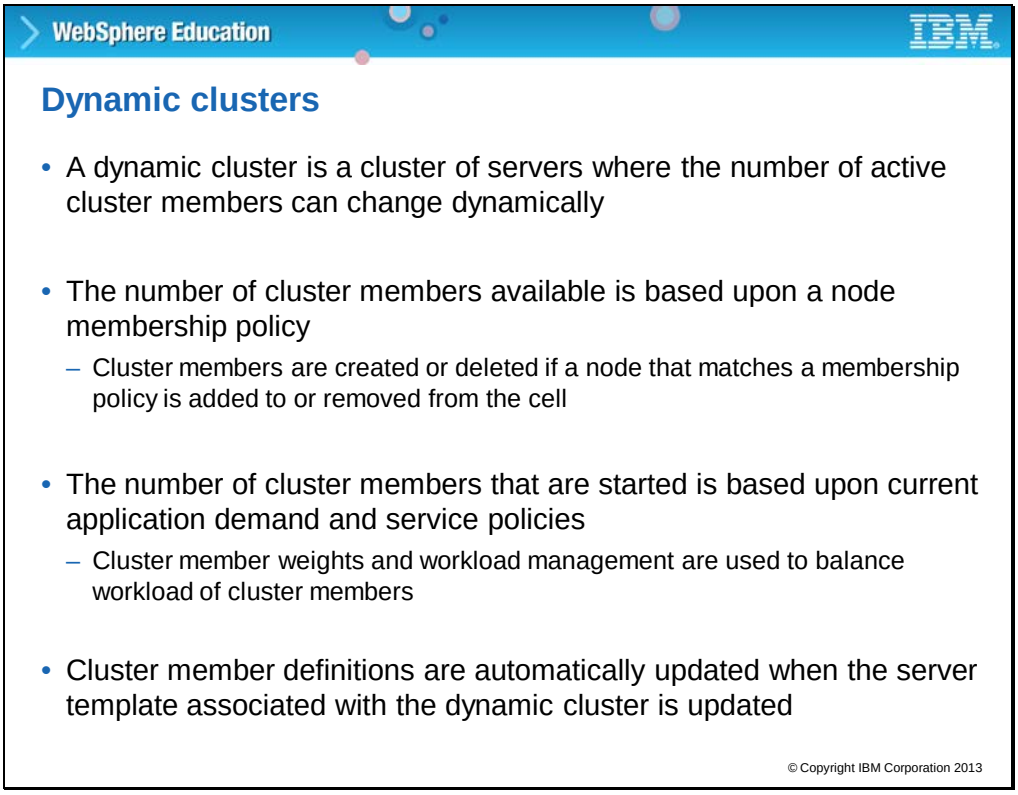
Slide 9



Components that Intelligent Management uses include dynamic clusters, service policies, autonomic managers and services, and intelligent routers. These components are described more fully in the next few slides.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 10

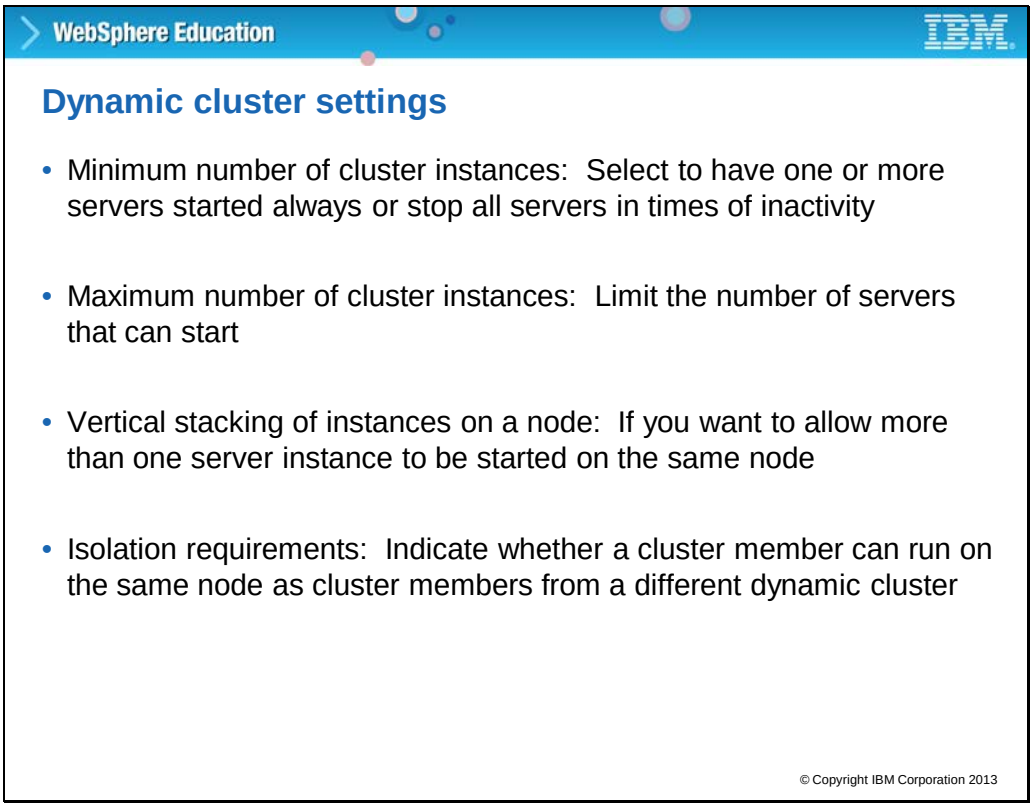


A dynamic cluster is a server cluster that enables application server virtualization. Members of a dynamic cluster are automatically created based on a membership policy, which is updated by using a server template, and started and stopped based on current demand, available resources, and service policies.

Dynamic clusters allow the application environment to dynamically expand and contract based on the amount of workload that can be handled at any time.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

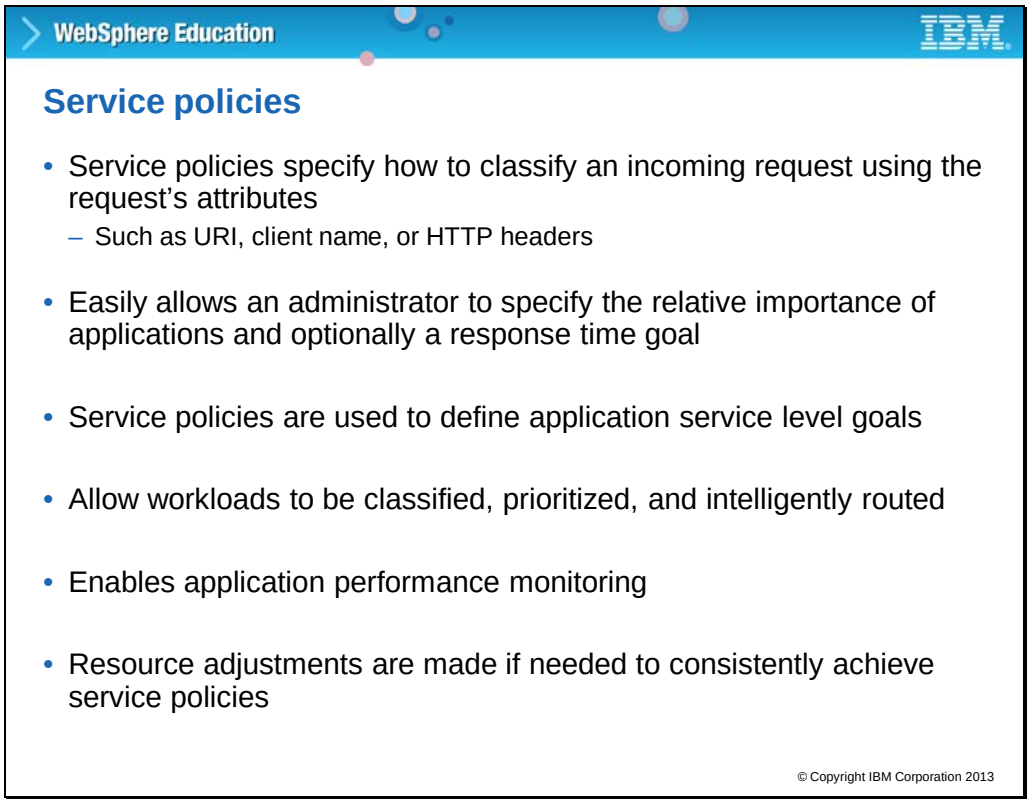
Slide 11



When you define a dynamic cluster, you can specify the minimum number of cluster members that must be started, and the maximum number of cluster members that can be started. You can specify how any cluster members for the dynamic cluster can exist on a single node. You can also specify which cluster members from other dynamic clusters can start on the same node with these cluster members.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

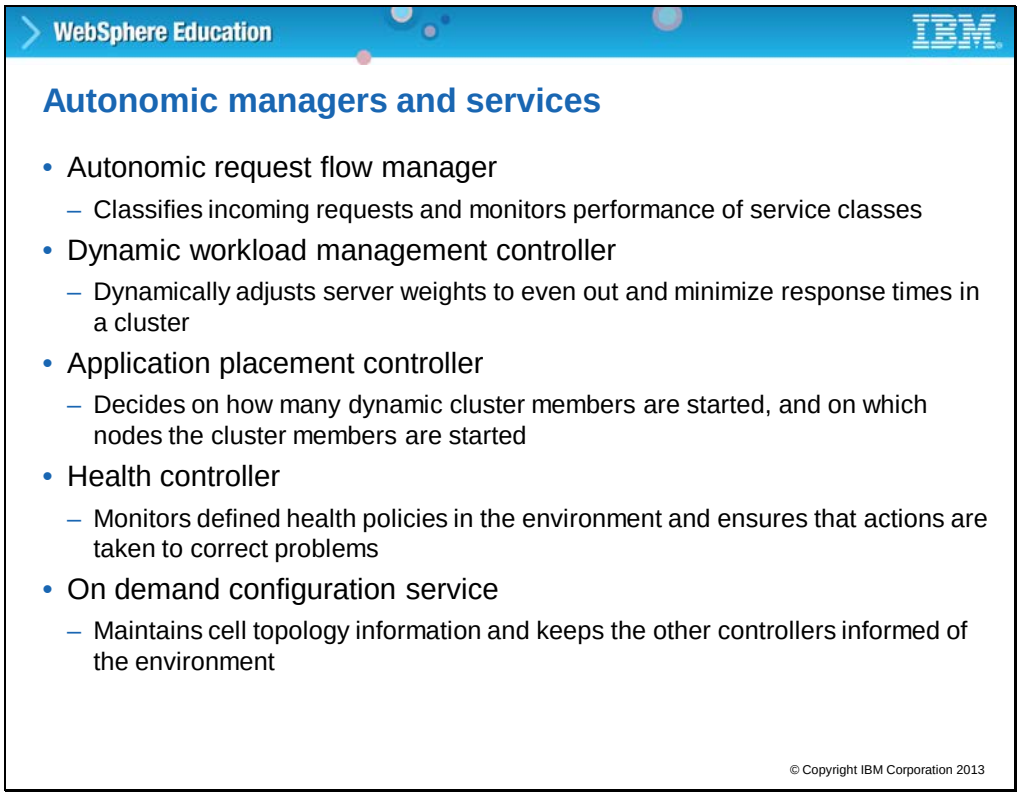
Slide 12



A service policy is a user-defined categorization that is assigned to potential work as an attribute that the application request flow manager (ARFM) reads. You can use a service policy to classify requests that are based on request attributes. These attributes include the Uniform Resource Identifier (URI), the client name and address, HTTP headers, query parameters, cookies, time of day, and others. By configuring service policies, you apply varying levels of importance to the actual work. You can use multiple service policies to deliver differentiated services to different categories of requests.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 13



With Intelligent Management, you can introduce autonomic capabilities into your infrastructure at your own pace. Autonomic capabilities are delivered in a set of components that are known as autonomic managers. Autonomic managers monitor performance and health statistics through a series of sensors, and optimize system performance and run traffic shaping.

The autonomic request flow manager manages traffic shaping of incoming requests. The dynamic workload controller dynamically adjusts server weights to even out and minimize response times across the cluster. There is one dynamic workload controller per cluster. The dynamic workload controller maintains a list of active server instances for each dynamic cluster, and assigns each a routing weight according to observed performance trends.

The application placement controller is responsible for the management of the location of an application within a node group. A single application placement controller exists in the cell and is hosted in the deployment manager or in a node agent process. The application placement controller starts and stops application server instances to manage HTTP, SIP, JMS, and IIOP traffic.

The health controller is used for managing health conditions and is covered in the next topic.

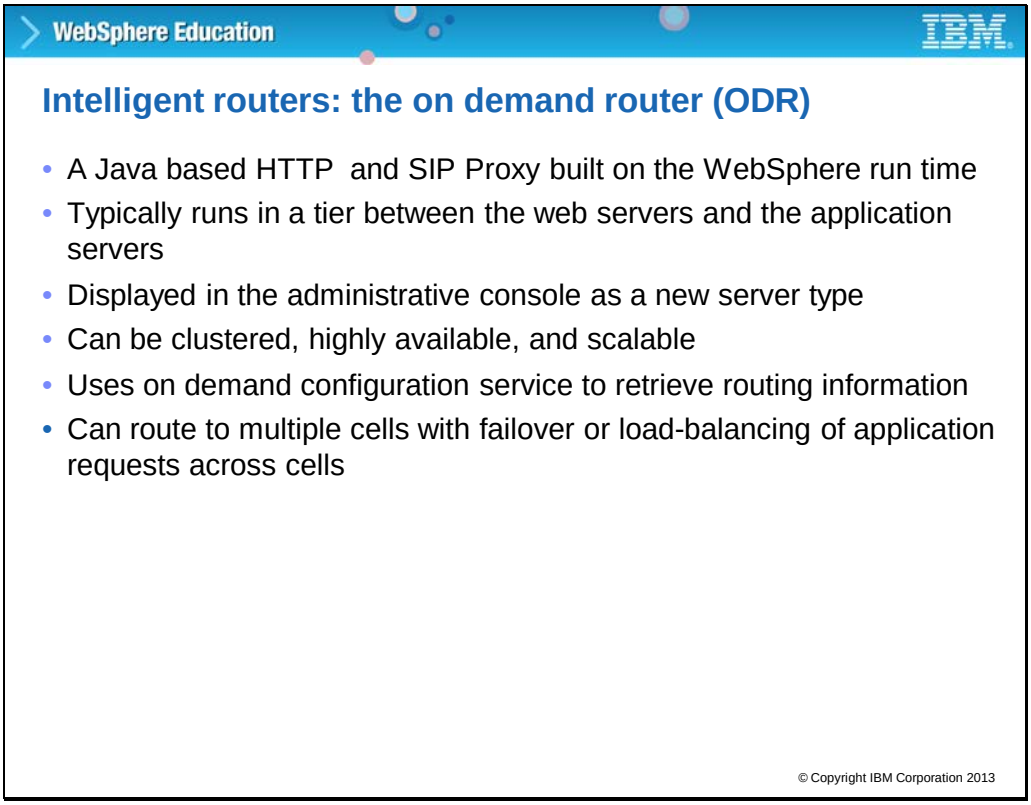
The on-demand configuration service maintains cell topology information and keeps the autonomic request flow manager and other controllers aware of the environment. It tracks updates in cell topology and state, including applications that are installed and removed, servers started and stopped, and others. The on-demand configuration service allows the on-demand

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

router to sense its environment. The on-demand router dynamically configures the routing rules at run time to allow the on-demand router to accurately route traffic to application servers.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

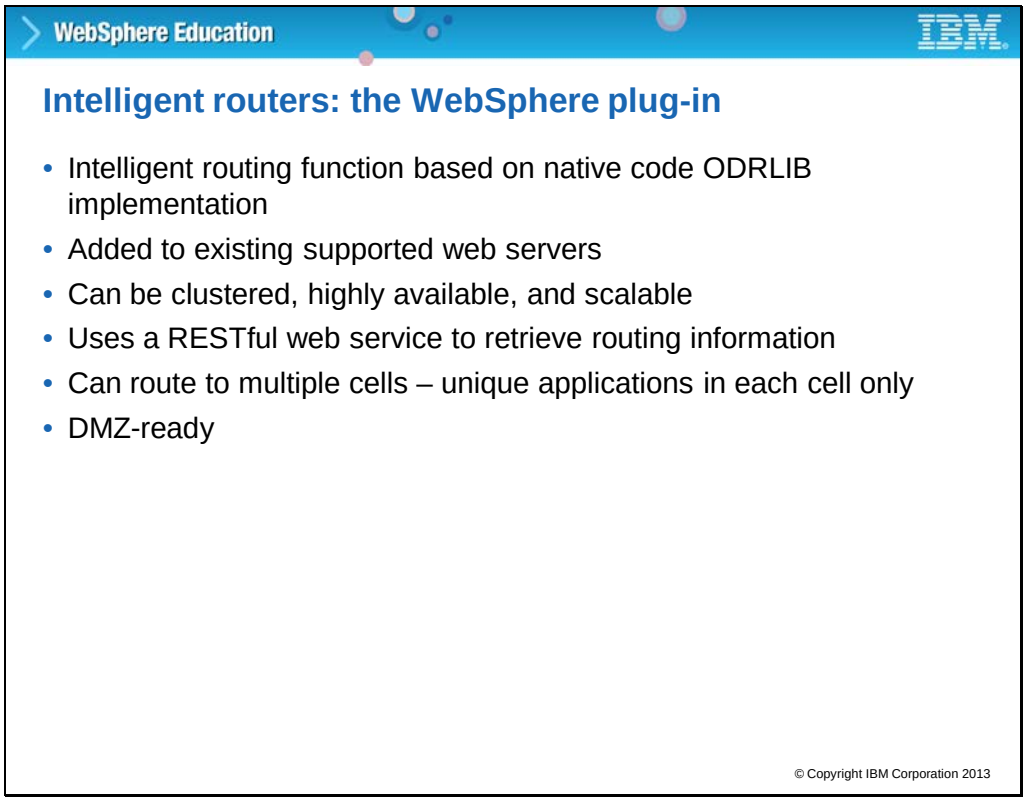
Slide 14



The on-demand router is an HTTP and SIP proxy that is part of a WebSphere Application Server cell. It typically sits between the web servers and the application servers. It is one of two types of intelligent routers that the Intelligent Management feature uses to route client requests to application servers. Because the on-demand router is part of the cell, it can use the on-demand configuration service to retrieve the information that is used to route requests. The on-demand router can route requests to multiple cells.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

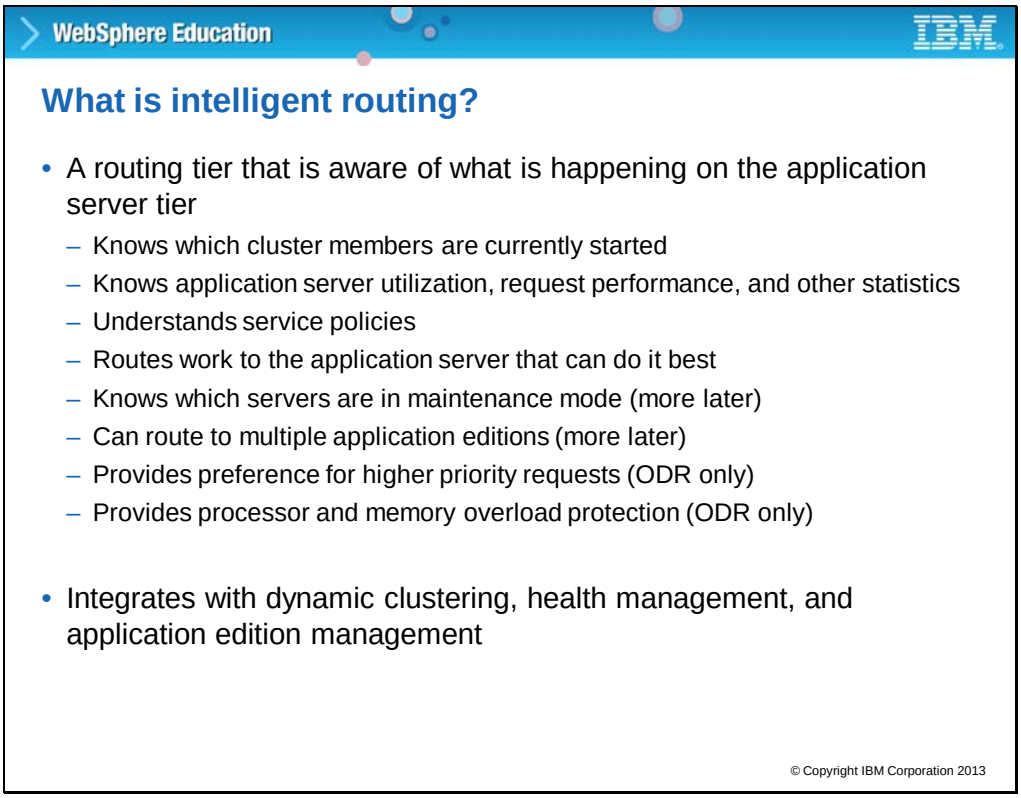
Slide 15



The WebSphere plug-in is the second type of intelligent router that the Intelligent Management feature uses. The plug-in runs in the web server tier, and routes requests directly to application servers. The information that is needed to route requests is retrieved from the cell with a RESTful web service. One advantage of the WebSphere plug-in intelligent router is that it is DMZ-ready. However, the plug-in intelligent router does not provide as much function as the on-demand router.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 16



With intelligent routing, the router is aware of what is happening in the application server tier. As dynamic cluster members are started and stopped, the intelligent router is informed so that the router knows what servers are available. The intelligent routers use performance characteristics of the application servers to make decisions. Service policy information is used along with performance information to route requests most effectively.

Intelligent routers understand when a server or node is in maintenance mode, and route only appropriate requests to those endpoints. Maintenance mode is presented more fully in the health management topic.

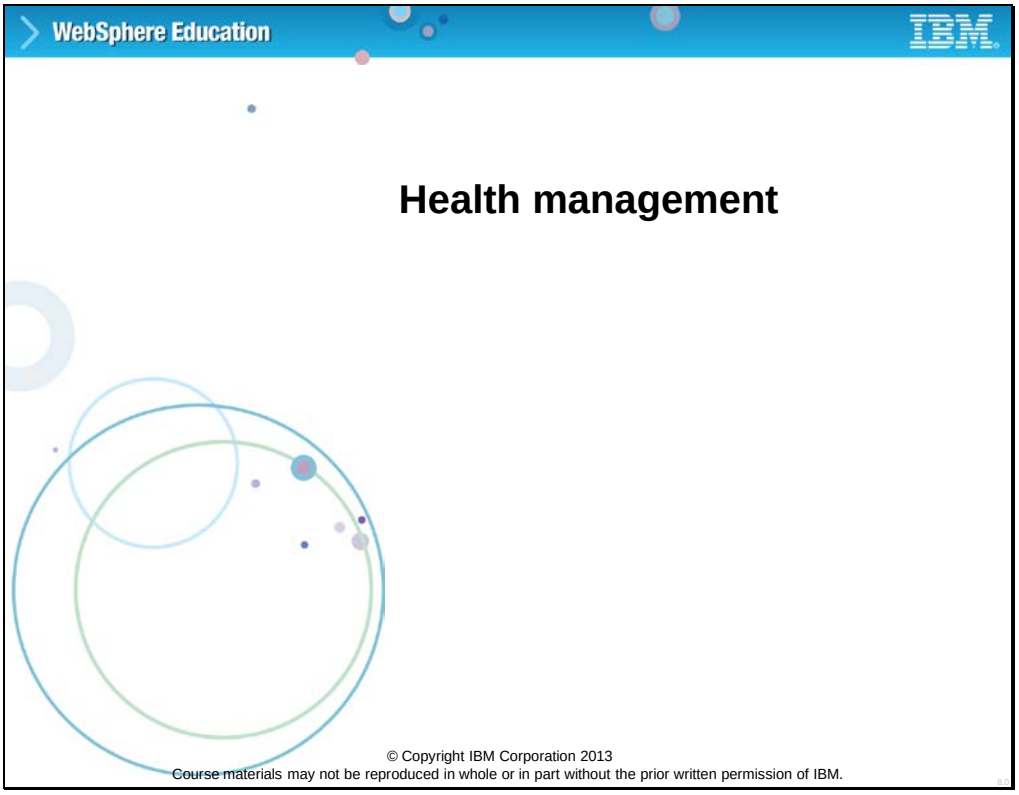
Intelligent routers can route to more than one edition of an application at the same time; or can route correctly when rolling out from one edition to another. Edition management is presented in more detail later in this unit.

The ODR intelligent router can prioritize requests so that lower priority requests are held back, allowing higher priority requests to use more application server resources. The ODR can also queue requests to prevent application server overload.

Intelligent routers work with autonomic managers, dynamic clusters, the health management system, and the application edition management system to route requests to the most appropriate resources.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

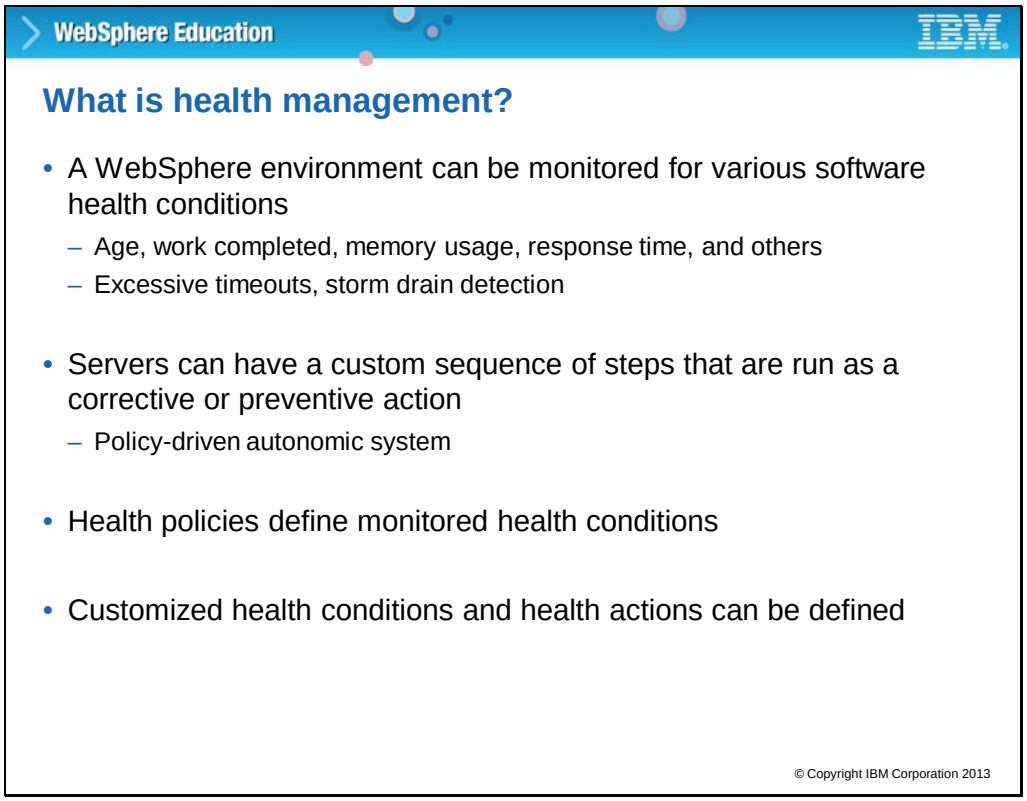
Slide 17



Topic: Health management. In this topic, you learn about the health management feature.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

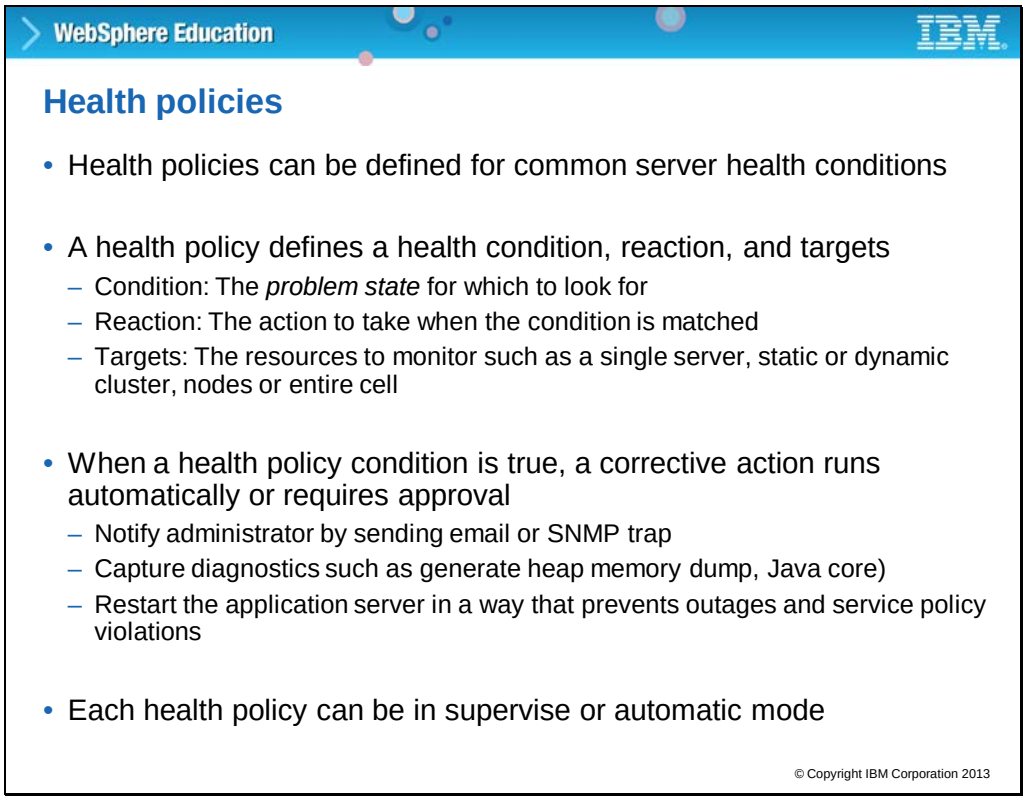
Slide 18



Intelligent Management provides a health management feature to monitor the status of your application servers, and sense and respond to problem areas before an outage occurs. You can manage the health of your application that is serving the environment with a policy-driven approach that enables specific actions to occur when monitored criteria are met. For example, when memory usage exceeds a percentage of the heap size for a specified time, health actions can run to correct the situation.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 19

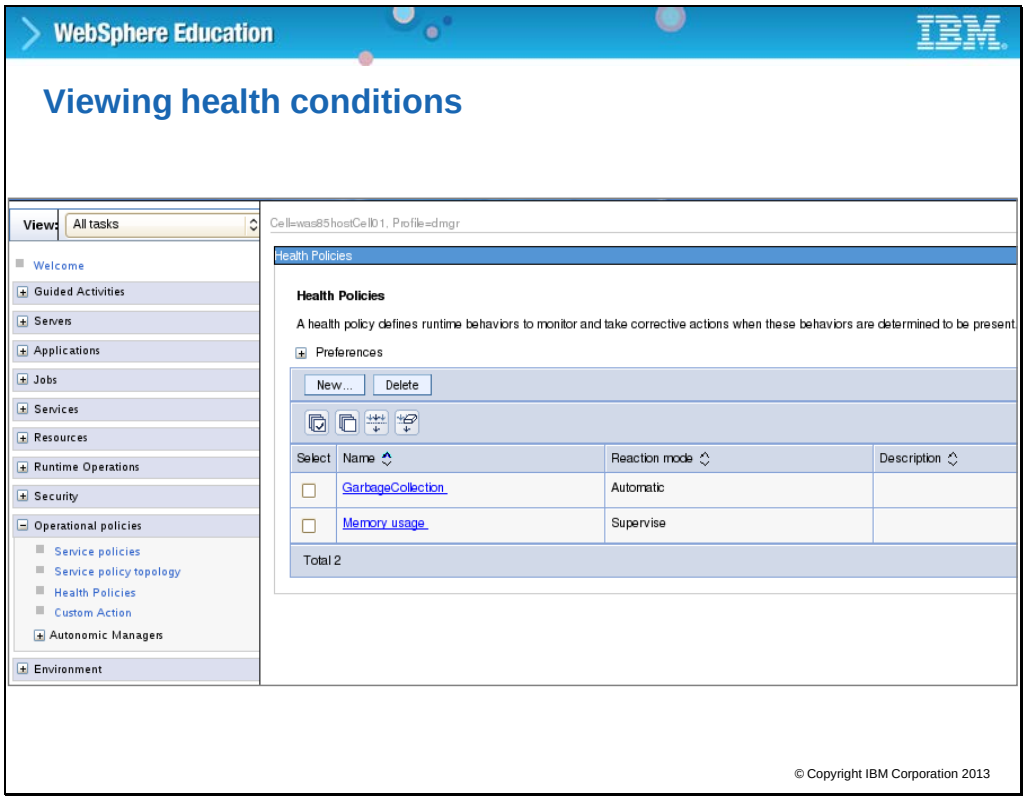


WebSphere can monitor servers for common health problems and take corrective action. Various health conditions can be defined by using health policies. When a health policies violation is detected, an action plan can be put into effect automatically. The plan can include to notify administrators (including by email), capture diagnostics information (Java thread or heap dump), or restart the server.

Application server restarts are intelligent and done in a way to prevent outage and service policy violations. WebSphere provides the “first line of defense” for poor application health by mitigating common health problems and routing around unhealthy servers. ITCAM for WebSphere extends WebSphere health management by adding in-depth application problem determination capabilities. ITCAM gives support teams the diagnostics tools they use find out at a granular level what went wrong and how to fix it fast. ITCAM provisions Rational and Eclipse developer and test tools with performance data captured in production, eliminating the need to attempt problem recreation. ITCAM integrates with the broader Tivoli Automation portfolio that enables customers to cost-effectively manage their IT infrastructure.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 20

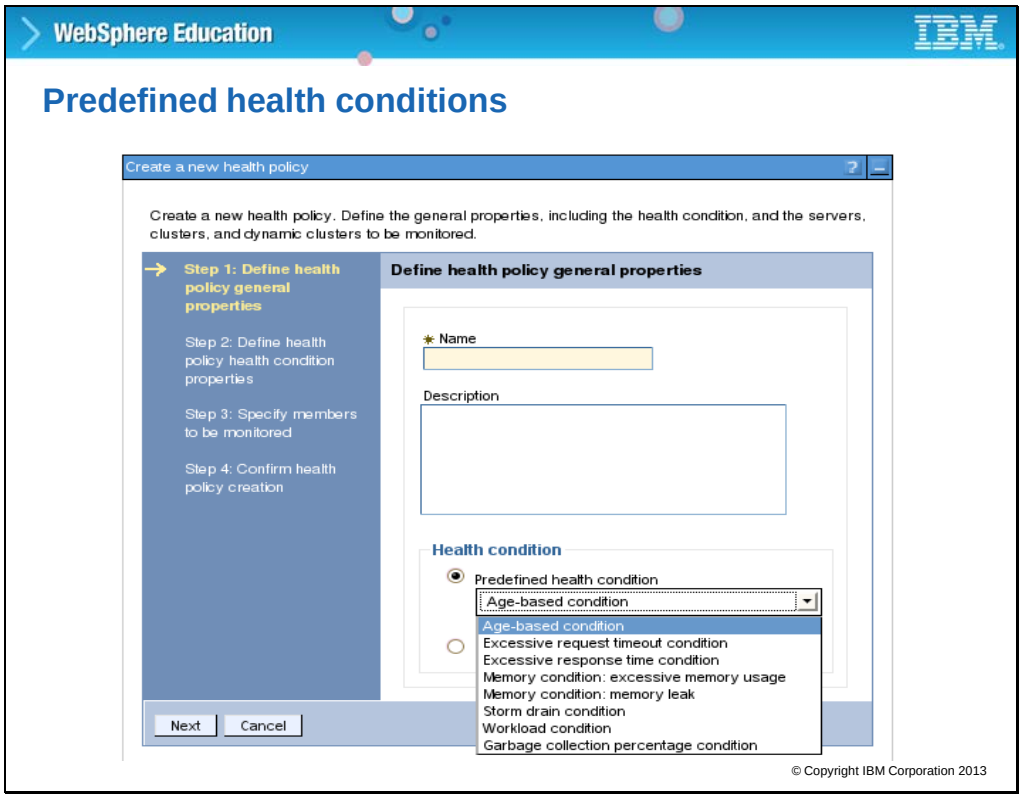


You can view existing and create new health policies by using the administrative console.

Navigate to **Operational policies > Health Policies** to view any defined policies.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

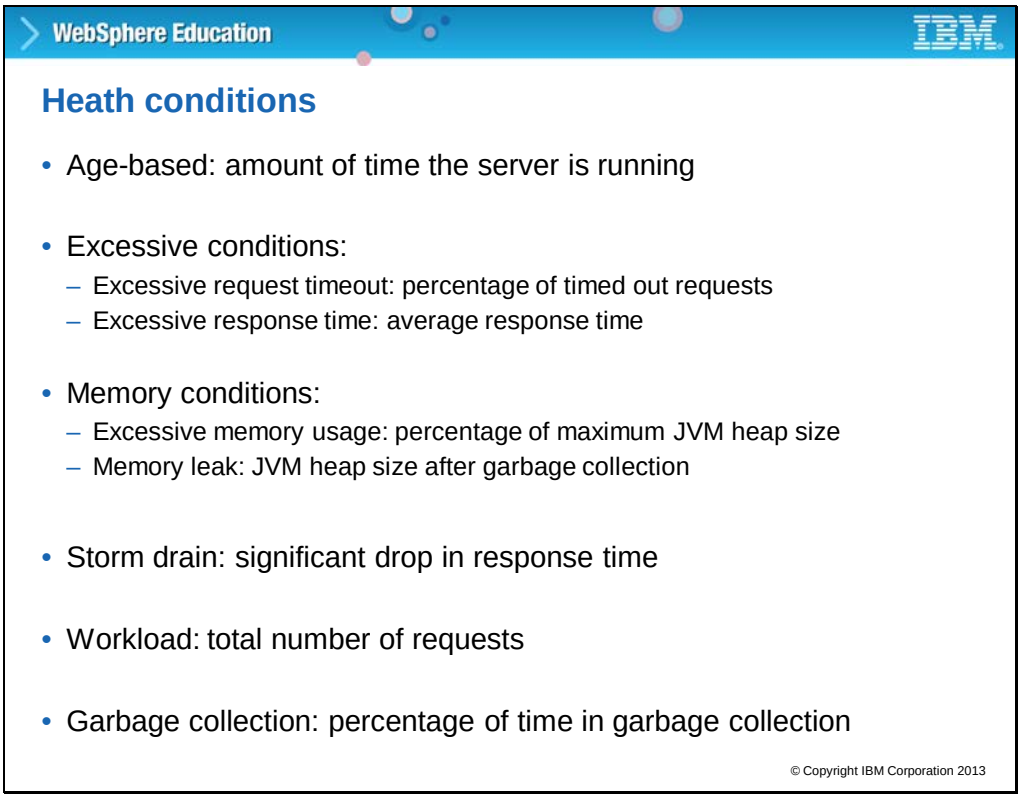
Slide 21



Health conditions define the variables that you want to monitor in your environment. Several categories of health policy conditions exist. You can choose from the list of predefined health conditions when you create a health policy

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 22



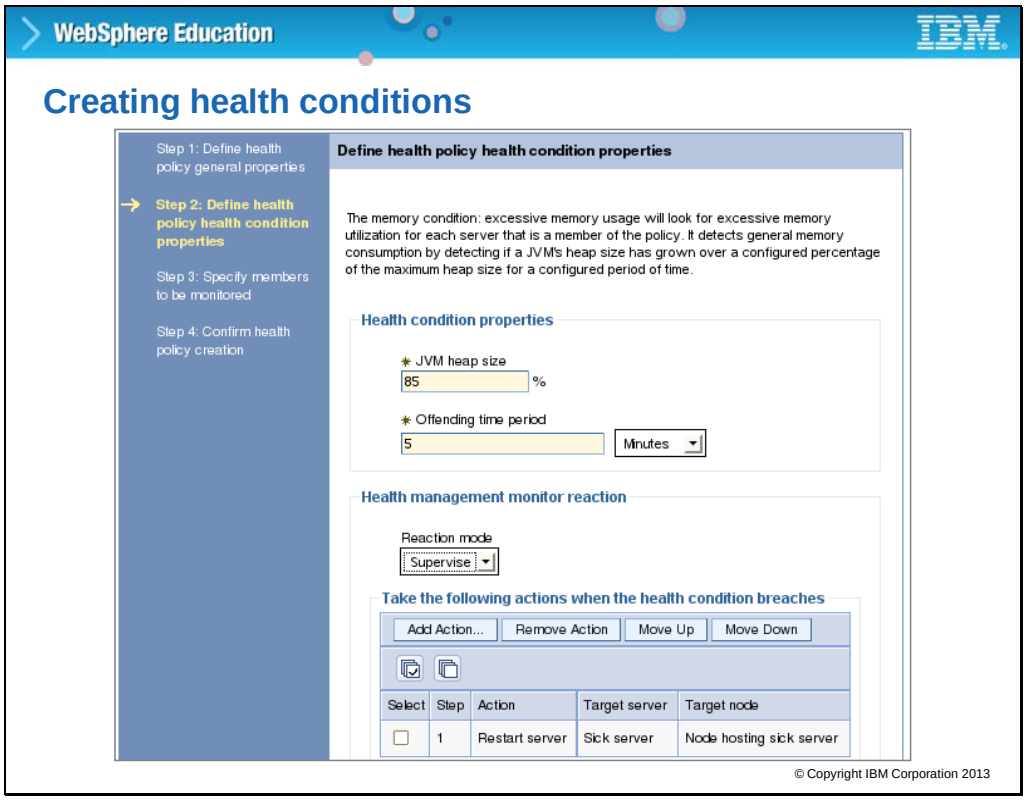
This slide introduces the pre-defined health conditions. Excessive garbage collection triggers when the Java virtual machine (JVM) spends more than a configured percentage of time when running garbage collections.

Storm drain detects situations where requests are shifted toward a faulty cluster member that advertises low response times. This condition is triggered when there is a significant drop in the average response time. This drop must be measured at the on-demand router, for a member of the cluster that is coupled with an increase in the dynamic weights for the cluster member.

Workload triggers when the members that are associated with this policy serve a user-defined number of requests. You can use the workload condition on all server types.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

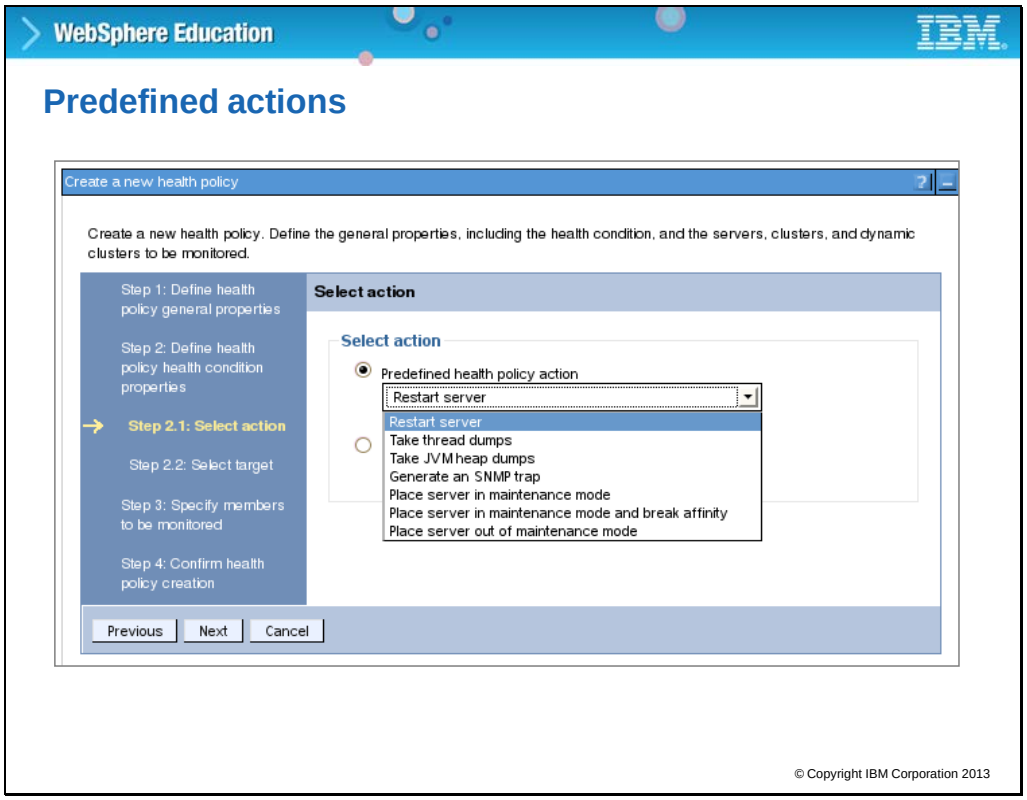
Slide 23



Actions can be taken automatically, or you can have them occur in supervised mode. Supervised mode requires an operator to confirm the action.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

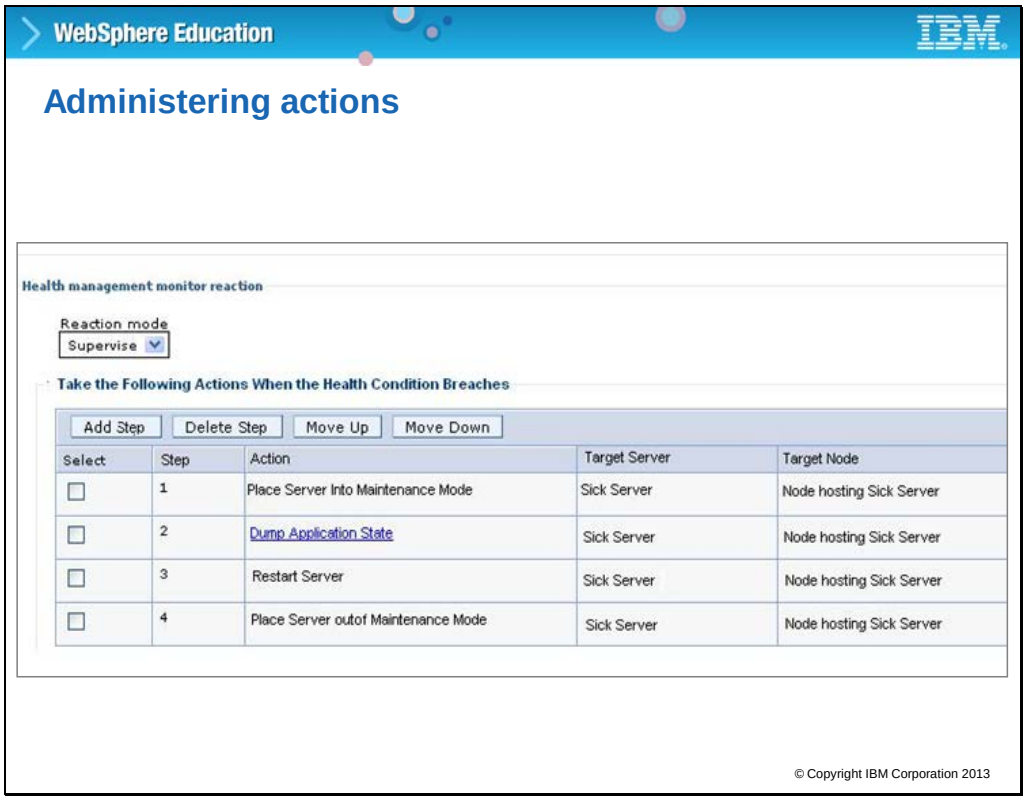
Slide 24



When a health policy violation is detected, an action plan can be put into effect automatically. Actions that are taken when a monitored condition is detected, are designed to bypass the problem and help in diagnosis. You can select from various predefined actions that are displayed on the screen capture.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

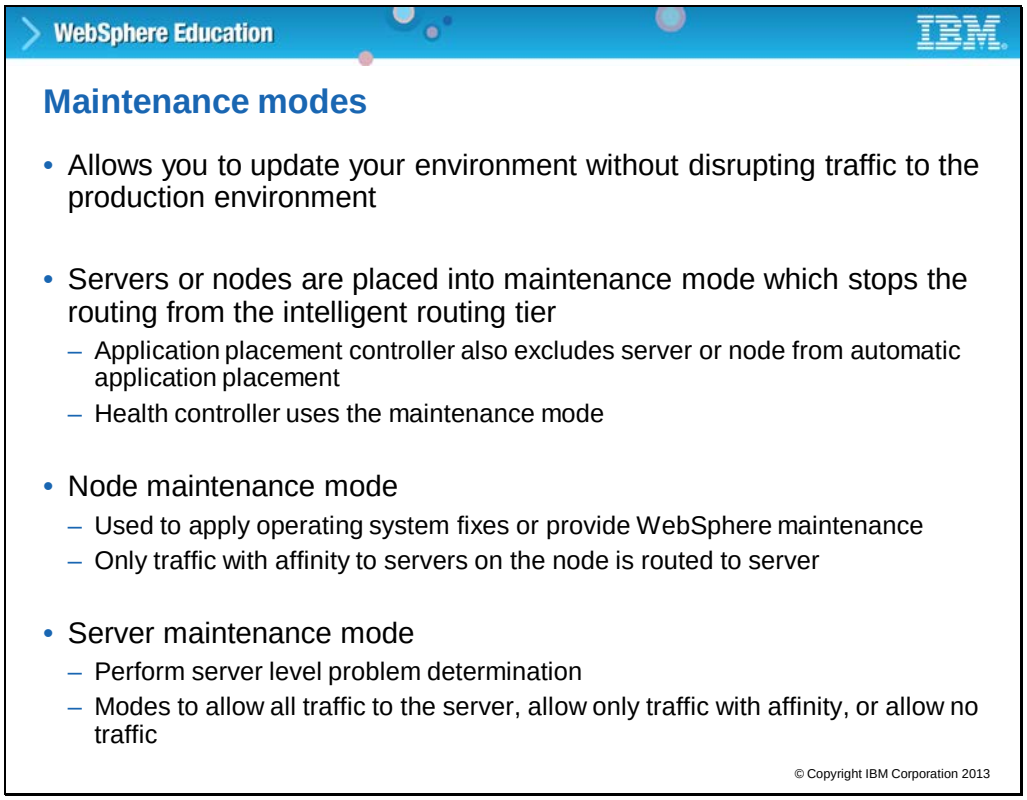
Slide 25



If you chose the **Supervise** reaction mode, then you receive recommendations to improve your health conditions. These recommendations display as runtime tasks that you can accept, deny, or close. To manage runtime tasks, click **System administration > Task management > Runtime tasks** in the administrative console. If you chose the **Automatic** reaction mode, actions to improve the health of your environment occur automatically.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 26



Periodic product maintenance is important to keep your system environment working correctly, and to avoid trouble from known issues. At some point in time, you might have a problem with a server and must complete diagnostic tests to troubleshoot a specific application server. These situations can lead to the disruption of client requests to servers in your environment.

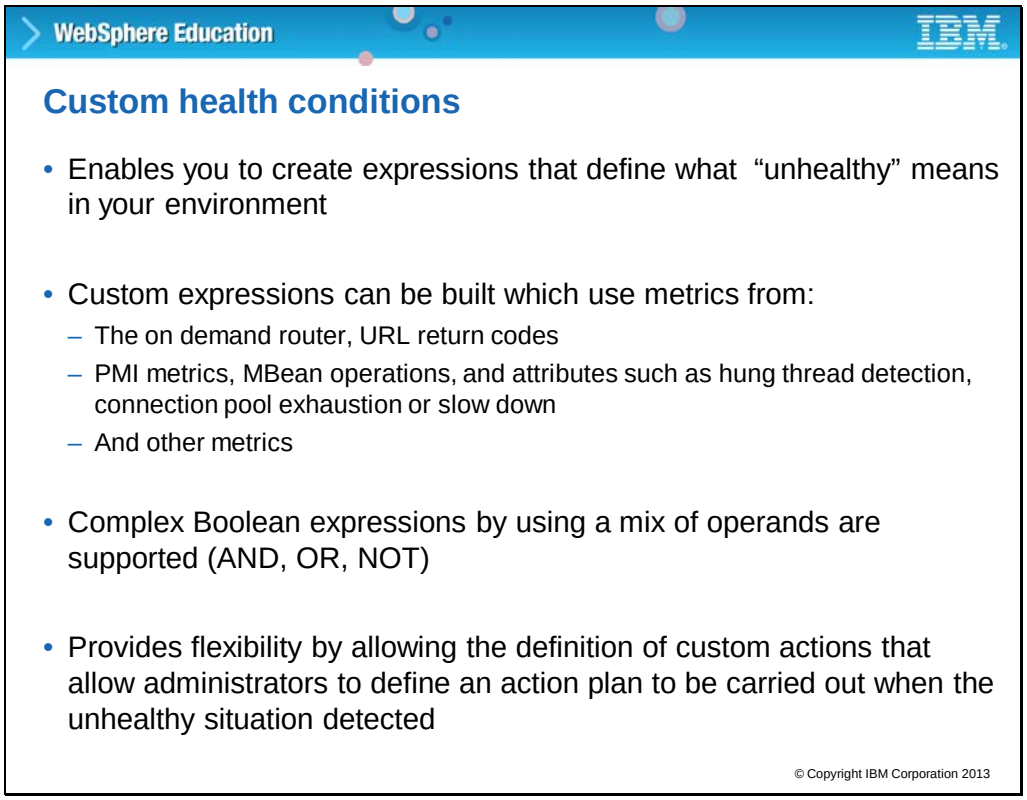
Using the Intelligent Management feature, you can maintain the environment without disrupting traffic to the production environment. You can use it to administratively put a server or node in the cell into maintenance mode.

You can put a node into maintenance mode when you must apply operating system fixes or complete WebSphere maintenance. When a node is in maintenance mode, only traffic with affinity to servers on the node is routed to the server by an intelligent router. A maintenance immediate stop mode can be set that immediately stops the servers on the node.

You can put a server into maintenance mode when you must complete server level problem determination. When an application server is placed into maintenance mode, you can indicate to allow all traffic to the server, allow only traffic with affinity, or allow no traffic during the maintenance period.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

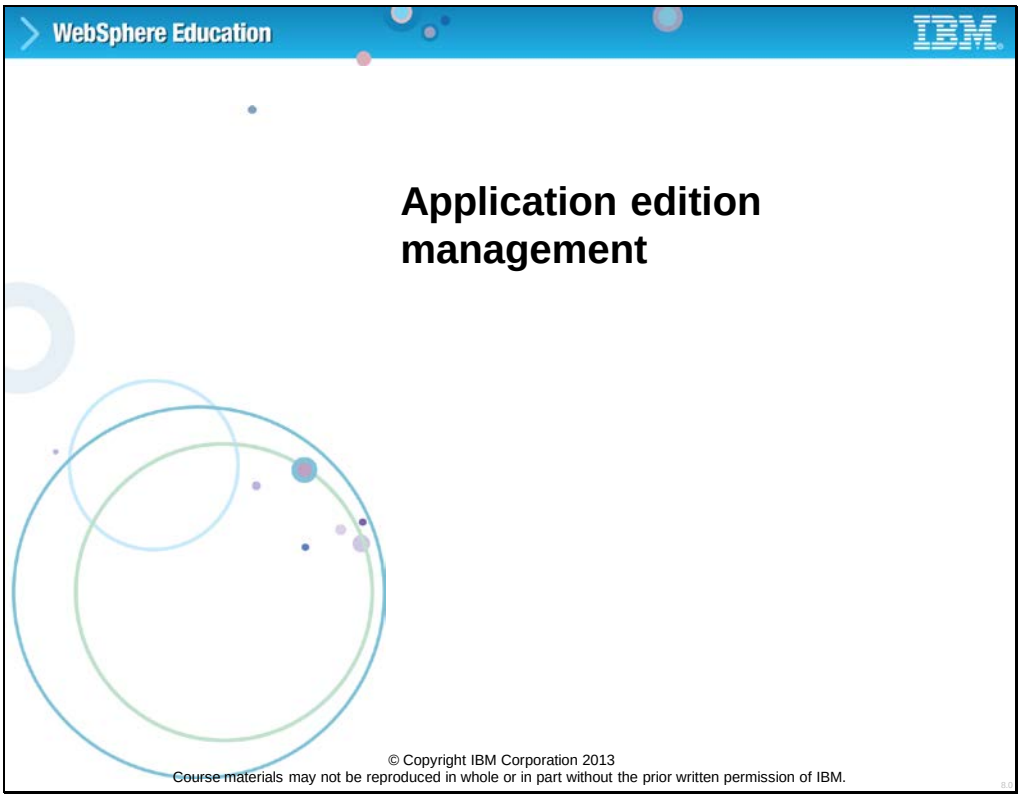
Slide 27



You can define custom conditions for your health policy if the predefined health conditions do not fit your needs. You define custom conditions as a sub expression that is tested against metrics in your environment. When you define a custom condition, consider the cost of collecting the data, analyzing the data, and if needed, enforcing the health policy. This cost can increase depending on the amount of traffic and the number of servers in your network. Analyze the performance of your custom health conditions before you use them in production.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

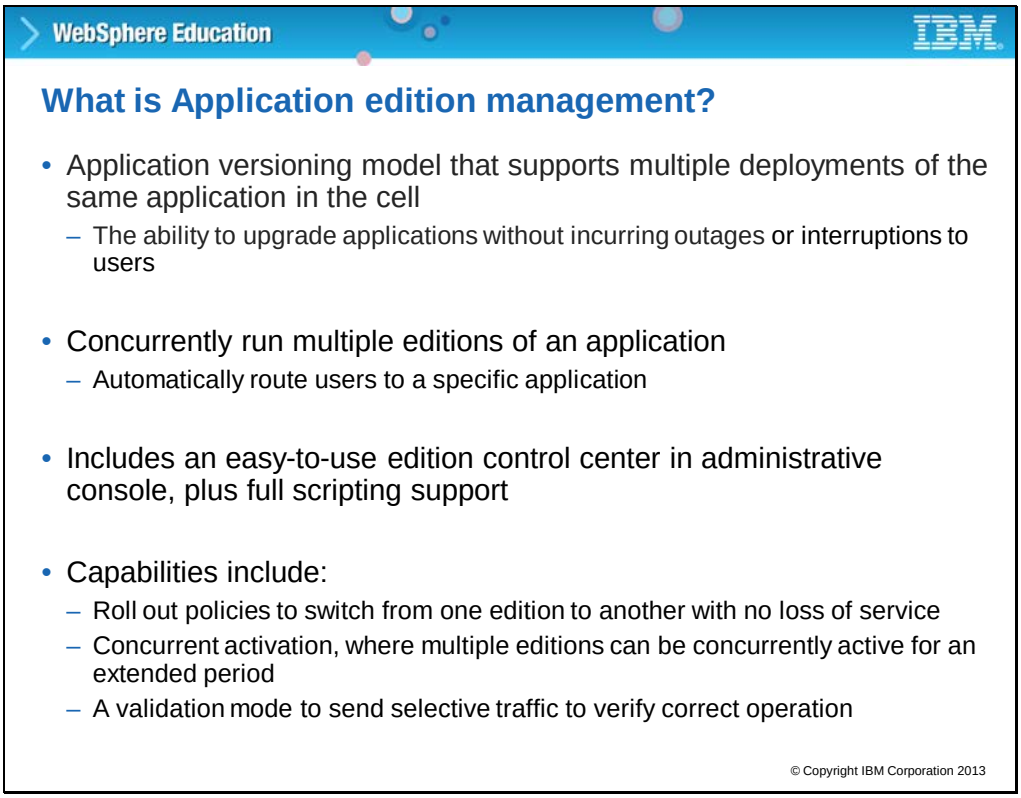
Slide 28



Topic: Application edition management. In this topic, you learn how to manage various application editions.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

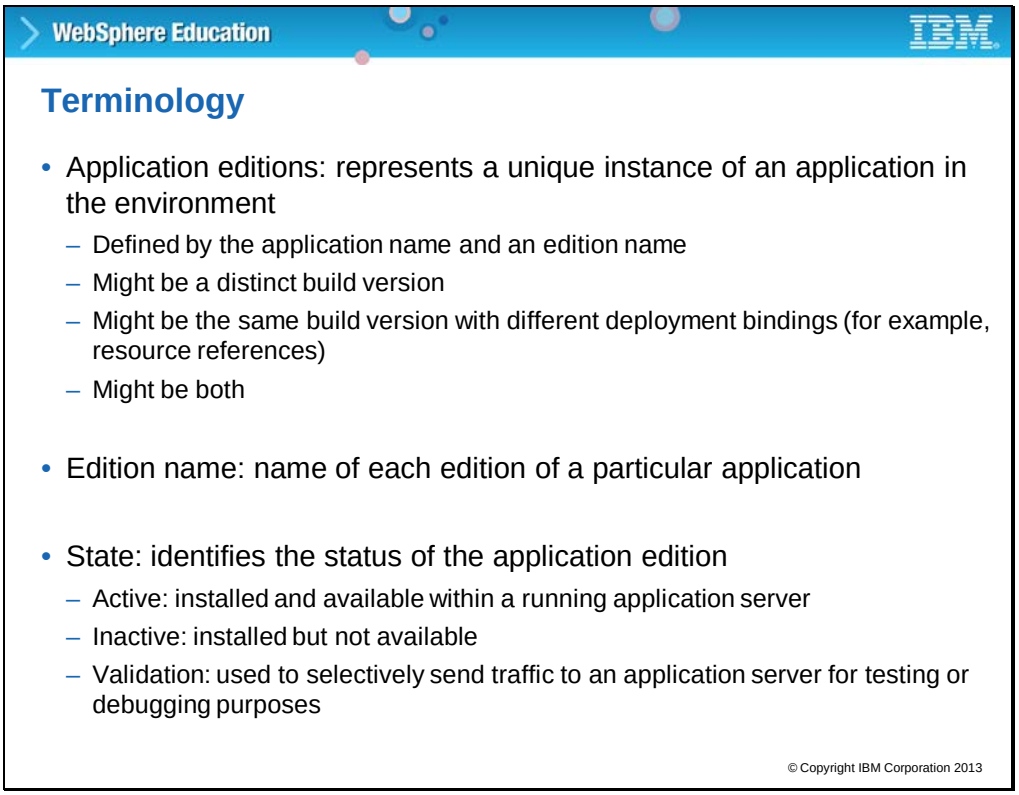
Slide 29



Application edition management enables management of interruption-free production application deployments. Using this feature, you can validate a new edition of an application in your production environment without affecting users, and upgrade your applications without incurring user outages. You can also run multiple editions of a single application concurrently, directing different users to different editions, as the ODR maintains not only traditional application state (for example, HTTP session) affinity, but also application version affinity. The ability to queue requests is also employed with the Intelligent Management application edition function. This ability is useful if an “atomic” application update that allows pre-provisioning of a new application version, and an “atomic” update of all users from the old application version to the new application version is wanted.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 30



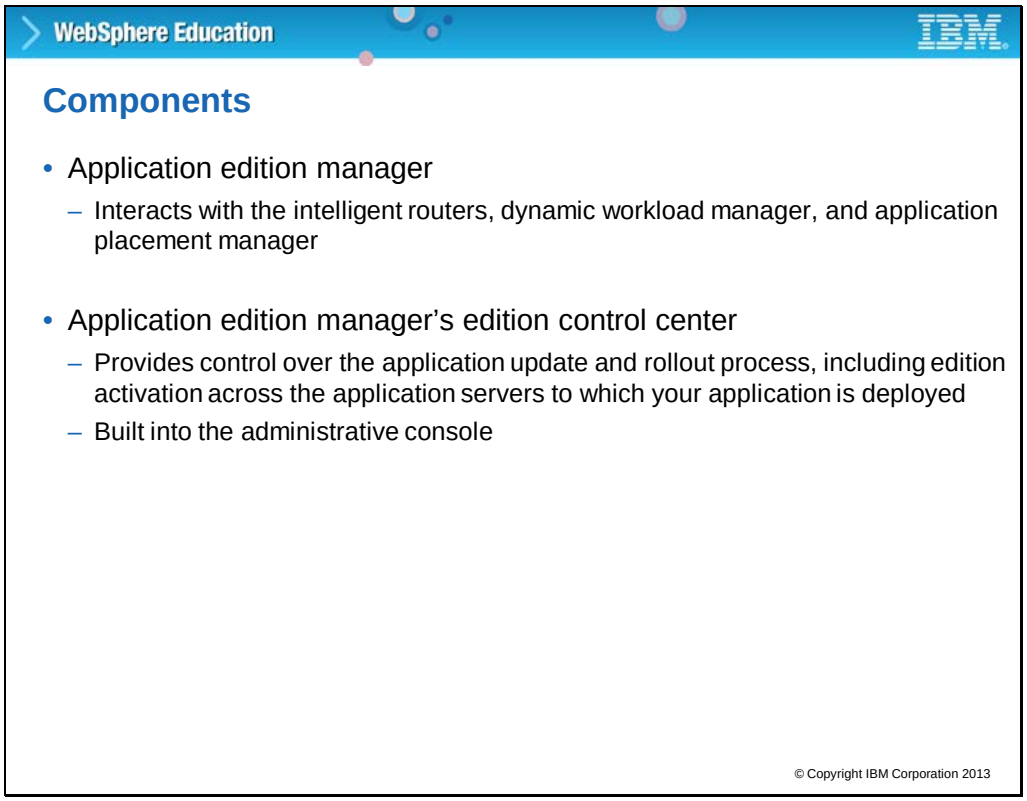
An application edition represents a unique instance of an application in the environment. An application edition encompasses both application versions and deployment bindings. The combination of an application name and an edition name uniquely identifies an edition of an application.

With application edition manager, you can install multiple editions of the same application. Each edition is identified with an application edition name and description. The edition name is a field in which you can specify a value to uniquely identify one application edition from other editions of the same application. Create a version number scheme for naming editions that is meaningful in your environment. Multiple editions of the same application have the same application name but different edition names. When deploying an application, you can also specify an edition description next to the edition name, which gives you the ability to store more information.

The existing application installation and update functions in Network Deployment are destructive. That is, they replace the old instance of the application with a new instance. Installing an application edition is non-destructive. You can install any number of application editions and keep them in the system management repository.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 31



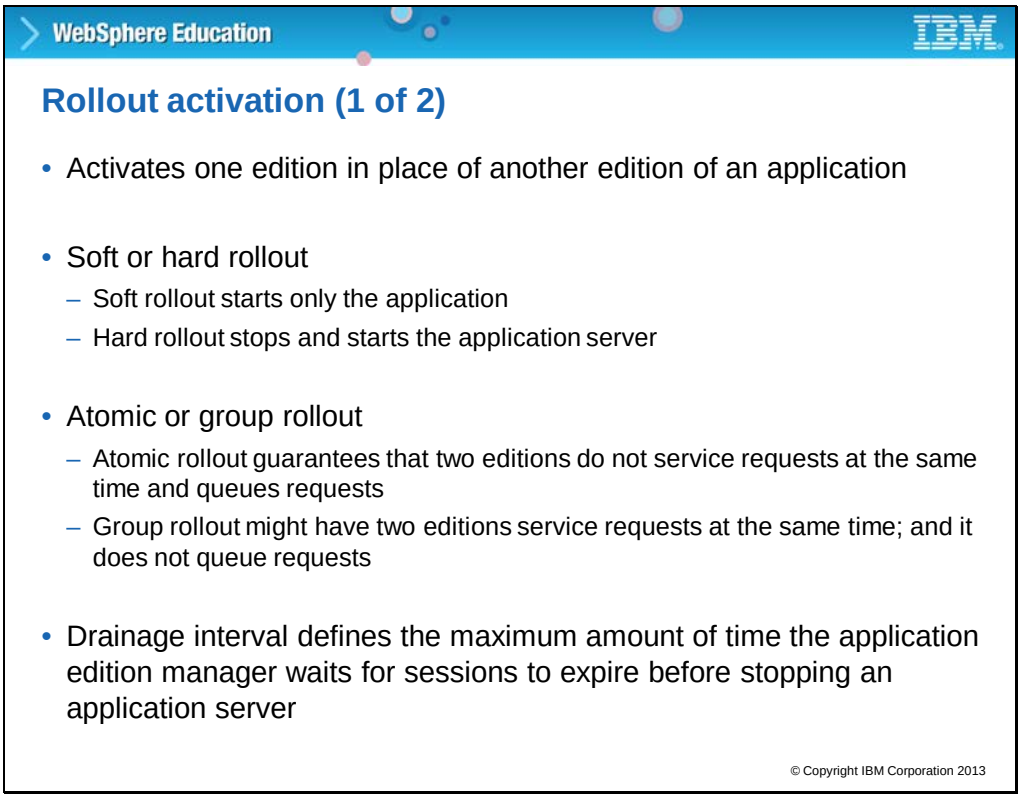
The application edition manager ensures interruption-free production application deployments. Interruption-free deployment prevents loss of service when you install an application update in your environment.

The application edition manager provides an application model for defining versions that supports multiple deployments of the same application in the Intelligent Management cell. Each deployment has a unique edition name. The application edition manager allows you to select the edition to activate on an Intelligent Management cluster so that you can roll out an application update or revert to a previous level.

The application edition manager is fully integrated with Intelligent Management, interacting with the on-demand router (ODR), dynamic workload balancing, and the application placement manager. This integration ensures predictable application behavior when you apply application updates, and a smooth transition from one application edition to another while the system continues to manage your application performance goals. You can access application update processes with the administrative console, including edition activation across the application servers.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

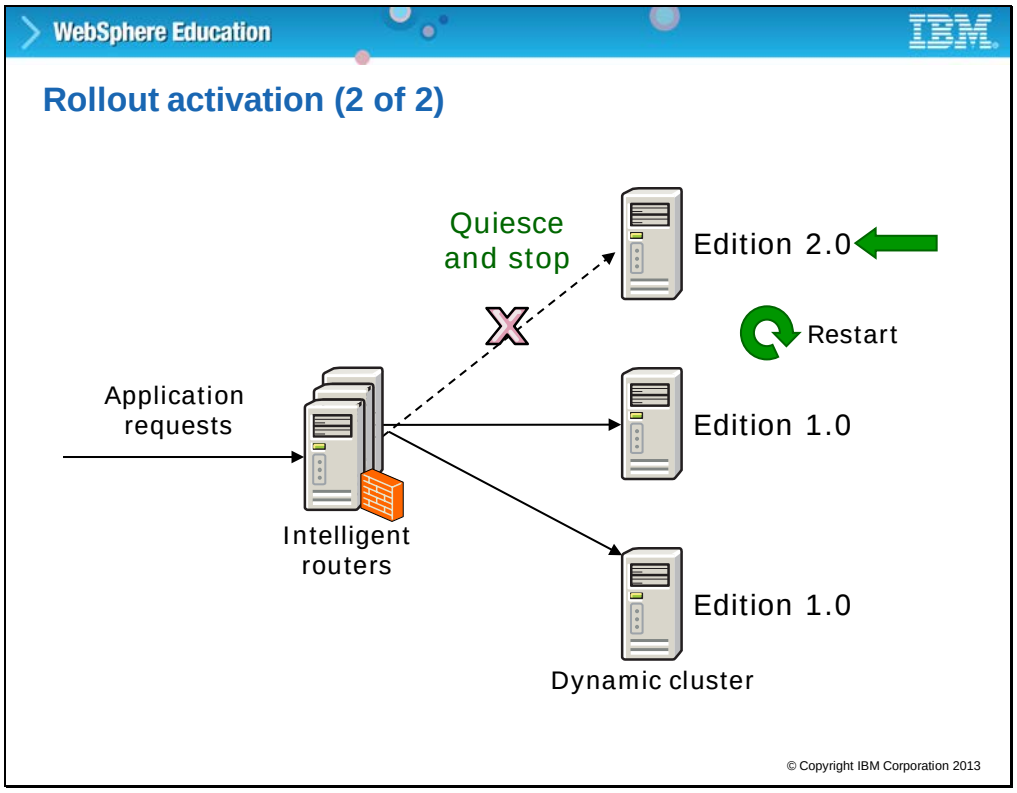
Slide 32



Rollout activation activates one edition in place of another, ensuring an interruption-free update in the process. Thus, all application requests are serviced during the rollout and none are lost. This process ensures continuous application operation from the perspective of the customers of that application. The application edition manager carefully coordinates the activation of the edition and the routing of requests to the application.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 33



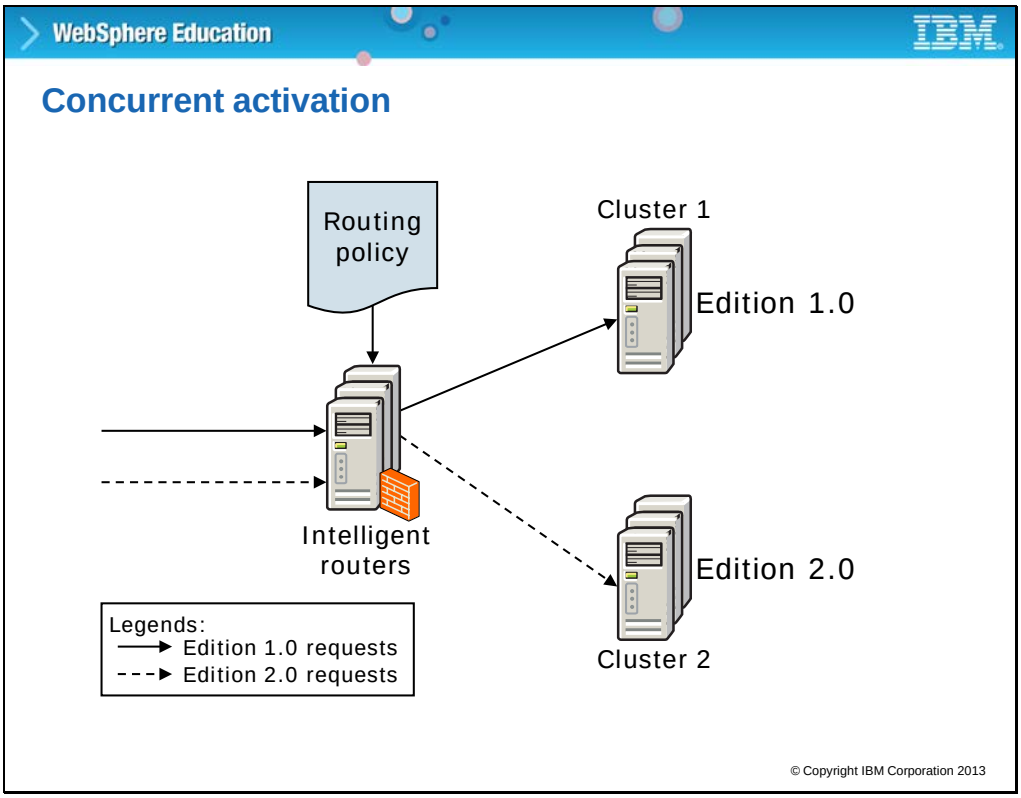
Replacement of one edition with another, in a production environment, requires certain discipline in the evolution of the application. Because edition replacement happens while application users are potentially accessing the previous application edition, the new edition must be compatible with earlier versions. Thus, the new edition cannot add or change any existing application interfaces, including essential behavior. New interfaces can be added. In addition, existing interfaces can be algorithmically corrected and, in some cases, even extended and remain compatible with existing application users.

This slide displays an example of a group rollout scenario. In the diagram, a dynamic cluster is created that consists of three servers. You first must divide the cluster into groups, which tells the application edition manager how many servers to update at the same time. Issuing a rollout to a group, results in the servers in each group that are being upgraded to the new edition at the same time. Each server in the group is quiesced, stopped, and reset.

As the rollout is run in the diagram on the slide, one server in the cluster is moved from Edition 1.0 to Edition 2.0. During this time, the server does not receive user requests that are directed from an intelligent router, and the server is stopped. All application requests are sent to the servers that are running Edition 1.0. After a server that is running Edition 2.0 is available, application requests are directed to that server. Any servers that are still running Edition 1.0 do not serve requests until the edition is updated to Edition 2.0.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

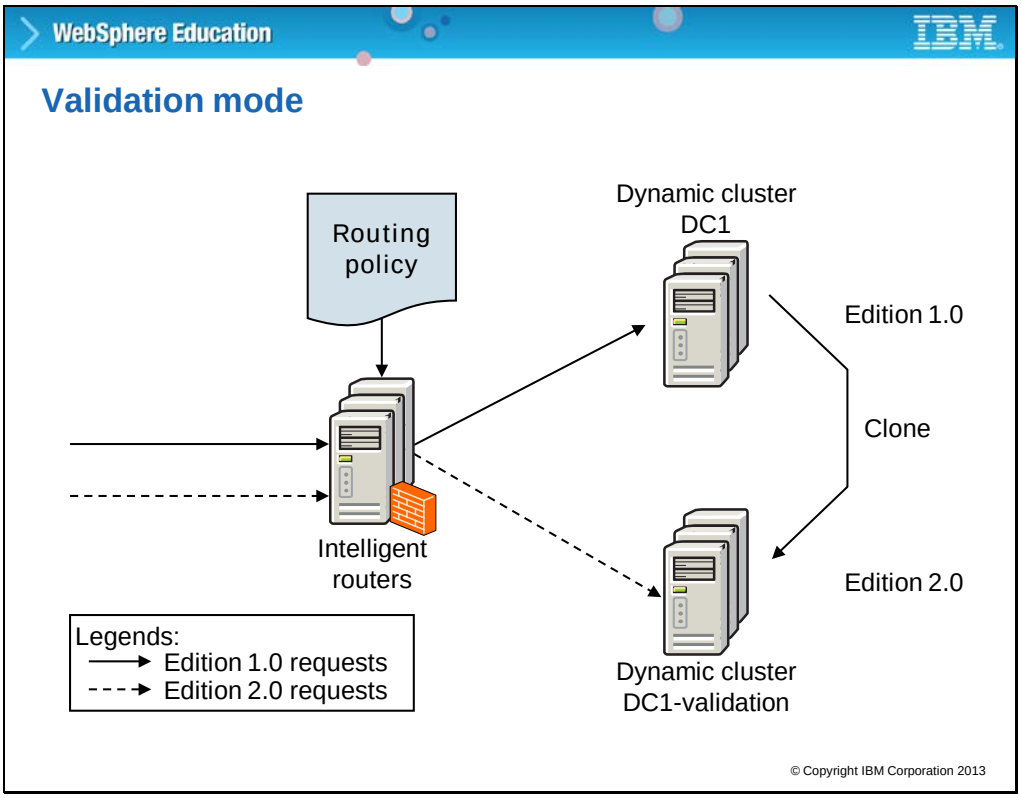
Slide 34



Concurrent activation enables you to activate the same edition on different servers or clusters. To use multiple editions concurrently, you must distinguish user requests from one another so that the requests are sent to the application server that hosts the appropriate edition. For example, if you introduce a new edition of an application, you might want only a select group of users to test the edition.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

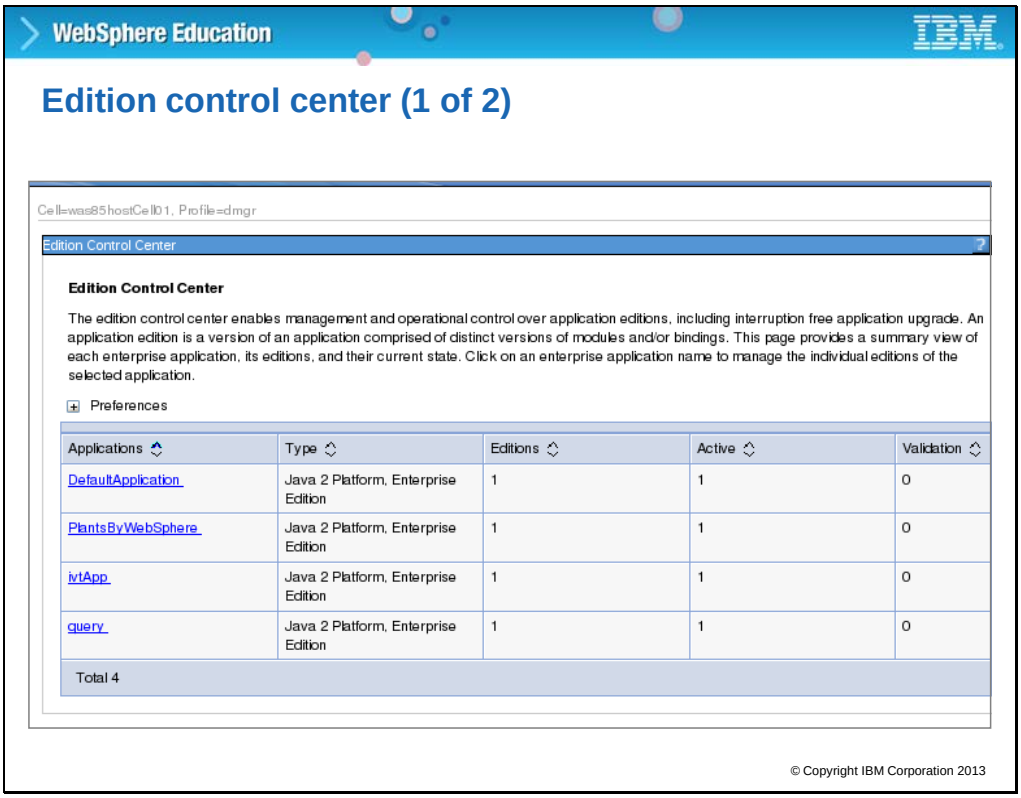
Slide 35



Validation activation is a special form of concurrent activation. It activates an edition on a clone of its original deployment target. The clone is created on activation of the edition. After the validation rollout to the original deployment target, the clone is removed automatically. This action allows you to complete final pre-production testing of an application edition in the actual production environment with a selected set of users.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

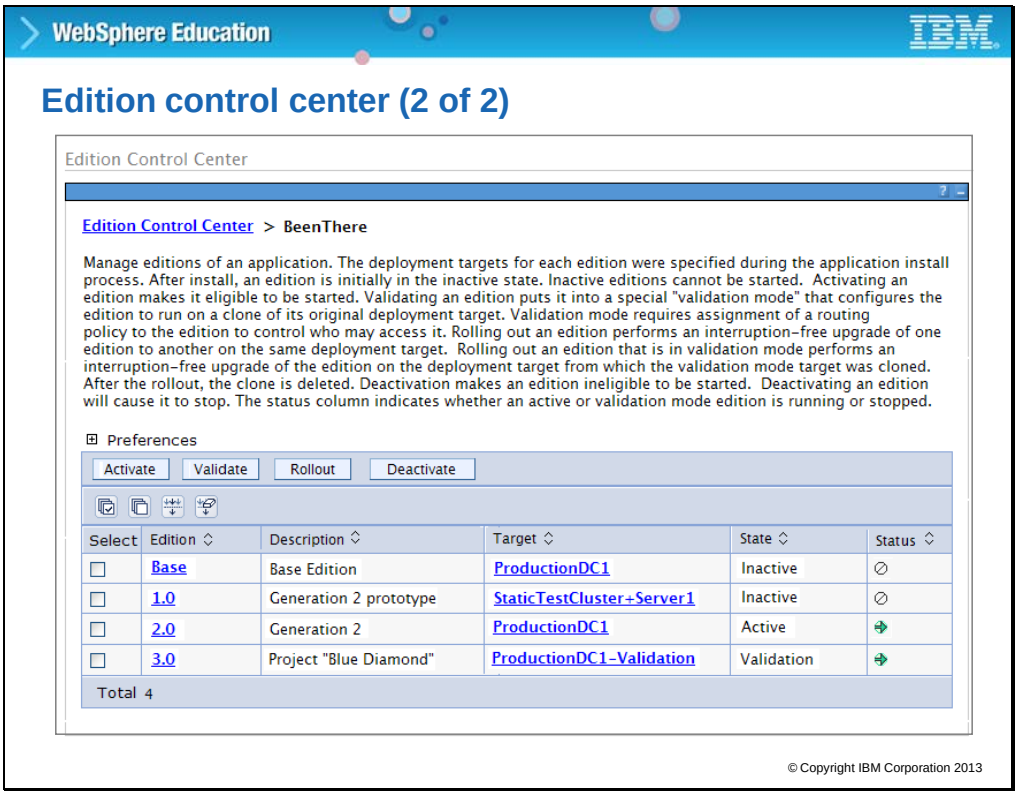
Slide 36



To validate the results, select **Applications** > **Edition Control Center** > ***application_name***.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

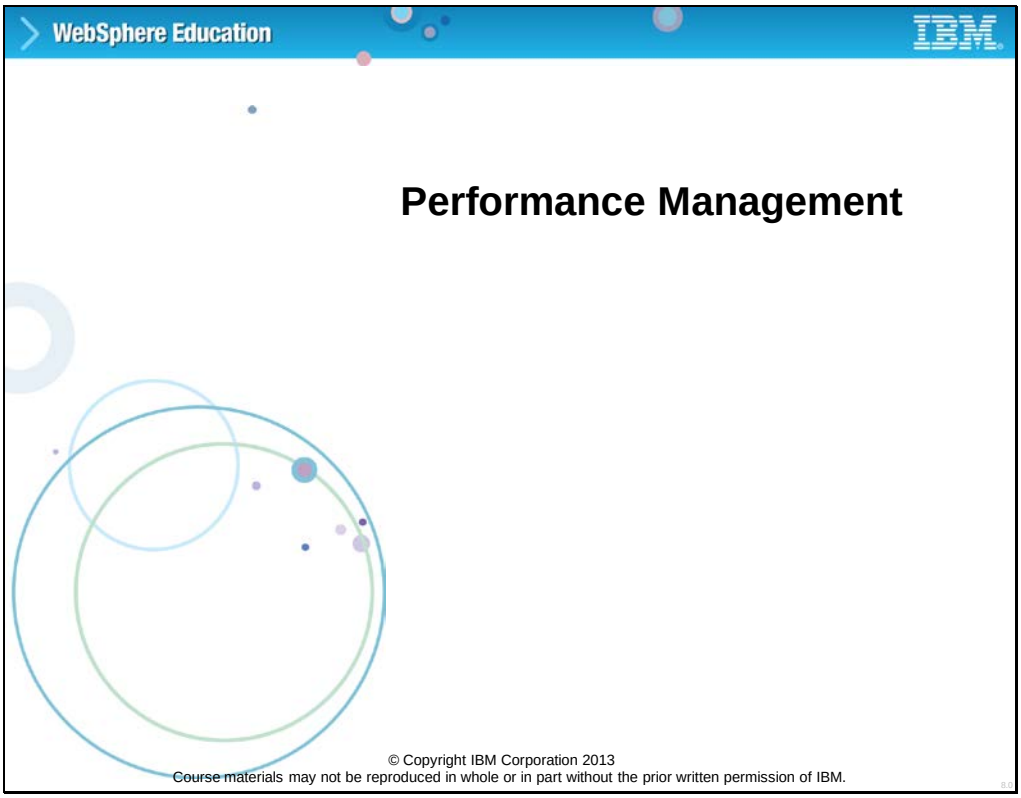
Slide 37



This slide displays editions for the BeenThere application.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

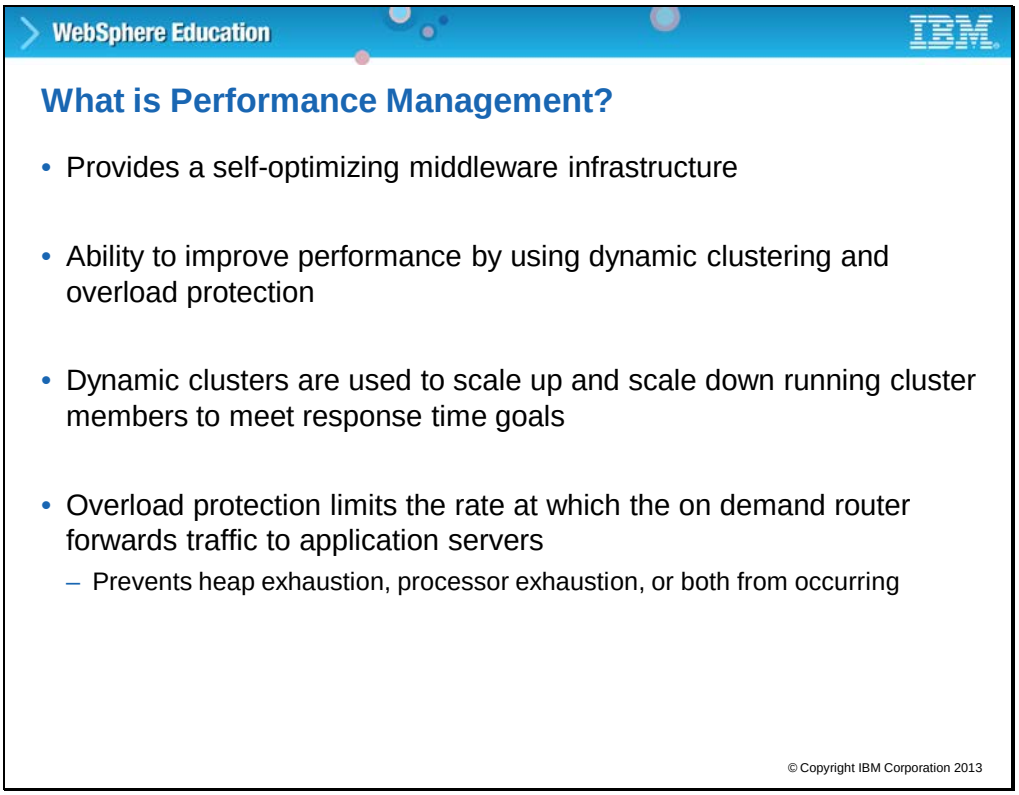
Slide 38



Topic: Performance Management. This topic introduces the concept of Performance Management.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

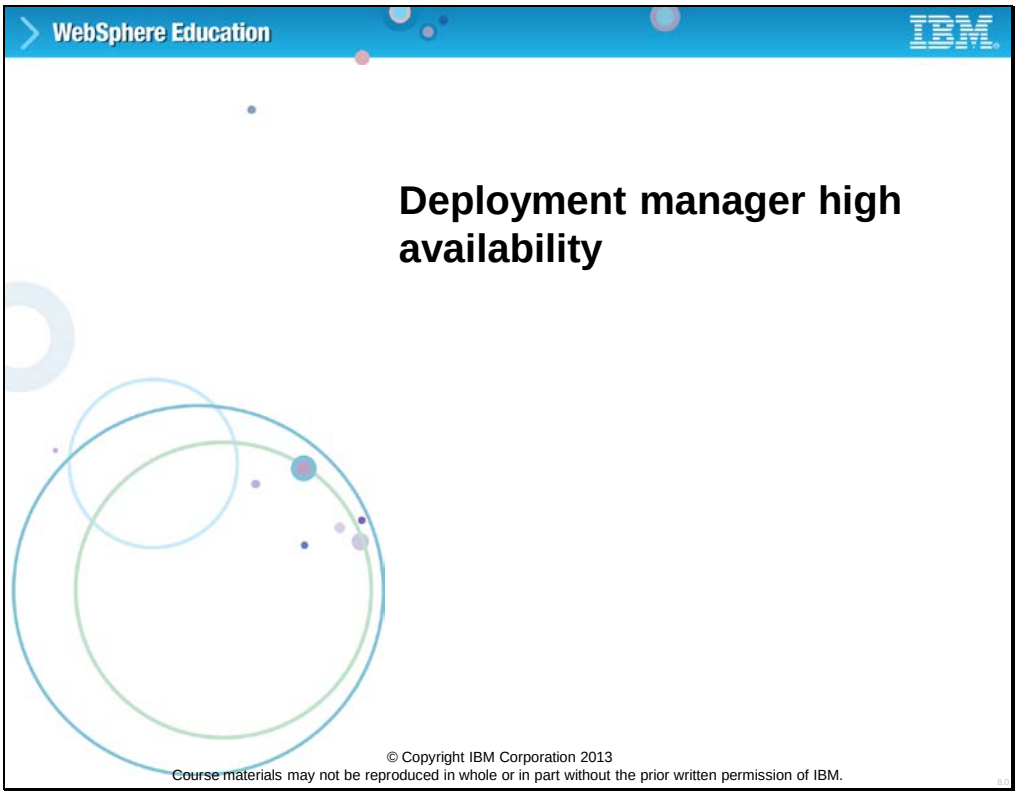
Slide 39



The Performance Management feature provides dynamic cluster capabilities and overload control. With dynamic clusters, you can automatically scale up and down the number of running cluster members as needed to meet response time goals for your users. You can use overload protection to limit the rate at which the on-demand router forwards traffic to application servers. Doing so helps prevent heap exhaustion, processor exhaustion, or both from occurring.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

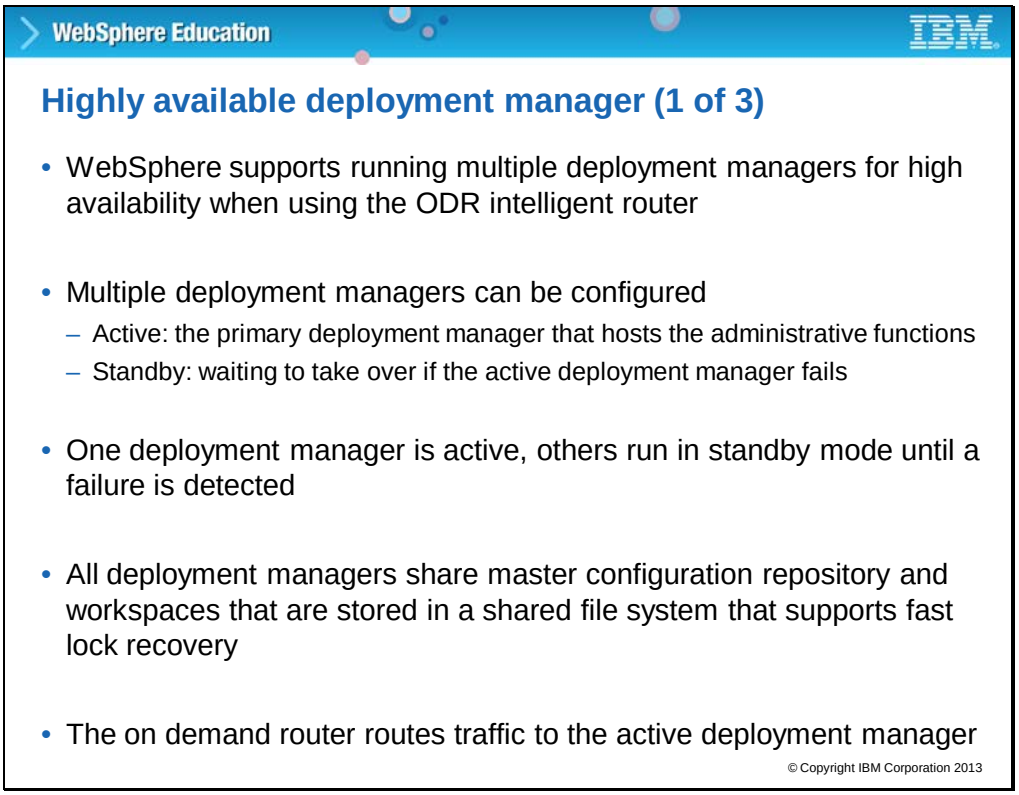
Slide 40



Topic: Deployment manager high availability. This topic introduces the concept of a highly available deployment manager.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

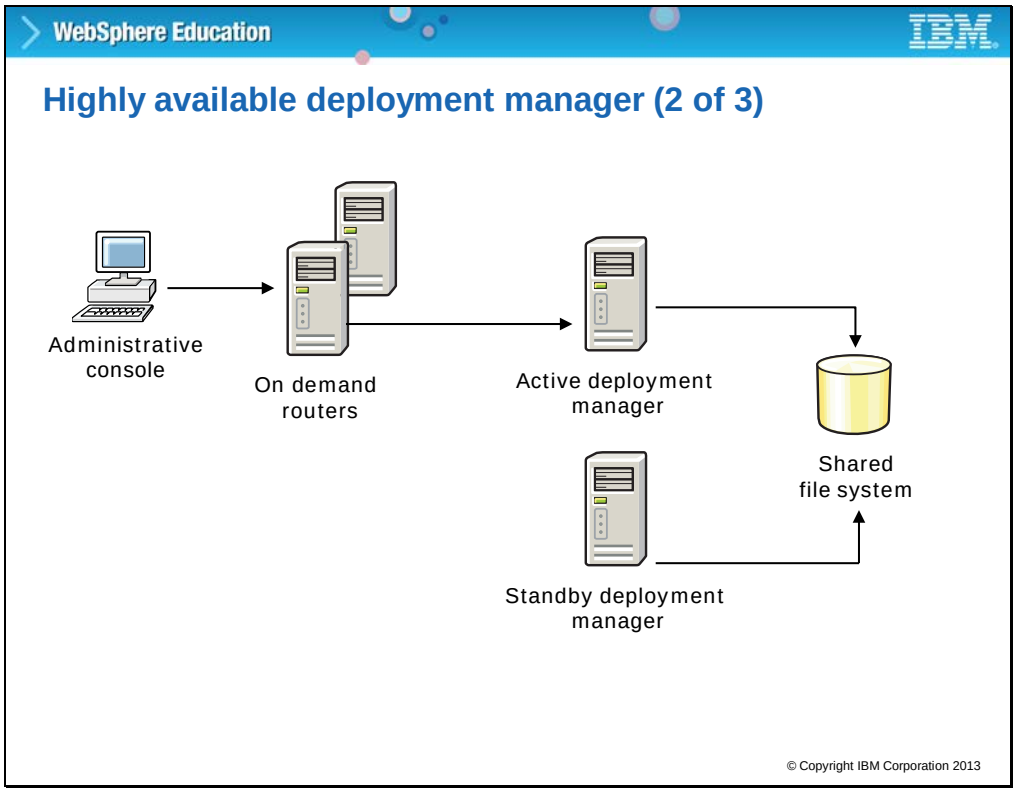
Slide 41



Although it is not required to have deployment manager that is always running, you might require highly available administrative capability, especially in environments that have significant number of new application deployments, updates, and server monitoring. Multiple instances of a deployment manager remove the single point of failure for cell administration; therefore assuring the attainability of the administrative console, wsadmin, and scripting features to manage your environment. WebSphere Application Server provides a mechanism for cloning your existing deployment manager, thus achieving high availability, by employing redundant deployment managers with a hot-standby model and the use of a shared file system.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

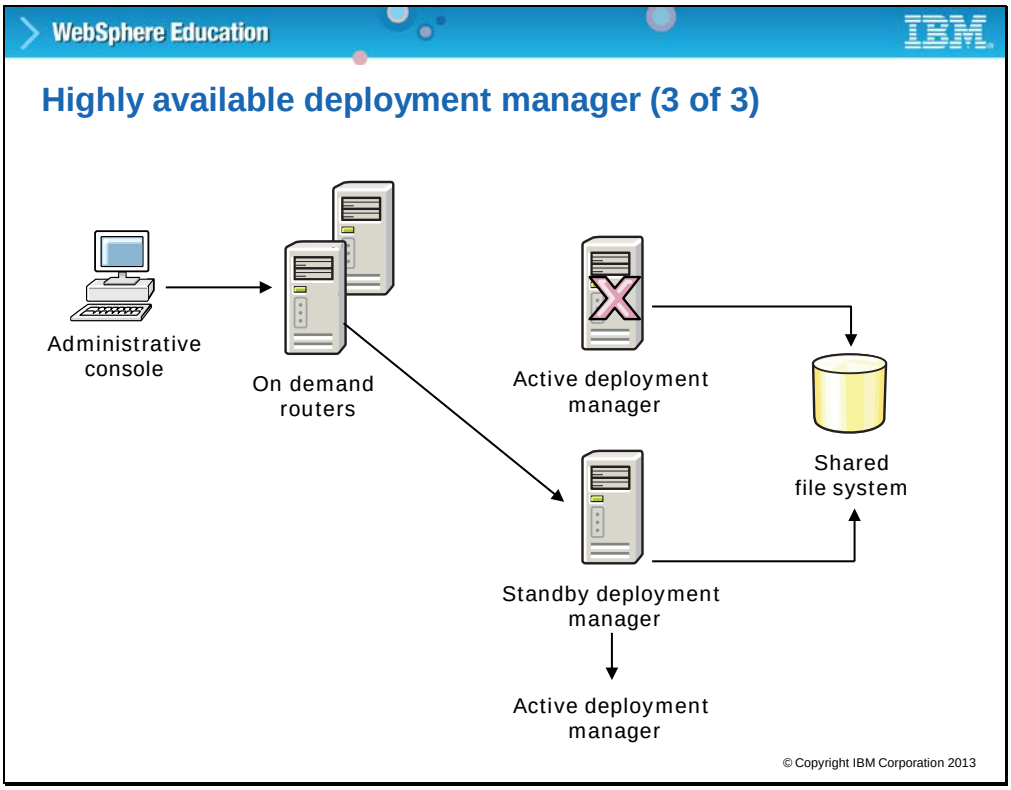
Slide 42



In this paradigm, one of the deployment managers is elected as primary. As primary it is considered an active deployment manager that is hosting the cell-wide endpoints for the administrative functions. Other deployment managers are considered backups and are kept in standby mode and are available to take over the active role in case of failure or termination of the primary manager.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

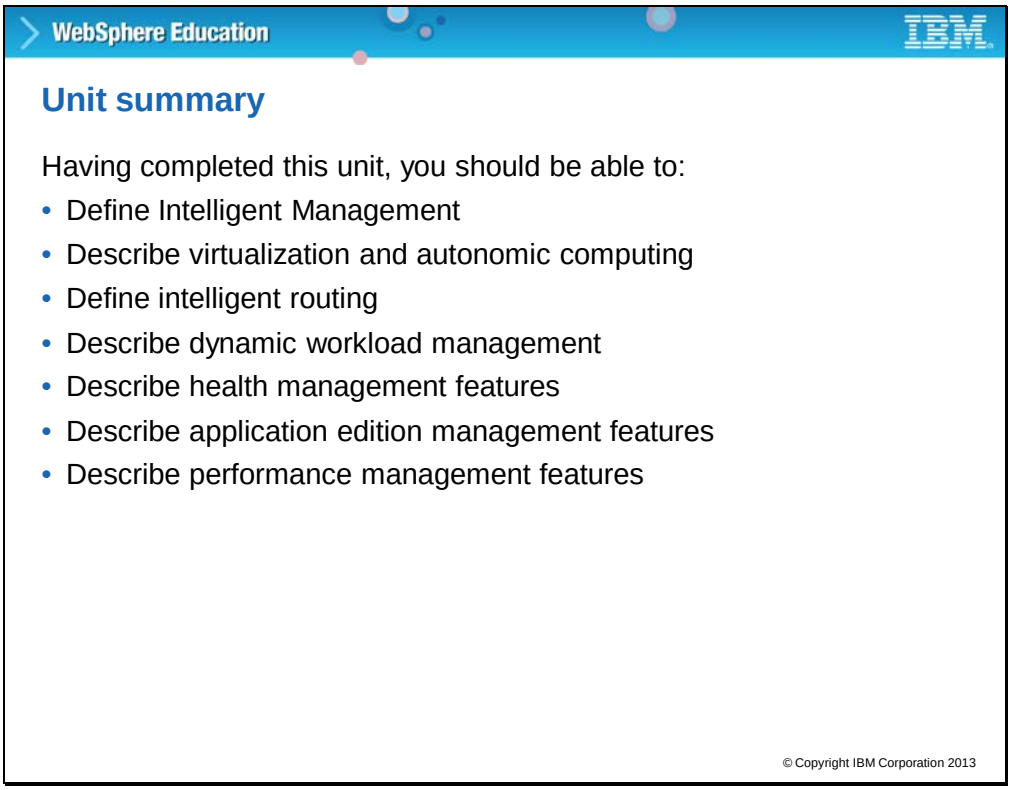
Slide 43



A highly available deployment manager component runs in each deployment manager to control which deployment manager is elected as the active one.

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.

Slide 44



Having completed this unit, you should be able to:

- Define Intelligent Management
- Describe virtualization and autonomic computing
- Define intelligent routing
- Describe dynamic workload management
- Describe health management features
- Describe application edition management features
- Describe Performance Management features

© Copyright IBM Corporation 2013
Course materials may not be reproduced in whole or in part without the prior written permission of IBM.