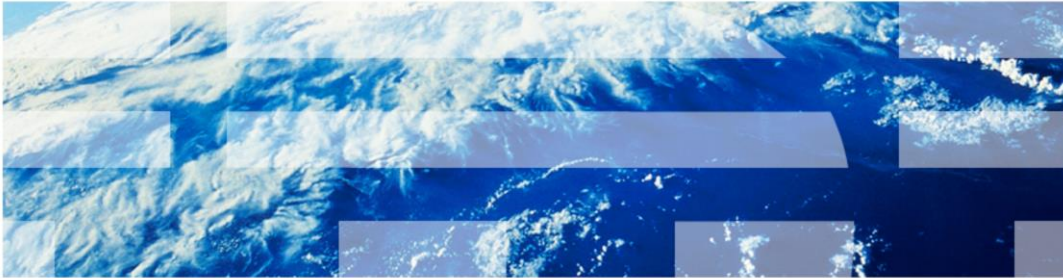

IBM PureApplication System

Overview of Intelligent Management



© 2013 IBM Corporation

This presentation presents an overview of the key concepts and features of Intelligent Management functionality in IBM PureApplication System.

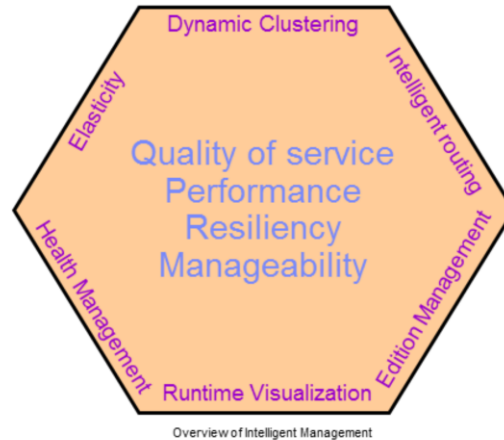
Table of contents

- Intelligent Management functionality overview
 - Dynamic clustering
 - Elasticity
 - Intelligent routing
 - Other important features
 - Health management
 - Application edition management
 - Runtime visualization

The Intelligent Management functionality in IBM PureApplication System provides you with additional capabilities for your WebSphere Application System environment. This presentation discusses Dynamic Clustering, Elasticity and Intelligent Routing, and a glimpse at three other important features provided by Intelligent Management: Health Management, Application Edition Management, and Runtime visualization.

Benefits of Intelligent Management

- Virtual resources in PureApplication System are:
 - Shared
 - Allocated and released dynamically
 - Controlled by performance goals that you set
- Extends your existing WebSphere Application Server middleware systems



Intelligent Management is a set of features focused around quality of service, performance, resiliency, and manageability for your WebSphere Application Server middleware systems.

It provides a virtualized environment that allows virtual resources to be shared and allocated dynamically based on performance goals that you define 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 use the virtual resources that you already have in your deployed systems.

Dynamic Clustering, in concert with Elasticity and Intelligent Routing, allows you to deploy your applications to a cluster of servers which can automatically adjust depending on your service goals. This allows you to take better advantage of hardware resources in a cloud environment and to improve performance. To enhance resiliency and manageability, Intelligent Management includes Health Management, Application Edition Management, and Runtime Visualization, which you hear more about in subsequent slides.

Intelligent Management automatically available in HVE 8.5

- Included as part of WebSphere Application Server Network Deployment Hypervisor Edition, beginning in V8.5.0.0
 - Implicitly included in the catalog image
 - ODR available as pattern part by default
 - Advanced Options in pattern editor available by default for:
 - Dynamic clustering
 - ODR clustering
 - Elasticity
 - Health policies and overload protection

WebSphere Application Server 8.5.0.0 64-bit RHEL 6 x86-64 (VM...) Refresh Export Clone Extend

Description: IBM WebSphere Application Server Hypervisor Edition 8.5.0.0 64-bit for RHEL 6 x86-64

Created on: Oct 31, 2012 1:13:01 AM

Current status: Read-only

Updated on: Oct 31, 2012 1:32:23 AM

License agreement: Accepted [view...]

Hypervisor type: PureSystems_ESX

Operating system: RedHat Enterprise Linux 64-Bit, version 6 (RedHat Enterprise Linux Server)

Version: 8.5.0.0

Image reference number: R3_mX_1222.02

Product IDs (e.g., 5724-X89): 5725-C04 (PVU license)
5725-A26 (PVU license)

Contains parts:

- Administrative agents [part product IDs...]
- Custom nodes [part product IDs...]
- Deployment manager [part product IDs...]
- IBM HTTP servers [part product IDs...]
- [show less]
- Job manager [part product IDs...]
- Liberty profile server [part product IDs...]
- On demand routers** [part product IDs...]
- Standalone server [part product IDs...]

No "Intelligent Management Pack" enablement is required

"On demand routers" part included by default

The Intelligent Management functionality is now fully integrated with IBM WebSphere Application Server Hypervisor Edition V8.5.0.0. You do not need to enable the functionality, which was called “Intelligent Management Pack” in previous WebSphere Application Server Hypervisor Edition releases. The example here shows the version 8.5.0.0 hypervisor edition catalog image, which includes Intelligent Management. Note that the “On demand router” part is included.

Enabling Intelligent Management functionality before HVE 8.5

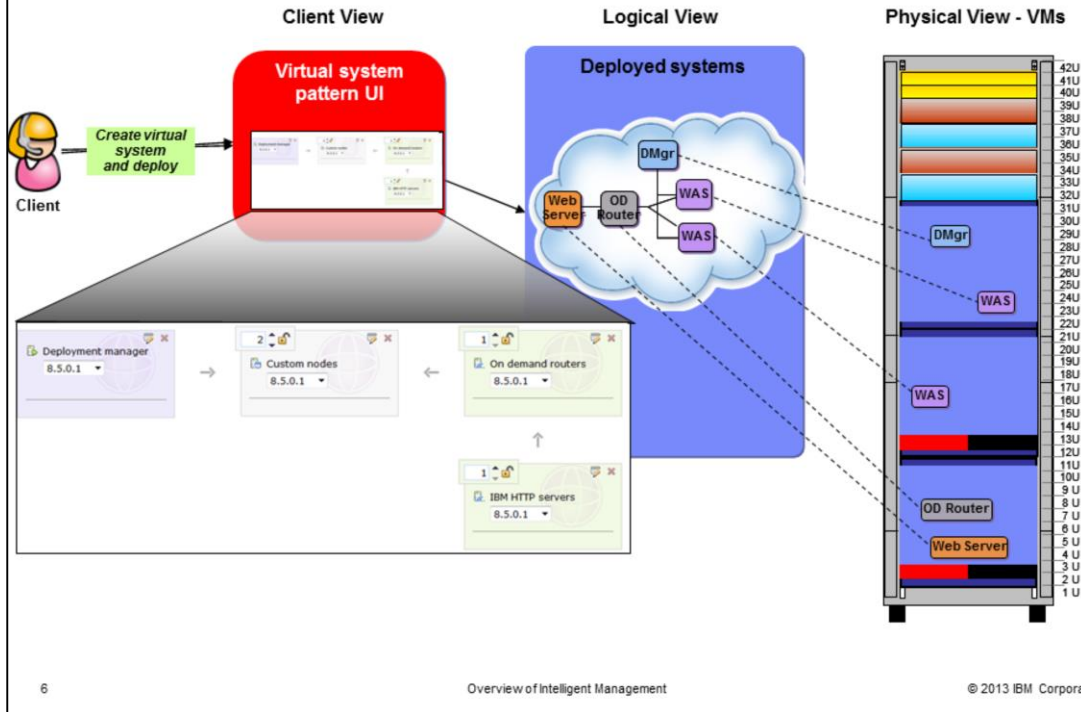
- For previous WebSphere Application Server Hypervisor Editions (V8.0.0.4 and older)
 - Similar Intelligent Management functionality as in V8.5
 - Enable “Intelligent Management Pack” in catalog image
 - Could incur additional cost for Intelligent Management Pack per license agreement

Example of enabling Intelligent Management functionality in a Hypervisor Edition before V8.5

Copy IBM Websphere Application Server 8.0.0.4		Refresh	Export	Clone	Extend	Capture	Lock	Delete
Description:	WAS 8.0.0.4 RHEL with Intelligent Management Pack enabled							
Created on:	Dec 12, 2012 2:25:23 PM							
Current status:	Read-only							
Updated on:	Dec 12, 2012 2:35:32 PM							
License agreement:	Accepted [view...]							
Intelligent Management Pack:	<input type="button" value="Enabled"/> Advanced features are enabled and may result in additional cost. Please refer to the license agreement.							
Hypervisor type:	PureSystems_ESX							
Operating system:	RedHat Enterprise Linux 64-Bit, version 6 (RedHat Enterprise Linux 6)							
Version:	8.0.0.4							

The Intelligent Management functionality was previously called “Intelligent Management Pack.” In releases previous to Hypervisor Edition V8.5 you needed to set the “Intelligent Management Pack” setting within the catalog image to “Enabled” to use the additional functionality. The functionality within the older “Intelligent Management Pack” is the same as that provided in Hypervisor Edition V8.5 and newer.

Logical to physical mapping for virtual systems in the cloud



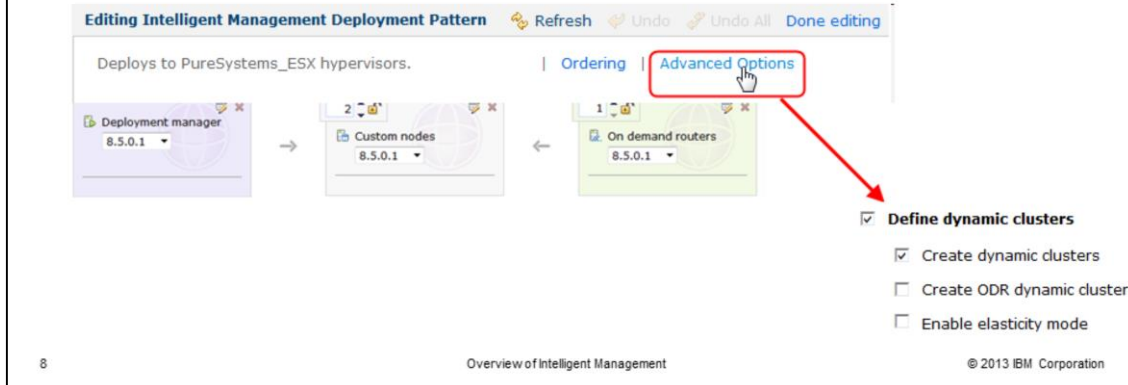
The Intelligent Management functionality is customized within the pattern editor. The “Client View” portion of this slide shows a pattern editor session for a virtual system pattern with Intelligent Management being used. The pattern contains a deployment manager part, a custom node part, an “on demand router” part, and an IBM HTTP Servers part, which is optional. To create a running virtual system, you simply deploy the pattern. The “Logical View” shows the parts and connections when the system is deployed. Notice that each part in the pattern results in a discrete virtual machine when the pattern is deployed. The “Physical View” at the right of this slide give you an idea of where the virtual machines might typically be deployed within the rack.

Dynamic clustering

In this next section you look at an overview of dynamic clustering.

Dynamic clustering overview

- Dynamic clustering – a key concept for dynamic workload management
 - Virtualized
 - Goal-oriented
 - Policy-based
- A “cluster” that can expand and contract to provide:
 - More consistent quality of service for critical applications
 - More efficient utilization of hardware resources
- Enabled in Advanced Options in pattern editor

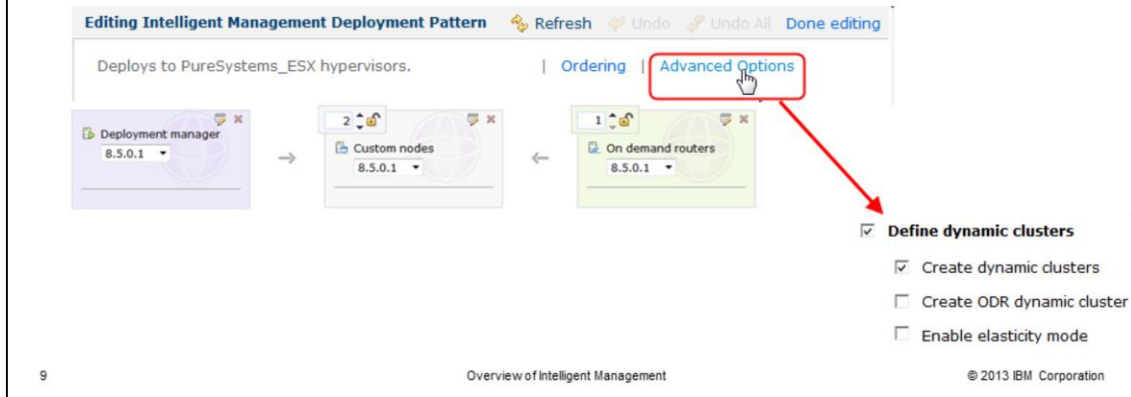


A key concept is the idea of dynamic clusters. These are similar to regular ‘clusters’ available in WebSphere Application Server, except that dynamic clusters can automatically grow or shrink across existing nodes in response to goals that you set. In this slide you see a pattern containing a deployment manager part, a custom nodes part, and an On demand routers part. Advanced Options is opened to show you the option “Create dynamic clusters.”

You deploy your applications to a dynamic cluster, either using scripting associated with the deployment, or using the WebSphere Application Server administrative console after deployment. You then define policies that dictate performance goals that govern the management of workload for that application. Intelligent Management manages work according to that goal. As work comes into the system, resources are dynamically allocated and balanced within the existing “dynamic” resource pool to maintain performance according to your goal. Using dynamic clusters, you might even find that you can afford to deploy more applications onto the same amount of resources, since they are now used in a more efficient manner.

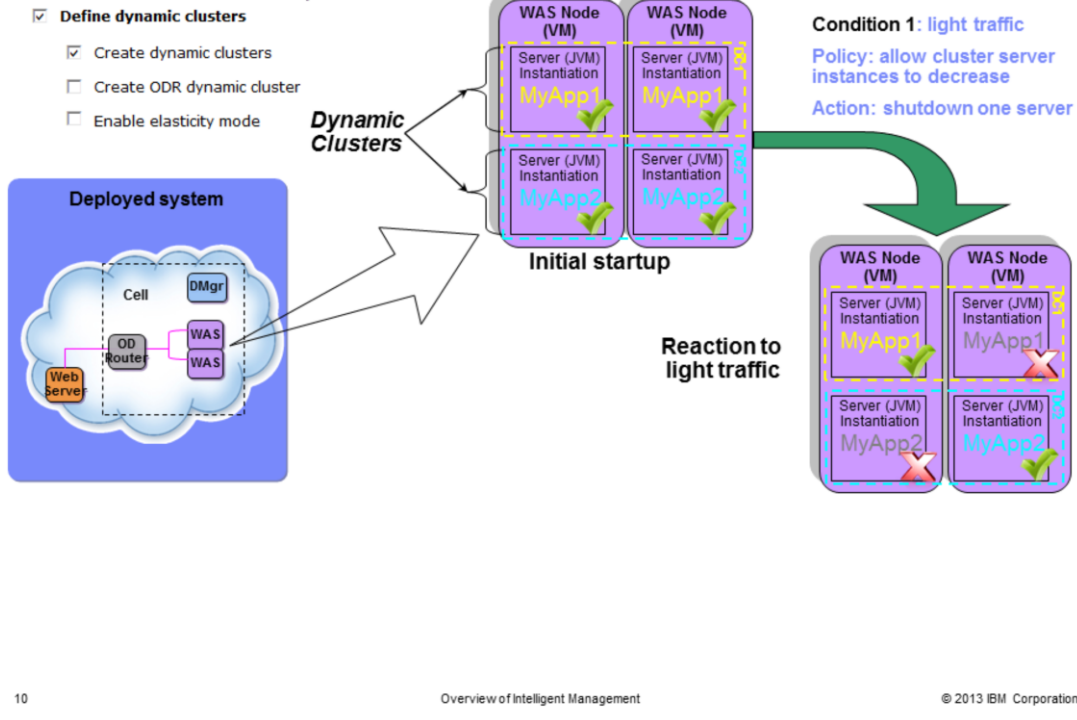
Dynamic clustering and automatic placement controller

- Dynamic application placement
 - Autonomic placement of instances of application servers on nodes
 - Starts and stops preconfigured server instances to meet defined performance goals



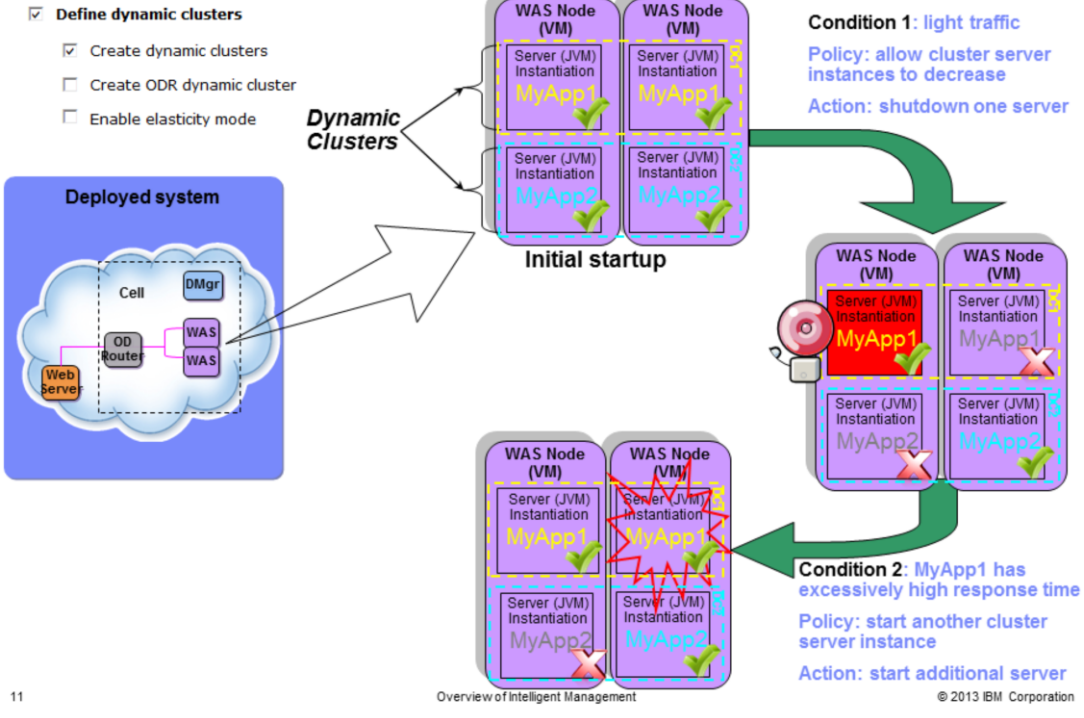
Another key feature is known as the application placement controller. It works with the dynamically sized server clusters to increase or decrease the number of server instances on which a particular application is running to help meet your performance goals. Each node within a dynamic cluster has at least one instance of an application server where that cluster's applications can be started and stopped dynamically as traffic for the application increases or decreases. This feature is automatically enabled when you define a dynamic cluster.

Dynamic clustering – starting and stopping cluster server instances



To get a better idea of how dynamic clustering and the automatic placement controller work together, look at this example. At the top of the slide, note the two virtual machines depicted. Within each virtual machine, the “MyApp1” and “MyApp2” applications have been deployed in their own respective dynamic clusters. Both server instances in each virtual machine are initially started, as defined in your initial startup configuration. If more servers are running than are needed to handle the load during very light load conditions, the system is permitted to shut down a server instance. How this is handled is defined according to a policy you configure. With “Condition 1” in effect, with light traffic spread across two sets of two server instances, Intelligent Management selects a server instance in each of the two dynamic clusters to shut down. It indicates to the on demand router to cease routing traffic to these server instances. Then using the application placement controller, Intelligent Management initiates a shutdown of the two selected application server instances. Look at the graphic on the right side of the screen to see the result. Now, only one application server instance is running for each application, which is a more efficient use of your virtual resources.

Dynamic clustering – starting and stopping cluster server instances



Look now at how Intelligent Management can start additional resources as needed. In the graphic at the right of the screen, the “alarm bell” indicates that MyApp1 activity has increased to the point where your response time goals are no longer being met. This leads to “Condition 2” (in the lower right corner of the screen) – MyApp1 has an excessively high response time. Intelligent Management then, according to your policy settings, starts another cluster server instance on the second Virtual Machine to run MyApp1, depicted in the graphic at bottom center of the screen. It then instructs the on demand router to begin routing traffic for MyApp1 to both server instances. Thus you can see that Intelligent Management can both stop and start cluster server instances as needed to react to your policy goals.

Section

Elasticity

In this next section you see an overview of elasticity.

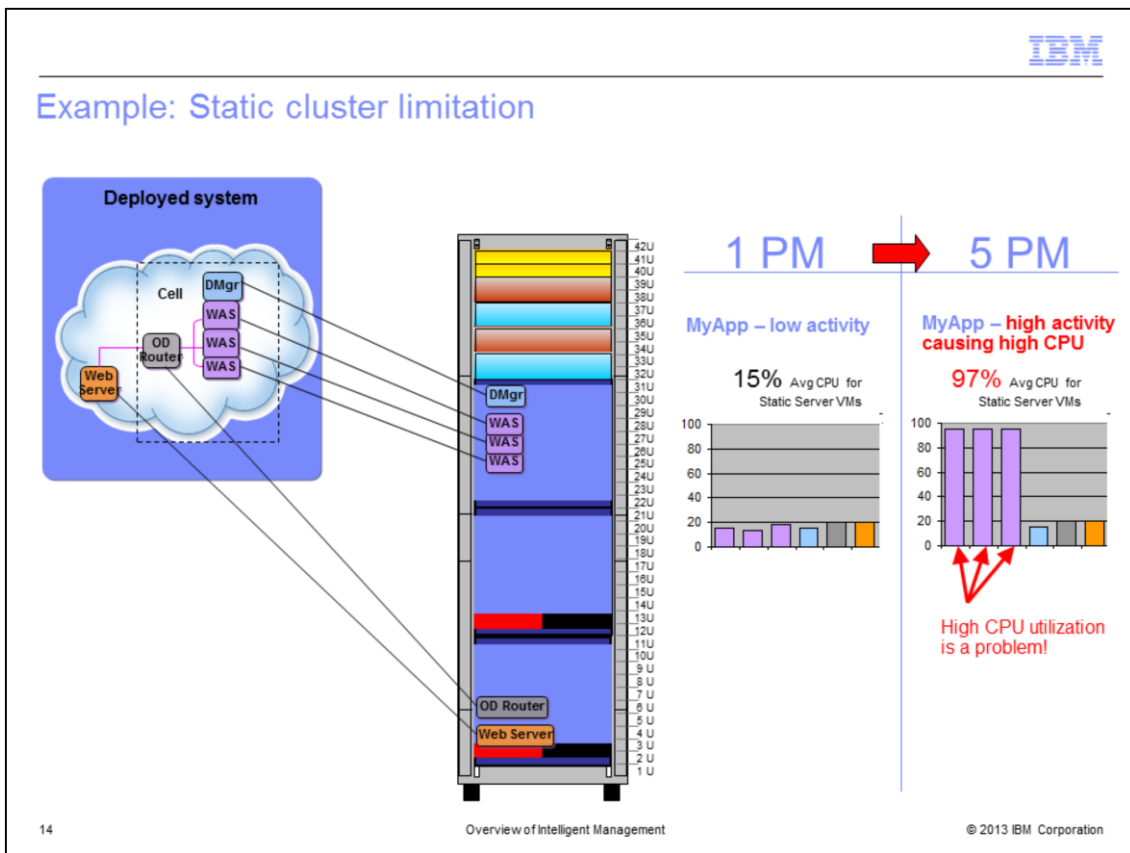
Elasticity overview

- Define dynamic clusters
 - Create dynamic clusters
 - Create ODR dynamic cluster
 - Enable elasticity mode

- Requires dynamic clusters
- Function: add or remove of a virtual machine within a WebSphere cell
 - Addition or removal determined by policy you set
 - Interacts with PureApplication System to manage VM creation and deletion
 - Attributes of the new VM:
 - A clone of the VM which started the “Add” action
 - Server instances have the cluster applications automatically installed
 - On demand router automatically begins sending html traffic to the applications on new server instances when the new VM is initialized
- Typical usage examples:
 - Add a VM when CPU load on the dynamic cluster is high
 - Remove a VM when CPU load on the dynamic cluster is low

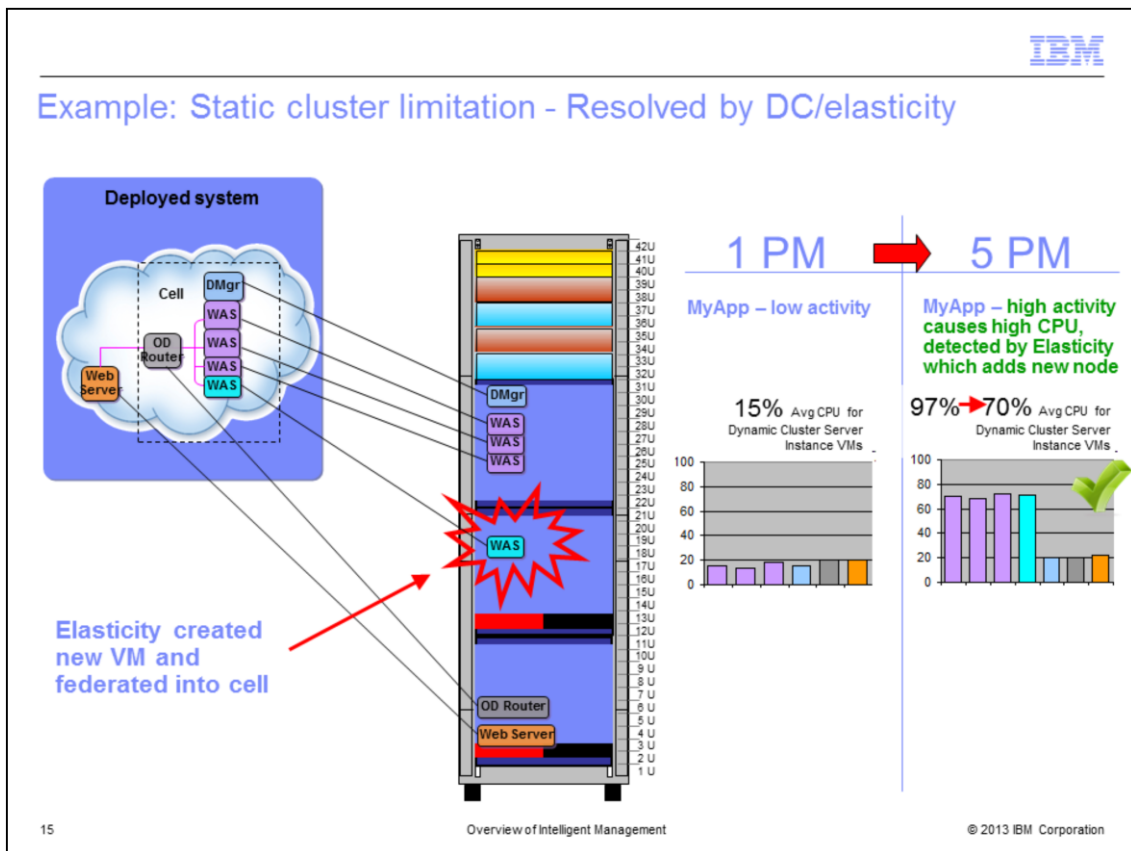
Elasticity further enhances Intelligent Management by providing a facility to autonomically create or remove a custom node virtual machine within the cell based on your policy goals. In this slide at the upper right, you see the Advanced Options settings that are required to enable VM elasticity. The “Create dynamic clusters” setting is required before you can select “Enable elasticity mode.” Once selected, these Advanced Options settings provide you additional parameters in your pattern editor session to control your dynamic cluster implementation and your elasticity settings. When you enable elasticity, the application placement controller communicates with PureApplication System to create a clone of a custom node virtual machine when additional compute power is required. Conversely, Intelligent Management can initiate the removal of a custom node virtual machine when excessive compute power is allocated. Intelligent Management coordinates the addition or removal of virtual machines with the on demand router so that HTML traffic is efficiently handled.

Example: Static cluster limitation



Here is an example of a static cluster limitation, which can be eliminated by the elasticity feature. At 1 PM the HTTP traffic going to the virtual machine running MyApp is very light and remains so for most of the afternoon. However at 5 PM, there was a sudden burst of HTTP activity for MyApp that drives the CPU usage for the virtual machines to 97 percent. At that point, MyApp response time suffers due to excessive CPU utilization. To alleviate this problem with static clusters, the administrator must increase the size of the static cluster to cater for this level of activity. He must do this manually by adding an additional node (an additional virtual machine) to the static cluster. This might mean that for large parts of the day when MyApp experiences light HTTP traffic, unneeded compute nodes are allocated to the cell.

Example: Static cluster limitation - Resolved by DC/elasticity



The use of dynamic clusters and elasticity resolves the static cluster problem on the previous slide. You can define a dynamic cluster with a minimum number of nodes (server instances) active. Then using elasticity, your goals can allow PureApplication System and Intelligent Management to create a new virtual machine when CPU utilization is above your goal levels. Conversely, the elasticity function can remove one or more custom node virtual machines when CPU utilization is below your goal levels.

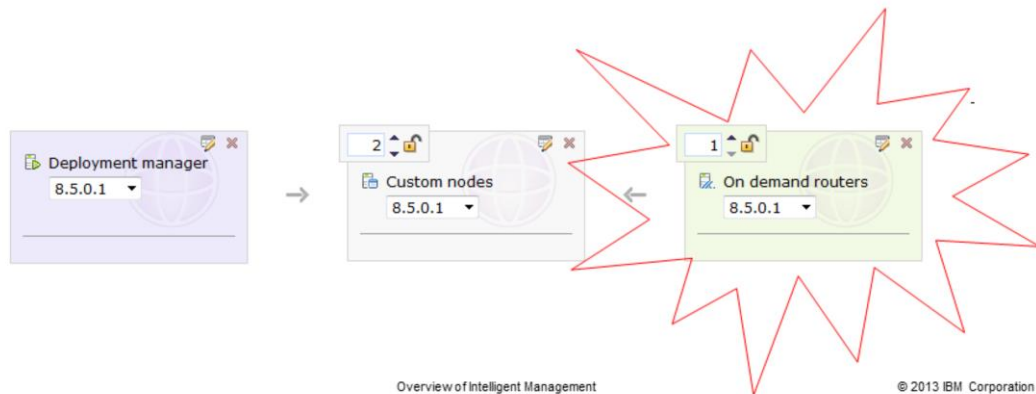
In the example above, when the high HTTP traffic at 5PM causes increased response times, the elasticity function and automatic placement controller creates a new virtual machine and custom node. The new node containing MyApp is federated into the cell, and the server instance started. Then Intelligent Management feature instructs the on demand router to begin sending traffic to the new server instance on the new virtual machine. The total average CPU utilization for all four custom nodes drops as the load spreads to the new node, providing adequate performance during this peak. When CPU activity reduces below your goal level, one or more of the custom node virtual machines can be removed from the cell, so that unneeded compute node resources can be returned to PureApplication System's pool of resources.

Intelligent routing

In this next section you see an overview of Intelligent Routing.

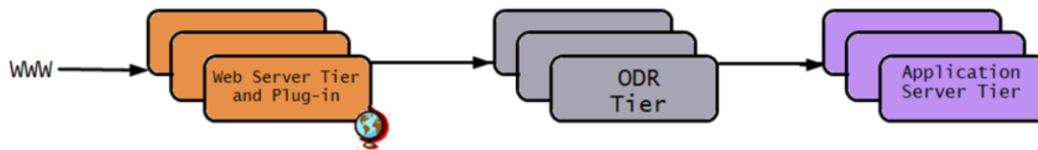
Intelligent routing

- On-demand router (ODR)
 - Functionality is built as a set of filters to the proxy server functions
 - Provides routing of HTTP traffic into a collection of cells
 - Provides three key on-demand features:
 - Flow control and queuing
 - Prioritization
 - Dynamic workload management



Intelligent Management provides a part called the on-demand router. This is an enhanced version of the WebSphere proxy server, which was originally included in WebSphere Application Server version 6.1. The on-demand router is an intelligent HTTP proxy that acts as the entry point for traffic coming into a WebSphere Application Server cell, performing request prioritization, flow control, and dynamic workload management. It works with dynamic clustering, elasticity, and application edition management to ensure that HTTP traffic is efficiently and seamlessly delivered in the dynamic environment provided by Intelligent Management.

Common topology



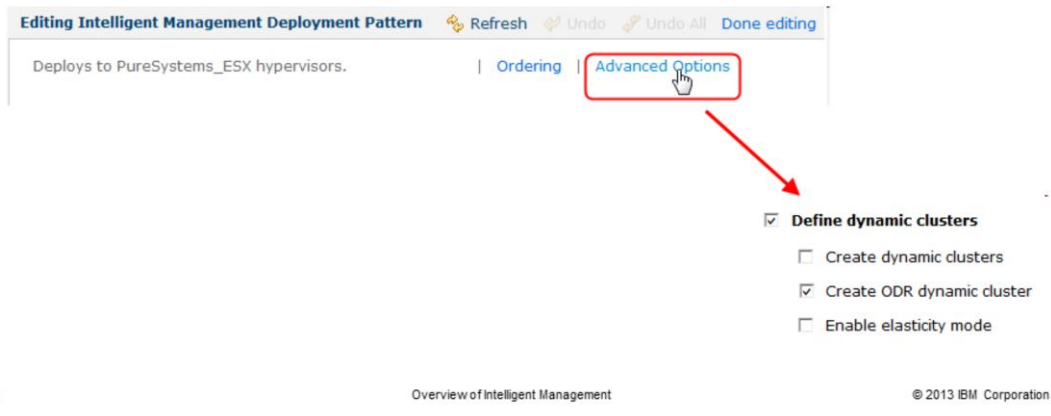
- Typical topology has multiple on-demand routers for scalability and availability
- On-demand router
 - Generates the plugin-cfg.xml file for web server plug-ins
 - Dynamically learns configuration of backend cells
 - Can be connected to multiple cells simultaneously
 - Supports routing to any HTTP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

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

Other topology options are also possible, including placing the On Demand Router in front of, or as a replacement for, the web server tier. The example here shows multiple on demand routers in a cluster, which is discussed on the next slide.

ODR dynamic cluster

- Define On Demand Routers dynamic cluster in the pattern
 - Controlled by Advanced Options in pattern editor
- Allows greater reliability and resiliency for routing functionality



To avoid a single point of failure, Intelligent Management allows you to create a dynamic cluster for on demand routers. As with dynamic clusters and elasticity, it is controlled in the pattern by settings in Advanced Options. By creating a dynamic cluster for the on demand routers, you can eliminate a single point of failure for the on demand routers and still efficiently use and release compute node resources with PureApplication System.

Other important features

The next section discusses other important features of Intelligent Management.

Health management

- An Intelligent Management environment can be monitored for various software health conditions
 - Memory usage, CPU usage, heap usage, memory leaks, age
 - Number of requests, stuck requests, response time, storm drain detection
- Health conditions and reactions are defined by health policies
 - Servers can be restarted as a corrective or preventive action
 - Java heap can be dumped for memory problems
 - Email notifications for policy violations

The screenshot shows the 'Editing Intelligent Management Deployment Pattern' interface. At the top, there are buttons for 'Refresh', 'Undo', 'Redo', and 'Done editing'. Below this, the main content area displays 'Deploys to PureSystems_ESX hypervisors.' and 'Updated on Dec 14, 2012 9:01:53 AM | Ordering'. A red box highlights the 'Advanced Options' link, with a red arrow pointing to a detailed configuration panel on the right.

The configuration panel on the right contains the following options:

- Enable overload protection**
 - Memory overload protection
 - CPU overload protection
- Configure standard health policies**
 - Excessive heap usage
 - Memory leak
 - Maximum server age
 - E-mail notification list
- Configure on demand router-dependent health policies**
 - Maximum requests served
 - Excessive number of timed out requests
 - Excessive average response time
 - Storm drain detection

Health management is another useful administration feature, providing the capability to configure “health policies.” Health policies define software conditions that might indicate a problem - such as excessive memory usage, stuck requests and excessive uptime without a restart - and automatically 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 can choose to automatically restart the server weekly. Similarly, you can also choose to generate a dump of the Java heap every time memory consumption exceeded a certain level, and restart the server as well.

Application edition management

- Intelligent Management 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 Java EE application (similar to *version*)
 - May be a distinct build version, or the same build with different deployment bindings

The screenshot shows the WebSphere Application Server administrative console. On the left, the 'Applications' tree is visible. The main area displays the 'Edition Control Center' for 'YourApplicationEdition'. A table shows the application details:

Applications	Type	Editions	Active	Validation
YourApplicationEdition	Java 2 Platform, Enterprise Edition	2	1	0
Total 1				

Below this, the 'Edition Control Center > Manage Editions' page is shown for 'YourApplicationEdition'. It includes buttons for 'Activate', 'Deactivate', 'Validate', 'Cancel Validation', and 'Rollout'. A table lists the editions:

Select	Editions	Description	Target	State
<input type="checkbox"/>	1.0.0	Initial deployment	WebSphere:cell=IntelligentManagementCell_1,cluster=DynCluster_0 WebSphere:cell=IntelligentManagementCell_1,cluster=DynCluster_1	ACTIVE
<input type="checkbox"/>	1.0.1	Bugfix Edition	WebSphere:cell=IntelligentManagementCell_1,cluster=DynCluster_0 WebSphere:cell=IntelligentManagementCell_1,cluster=DynCluster_1	INACTIVE
Total 2				

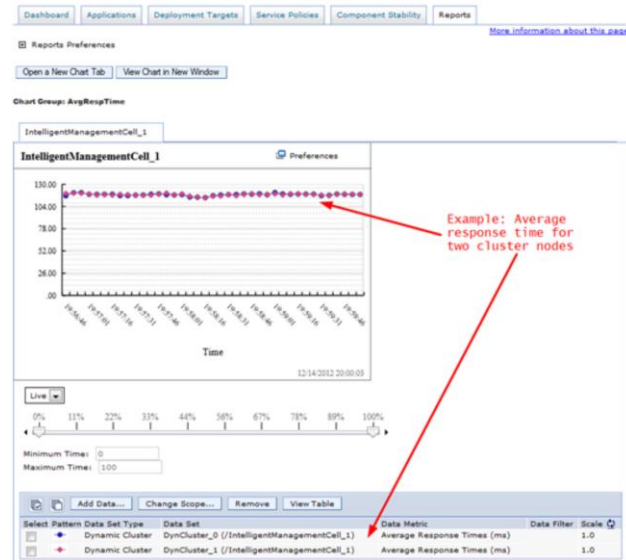
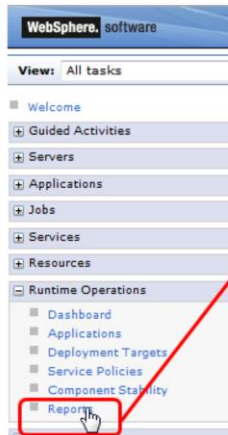
At the bottom of the console, it says 'Overview of Intelligent Management' and '© 2013 IBM Corporation'.

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

Runtime visualization

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

WebSphere Application Server administrative console



23

Overview of Intelligent Management

© 2013 IBM Corporation

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

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

Section

Summary

The following section provides a summary of this presentation.

Summary

- **Dynamic clustering** creates a virtualized, goal-based environment for workload management
- **Elasticity** function works with dynamic clustering to create or remove custom node virtual machines as workload varies
- **Intelligent routing** provides HTTP traffic support for the dynamic workload environment created by Intelligent Management
- Other important features
 - **Health management** can trigger alerts and take action when software health problems are detected
 - **Application edition management** enables seamless rollout of multiple editions of an application
 - **Runtime visualization** provides live data on the performance and health characteristics of the entire cell

The advanced features provided by Intelligent Management includes many tools for more automation, better workload management, more robust health monitoring, and easier administration for large systems. Dynamic clustering provides a goal-based environment for workload management. Working in concert with dynamic clustering, the elasticity feature permits the PureApplication System to create or remove custom node virtual machines as your workload varies. Intelligent routing supports the dynamic workload environment provided by Intelligent Management by appropriately prioritizing, queuing, and routing HTTP traffic. Health management triggers alerts and takes action when software health issues are detected. Application edition management enables the seamless control and rollout of multiple application editions. And finally, runtime visualization provides you live data on the performance and health characteristics of the entire cell.

Trademarks, disclaimer, and copyright information

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)" at <http://www.ibm.com/legal/copytrade.shtml>

Other company, product, or service names may be trademarks or service marks of others.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2013. All rights reserved.