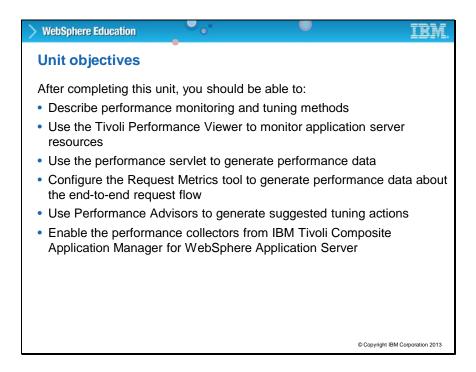


Unit 20: Performance monitoring

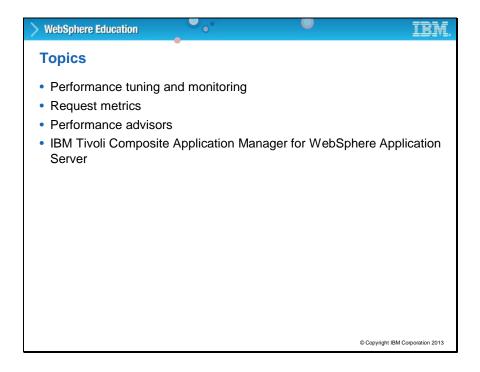
This unit describes performance monitoring methods and tools available through the administrative console.



Title: Unit objectives

After completing this unit, you should be able to:

- · Describe performance monitoring and tuning methods
- Use the Tivoli Performance Viewer to monitor application server resources
- Use the performance servlet to generate performance data
- Configure the Request Metrics tool to generate performance data about the end-to-end request flow
- Use Performance Advisors to generate tuning actions
- Enable the performance collectors from IBM Tivoli Composite Application Manager for WebSphere Application Server

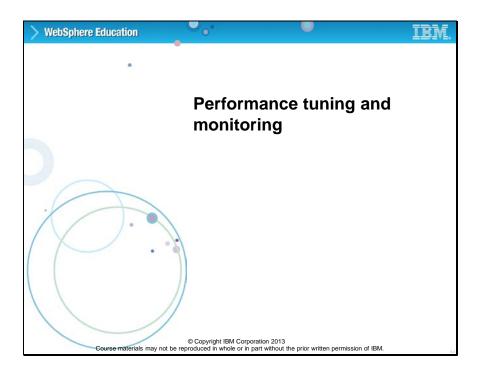


Title: Topics

This unit describes the following topics.

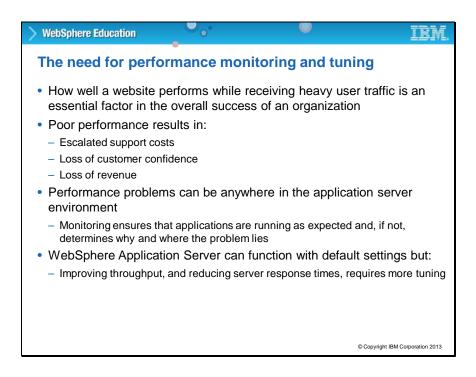
- Performance tuning and monitoring
- Request metrics
- Performance advisors
- ITCAM for WebSphere Application Server

Slide 4



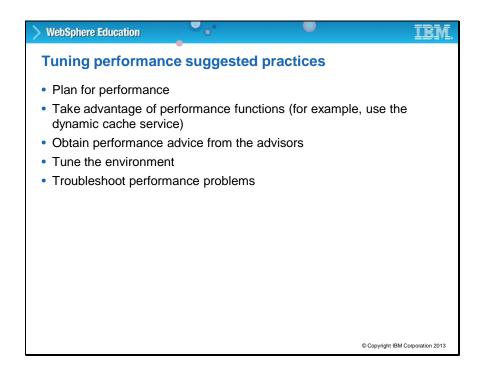
Topic: Performance tuning and monitoring

This topic covers some methods for tuning and monitoring application server performance.



Title: The need for performance monitoring and tuning

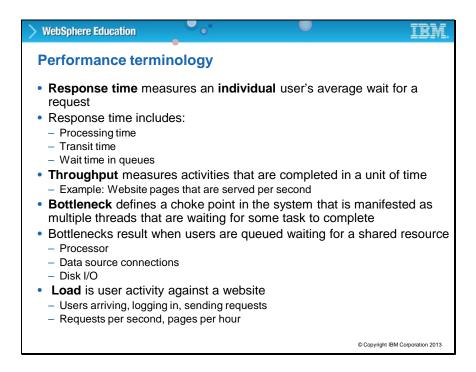
The goal of performance monitoring is to collect runtime statistics on your application and its environment to quantify their performance behavior. It allows you to determine whether your application meets its performance objectives and helps to identify any performance bottlenecks. It is important to monitor system performance because poor performance can result in higher support costs, loss of customer confidence, and loss of revenue. Monitoring ensures that applications are running as expected, and if not, the cause of performance problems can be investigated. WebSphere Application Server can function well with default settings, but some applications can require further tuning for optimal performance.



Title: Tuning performance best practices

Tuning WebSphere Application Server is a critical part of getting the best performance from your website. But tuning WebSphere Application Server involves analyzing performance data and determining the optimal server configuration. This determination requires considerable knowledge about the various components in the application server and their performance characteristics. The performance advisors encapsulate this knowledge and analyze the performance data. The advisors provide configuration recommendations to improve the application server performance. Therefore, the performance advisors provide a starting point for tuning the application server. Keep in mind the following suggestions:

- Take advantage of performance functions.
- Obtain performance advice from the advisors.
- Tune the environment.
- Troubleshoot performance problems.



Title: Performance terminology

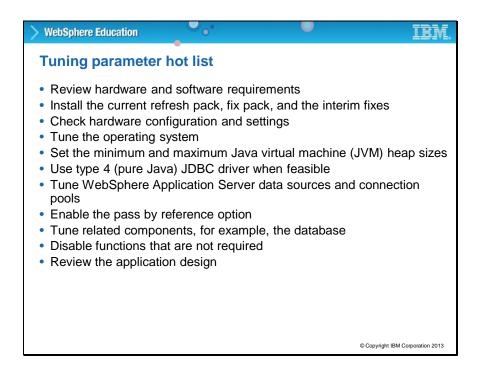
This slide displays a list of terms that are commonly associated with performance. Creating a glossary of performance terms is essential for good performance monitoring. The glossary is used to ensure that everyone who is involved understands the meaning of each term, decreasing the chance of inaccurate performance statistics.

Response time measures average wait for a request for an individual user. Response times include the cumulative time that is spent in processing the request, time that is spent in transit between systems, and the wait time that is spent in queues.

Throughput measures activities that are completed in a unit of time. For example, the number of website pages that are served per second.

Bottleneck defines a choke point in the system that is manifested as multiple threads that are waiting for some task to complete. A bottleneck results when user requests are queued waiting for a shared resource such as processors, data source connections, and disk I/O.

Load is user activity against a website. Some examples are the number of users that arrive, log in, and send requests.



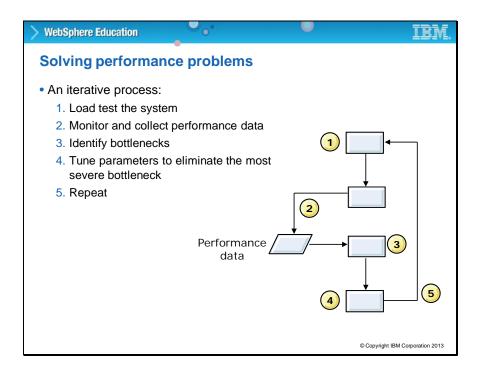
Title: Tuning parameter hot list

This hot list contains recommendations that can improve performance, scalability, or both, for many applications.

WebSphere Application Server provides several tunable parameters and options to match the application server environment to the requirements of your application. The details of each of these items are documented in the information center.

The last item is especially important. Review your application design. You can track many performance problems back to the application design. Review the design to determine whether it is the source of performance problems.

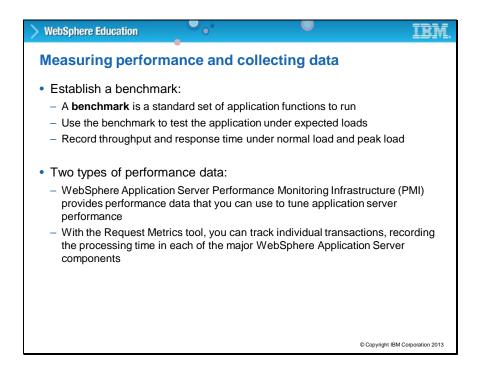
Slide 9



Title: Solving a performance problem

Your application and its runtime environment must also be tuned optimally. This process entails conducting many iterations of a monitor, tune, and test cycle. In short, monitoring, performance testing, and tuning are essential tasks for ensuring a well-performing, application-serving environment.

This process is often iterative because when one bottleneck is removed, some other part of the system now constrains the performance. For example, replacing slow hard disks with faster ones might shift the bottleneck to the processor of a system.



Title: Measuring performance and collecting data

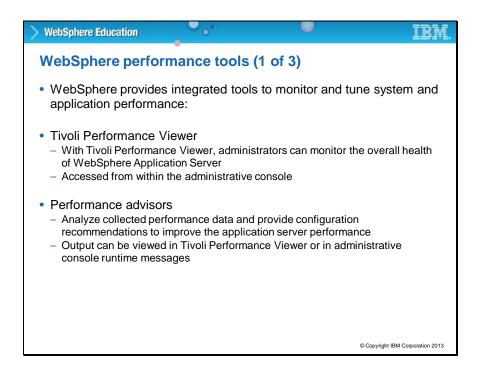
To measure performance, begin by choosing a benchmark. A benchmark is a standard set of operations to run. This benchmark exercises those application functions experiencing performance problems. Complex systems frequently need a warm-up period to cache objects and optimize code paths. System performance during the warm-up period is much slower than after the warm-up period. The benchmark must be able to generate work that warms up the system before recording the measurements that are used for performance analysis. Depending on the system complexity, a warm-up period can range from a few thousand transactions to longer than 30 minutes.

If the performance problem under investigation occurs only when many clients use the system, then the benchmark must also simulate multiple users. Another key requirement is that the benchmark must be able to produce repeatable results. If the results vary more than a few percent from one run to another, consider the possibility that the initial state of the system might not be the same for each run. It might also be that the measurements are made during the warm-up period, or that the system is running several workloads.

Several tools facilitate benchmark development. The tools range from tools that merely call a URL, to script-based products that can interact with dynamic data that the application generates. IBM Rational has tools that can generate complex interactions with the system under test and simulate thousands of users. Producing a useful benchmark requires effort and must be part of

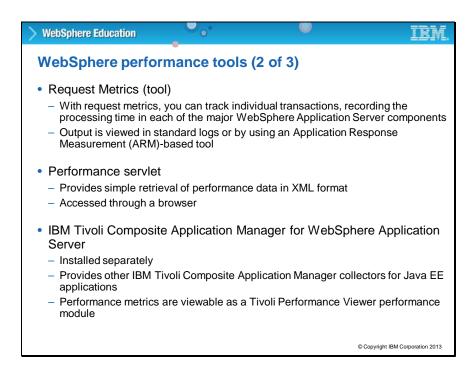
the development process. Do not wait until an application goes into production to determine how to measure performance.

The benchmark records throughput and response time results in a form to allow graphing and other analysis techniques. The performance data that WebSphere Application Server Performance Monitoring Infrastructure (PMI) provides helps to monitor and tune the application server performance. Request metrics are another source of performance data that WebSphere Application Server provides. Request metrics allow a request to be timed at WebSphere Application Server component boundaries, enabling a determination of the time that is spent in each major component.



Title: WebSphere performance tools (1 of 3)

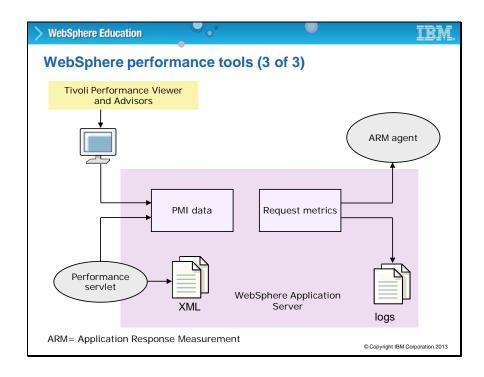
WebSphere Application Server provides the following tools, or facilities, to monitor and tune system performance: The Tivoli Performance Viewer or TPV, request metrics, performance advisors, and the performance servlet, all of which are described in more detail in this unit.



Title: WebSphere performance tools (2 of 3)

PMI data can be viewed by using tools that are built into the administrative console, or can be viewed in XML form by using the performance servlet. Request metrics are written to the logs, and can also be viewed by using an ARM agent. IBM Tivoli Composite Application Manager (ITCAM) for WebSphere is introduced in version 7. ITCAM for WebSphere gives you some additional monitoring capabilities such as CPU usage per application.

Slide 13



Title: WebSphere performance tools (3 of 3)

WebSphere provides integrated tools to monitor and tune system and application performance.

The **Tivoli Performance Viewer**, which is accessed from within the administrative console, gives administrators the ability to monitor the overall health of WebSphere Application Server.

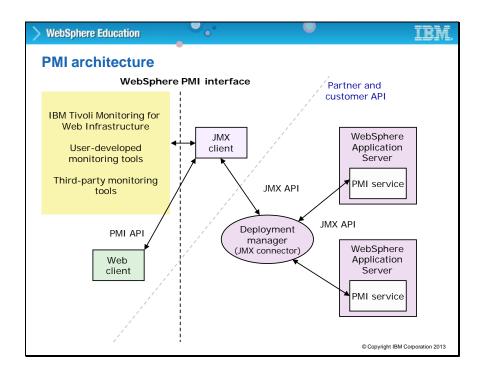
The **Request metric (tool) gives** you the ability to track individual transactions, and record the processing time in each of the major WebSphere Application Server components.

Output is viewed in standard logs or by using an Application Response Measurement (ARM)-based tool.

Performance advisors analyze collected performance data and provide configuration recommendations to improve the application server performance. Output is viewed in Tivoli Performance Viewer or in administrative console runtime messages.

The **Performance servlet** provides simple retrieval of performance data in XML format, which is accessed through a browser.

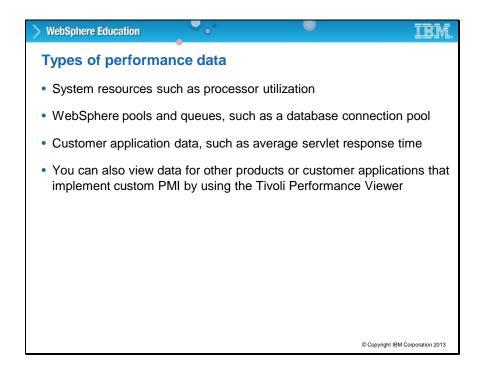
Slide 14



Title: PMI architecture

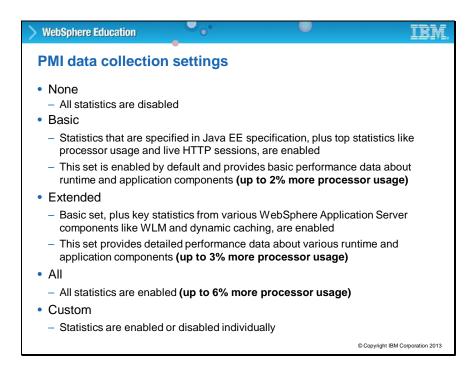
The Performance Monitoring Infrastructure (PMI) uses client/server architecture.

The figure shows the overall PMI architecture. On the right side, the server updates and keeps PMI data in memory. The left side displays a web client, a Java client, and a JMX client that retrieves the performance data. This data consists of counters such as servlet response time and data connection pool usage. The data points are then retrieved by using a web client, a Java client, or a Java Management Extensions (JMX) client. WebSphere Application Server contains Tivoli Performance Viewer, a Java client, which displays and monitors performance data. The server collects performance data from various WebSphere Application Server components. A client retrieves performance data from one or more servers and processes the data. WebSphere Application Server supports the Java EE Management Reference Implementation (JSR-77).



Title: Types of performance data

Tivoli Performance Viewer is used to help manage configuration settings by viewing the various graphs or by using the Tivoli Performance Advisor. For example, by looking at the summary chart for thread pools, you can determine whether the thread pool size must be increased or decreased by monitoring the percent usage. After configuration settings are changed based on the data that is provided, you can determine the effectiveness of the changes. To help with configuration settings, use the Tivoli Performance Advisor. The Advisor assesses various data while your application is running, and provides advice about configuration settings to improve performance.



Title: PMI data collection settings

PMI uses statistics sets to specify the type and amount of performance data to collect.

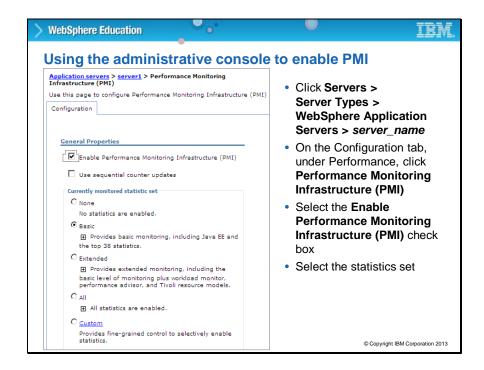
PMI counters are enabled, based on a monitoring or instrumentation level. The levels are None,

Basic, Extended, All, and Custom. These levels are specified in the PMI module XML file.

Enabling the module at a certain level includes all the counters at that level plus counters from levels below that level. So, enabling the module at the extended level enables all the counters at that level plus all the basic level counters as well.

PMI is set to monitor at a Basic level by default.

Slide 17

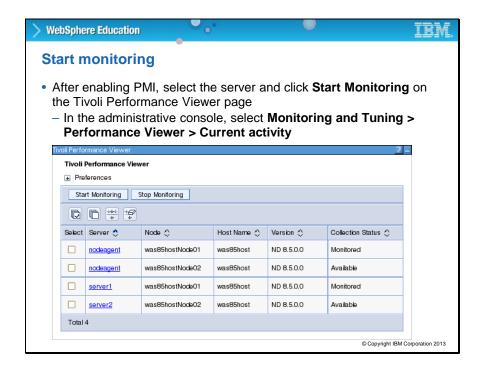


Title: Using the administrative console to enable PMI

For a particular server, navigate to its **Performance Monitoring** Infrastructure configuration panel.

Select the **Enable Performance Monitoring Infrastructure (PMI)** check box and select a statistics set.

Slide 18

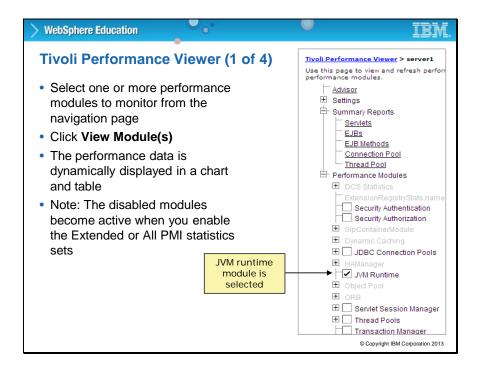


Title: Start monitoring

After enabling PMI for the server, in the console, select **Monitoring and Tuning > Tivoli**performance Viewer > server_name > Start Monitoring. Then, go back to Monitoring and

Tuning > Tivoli performance Viewer > current activity.

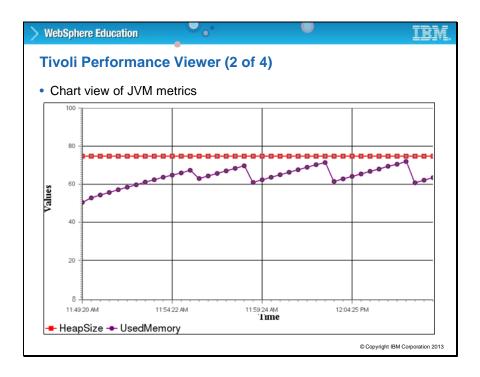
Slide 19



Title: Tivoli Performance Viewer (1 of 4)

In the Tivoli Performance Viewer, you can view graphs and reports that display PMI statistics. This area of the screen shows the navigation tree in Tivoli Performance Viewer where you can select which components to monitor. In this example, the JVM runtime module is selected.

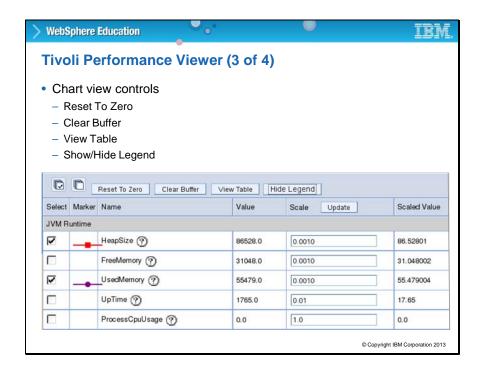
Slide 20



Title: Tivoli Performance Viewer (2 of 4)

Statistics for the selected modules are displayed as a line graph. You can select which metrics you want display in the graph, and you can optionally show the legend. In this example, only the **heap size** and **used memory** metrics are displayed for the JVM run time. The saw tooth pattern of the used memory graph is typical of a steady state JVM. The periodic reductions in used memory correspond to JVM garbage collections, which return unused memory to the heap.

Slide 21

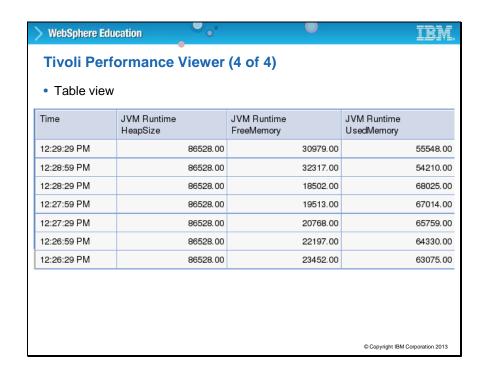


Title: Tivoli Performance Viewer (3 of 4)

The **Reset to zero** button sets a new baseline by using the current counter readings when the button is clicked. Future data points are plotted on the graph relative to their position at the time **Reset to zero** is clicked. Data points that are gathered before the time **Reset to zero** is clicked are not displayed, although they are still held in the Tivoli Performance Viewer buffer. If **Undo Reset to zero** is clicked again, Tivoli Performance Viewer displays all data that is recorded from the original baseline, not from the **Reset to zero** point.

Click the Clear Buffer button to remove the PMI data from a table or chart.

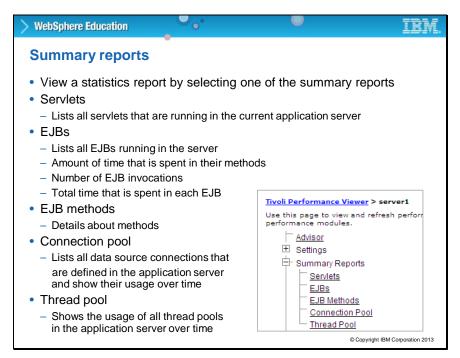
Slide 22



Title: Tivoli Performance Viewer (4 of 4)

Statistics can also be viewed in a table format, by clicking the **View Table** button in the Tivoli Performance Viewer. This screen is an example of what the table view looks like for the JVM runtime module statistics.

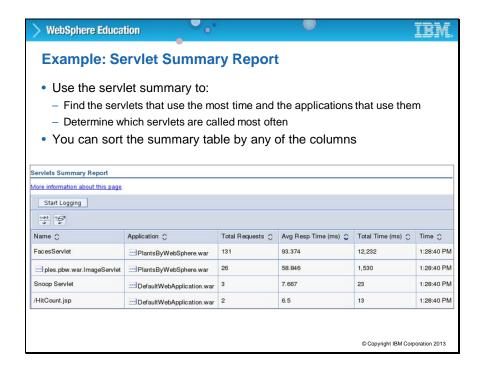
Slide 23



Title: Summary reports

Another option for viewing PMI data is selecting one of the summary reports available. These summary reports collect a number of statistics for a broader group of related components into a report form. Summary reports are available for:

- Servlets
- EJBs
- EJB methods
- Connection pools
- Thread pools



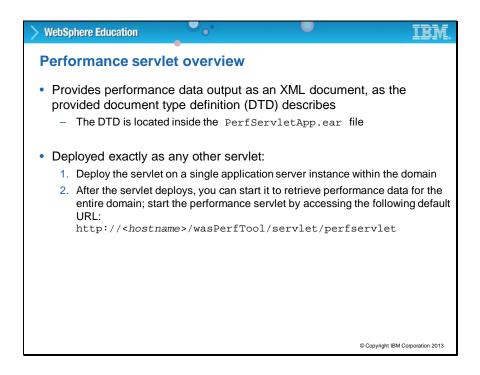
Title: Example: Servlet Summary Report

This screen shows an example of the servlet summary report. The servlet summary lists all servlets that are running in the current application server. Use the servlet summary view to quickly find the most time-intensive servlets and the applications that use them, and to determine which servlets are called most often.

You can sort the summary table by any of the columns. Some tips include:

- Sort by Avg Response Time to find the slowest servlet or JSP page.
- Sort by Total Requests to find the servlet or JSP used the most.
- Sort by Total Time to find the most costly servlet or JSP.

In this screen, the servlets report is shown with the total requests column sorted. You can see which two servlets in the PlantsByWebSphere application are used most frequently.

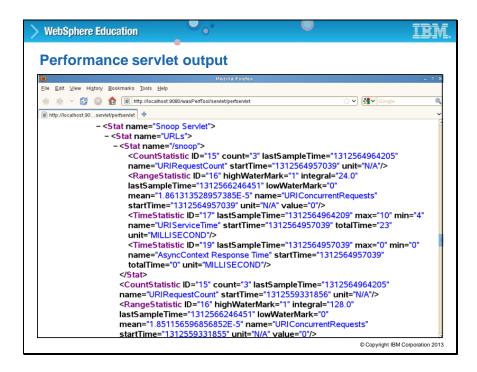


Title: Performance servlet overview

The performance servlet provides a way to use an HTTP request to query the performance metrics for an entire WebSphere Application Server administrative domain. Because the servlet provides the performance data through HTTP, issues such as firewalls, are trivial to resolve. The performance servlet provides the performance data output as an XML document, as described in the provided document type definition (DTD). In the XML structure, the leaves of the structure provide the actual observations of performance data and the paths to the leaves that provide the context.

The performance servlet uses the JMX Perf MBean interface to retrieve the PMI data and outputs an XML document that uses the Java EE Performance Data Framework to describe the statistics. The performance servlet EAR file **PerfServletApp.ear** is in the **WAS_HOME/installableApps** directory, where **WAS_HOME** is the installation path for WebSphere Application Server.

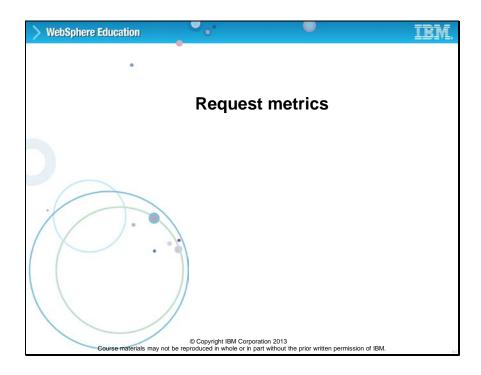
Slide 26



Title: Performance servlet output

This slide shows you the output of a performance servlet request. The snapshot shows the performance statistics about Snoop servlet.

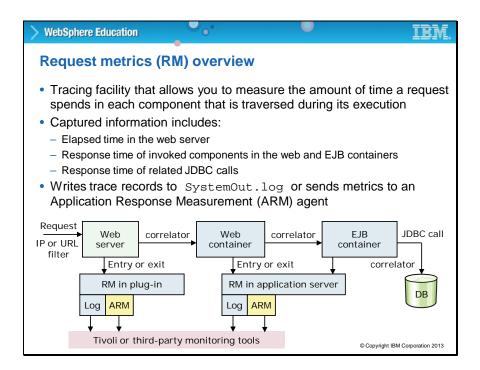
Slide 27



Topic: Request metrics

Request metrics are covered in this topic.

Slide 28

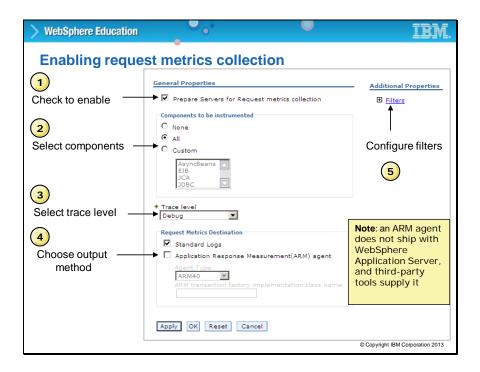


Title: Request metrics (RM) overview

Request metrics allow you to monitor the transaction flow and analyze the response time of the components that are involved in processing it. This analysis can help you target performance problem areas and debug resource constraint problems. For example, it can help determine whether a transaction spends most of its time in the web server plug-in, the web container, the Enterprise JavaBeans (EJB) container, or the back-end database. The response time that is collected for each level includes the time that is spent at that level and the time that is spent in the lower levels. For example, if the total response time for the servlet is 130 milliseconds, and it includes 38 milliseconds from the enterprise beans and JDBC calls, then 92 milliseconds can be attributed to the servlet process.

An ARM agent does not ship with WebSphere Application Server, but third-party tools can provide it.

Slide 29



Title: Enabling request metrics collection

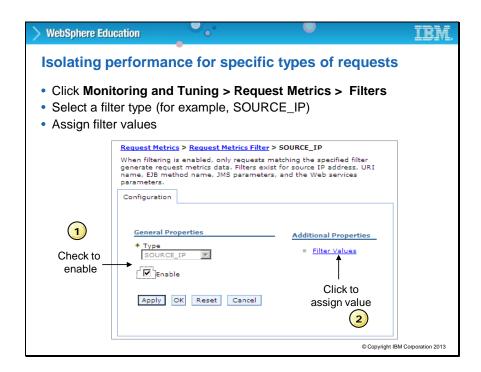
In the administrative console, select **Monitoring and Tuning > Request Metrics** and select the check box to **Prepare Servers for Request metrics collection**.

Trace level specifies how much trace data to accumulate for a particular transaction. **Trace level** and **Components to be instrumented** work together to control whether a request is instrumented or not. The trace level can be set to one of the following values:

- None: No instrumentation.
- Hops: Generates instrumentation information about process boundaries only. When this
 setting is selected, you see the data at the application server level, not the level of
 individual components such as enterprise beans or servlets.
- **Performance_debug**: Generates the data at Hops level and the first level of the intraprocess servlet and Enterprise JavaBeans (EJB) call (for example, when an inbound servlet forwards to a servlet and an inbound EJB calls another EJB). Other intra-process calls, like naming and service integration bus (SIB), are not enabled at this level.
- Debug: Provides detailed instrumentation data, including response times for all intraprocess calls. Note: Requests to servlet filters are only instrumented at this level.
- Standard logs: Enables the request metrics feature for logging. Select this check box to trigger the generation of request metrics logs in the SystemOut.log file. Enabling the request metrics feature for logging increases processor usage; therefore it is suggested to use this feature together with filters so that only selected requests are instrumented.

Application Response Measurement (ARM) agent: Allows request metrics to call an
underlying Application Response Measurement (ARM) agent. Before enabling ARM, you
must install an ARM agent and configure it to the appropriate class path and path,
following the instructions of the ARM provider.

Slide 30



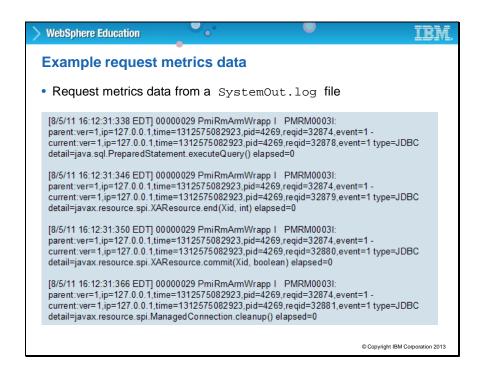
Title: Isolating performance for specific types of requests

The request metrics filters are enabled according to your configuration. For example, if you enabled source IP, only requests whose source IP matches the one specified in the filter are instrumented.

Note: Filters are only checked for edge transactions. An edge transaction is the transaction that first enters an instrumented system. For example, if a servlet calls an EJB component, the servlet is the edge transaction. The servlet must not be instrumented at the web server plug-in, and the URI and SOURCE_IP filters must be checked for the servlet request. However, when the request comes to the EJB container, the EJB filter is not checked because it is no longer an edge transaction.

You must regenerate the web server plug-in configuration file after modifying the request metrics configuration.

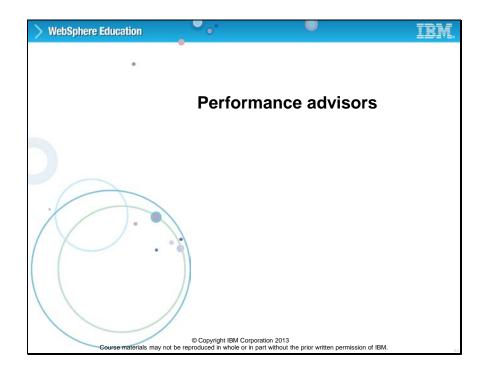
Slide 31



Title: Example of request metrics data

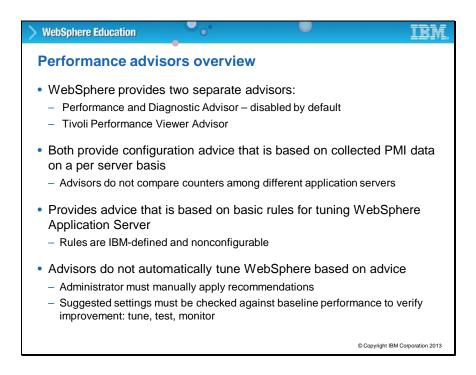
The example of request metrics data that is shown on this slide, shows detail of an SQL call that is being made by using the prepared statement cache. You can trace the steps that are involved, and the timings for this database transaction.

Slide 32



Topic: Performance advisors

This topic describes performance advisors.



Title: Performance advisors overview

The advisors provide advice on the following application server resources:

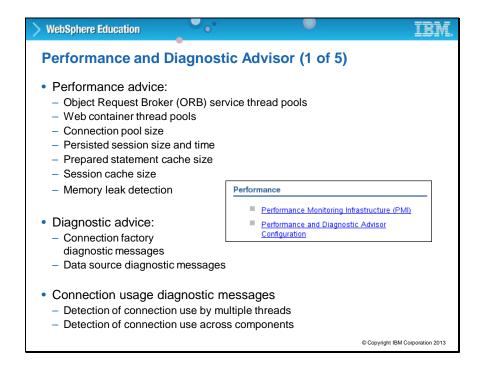
- Thread pools
- Persisted HTTP session sizes
- Cache sizes
- JVM heap size

For example, consider the data source statement cache. It optimizes the processing of *prepared statements* and *callable statements* by caching those statements that are not used in an active connection. (Both statements are SQL statements that essentially run repeatable tasks without the costs of repeated compilation.) If the cache is full, an old entry in the cache is discarded to make room for the new one. The best performance is generally obtained when the cache is large enough to hold all of the statements that are used in the application. The PMI counter called "prepared statement cache discards", indicates the number of statements that are discarded from the cache.

The performance advisors check this counter and provide recommendations to minimize the cache discards.

The advisors can also issue diagnostic advice to help in problem determination and health monitoring. For example, if your application requires more memory than is available, the diagnostic adviser tells you to increase the size of the heap for the application server.

Slide 34

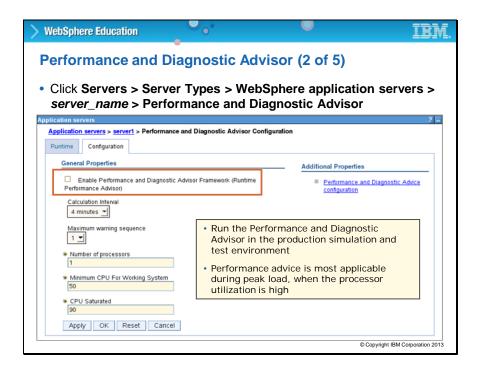


Title: Performance and Diagnostic Advisor (1 of 5)

The Performance and Diagnostic Advisor runs in the Java virtual machine (JVM) process of the application server; therefore, the performance cost is minimal.

To access the Performance and Diagnostic Advisor Configuration, click Servers > Server Types > WebSphere application servers > server_name > Performance and Diagnostic Advisor Configuration.

Slide 35



Title: Performance and Diagnostic Advisor (2 of 5)

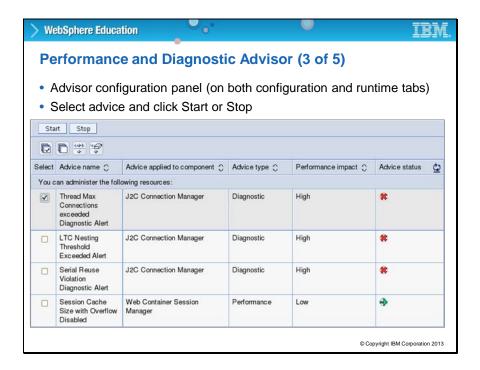
The Performance and Diagnostic Advisor analyzes PMI data and receives notifications about performance and diagnostic information from components. Use this page to specify settings for the Performance and Diagnostic Advisor. Performance issues can be related to memory leaks in the system.

The Performance and Diagnostic Advisor Framework is disabled by default. Each time that you enable it for an application server, you see the warning message:

"Run the Performance and Diagnostic Advisor in the Production Simulation and Test environment."

Performance advice is most applicable during peak load, when the processor use is high.

Slide 36



Title: Performance and Diagnostic Advisor (3 of 5)

Advice type categorizes the primary intent of a piece of advice.

Use Advice type for grouping, and then enable or disable sets of advice that are based on your performance goal. Advice has the following types:

Performance: Performance advice provides tuning recommendations, or identifies problems with your configuration from a performance perspective.

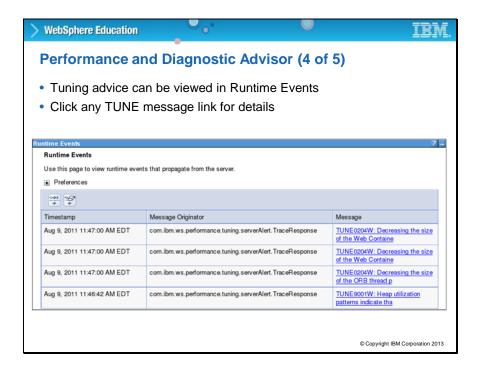
Diagnostic: Diagnostic advice provides automated logic and analysis that relates to problem identification and analysis. These types of advice are issued when the application server encounters unexpected circumstances.

Performance impact generalizes the negative effect on performance that an alert might incur.

The performance impact of a particular piece of advice is highly dependent upon the scenario that is run and upon the conditions that are met. The performance categorization of alerts is based on worst case scenario measurements. The performance categorizations are:

- Low
- Medium
- High

Slide 37



Title: Performance and Diagnostic Advisor (4 of 5)

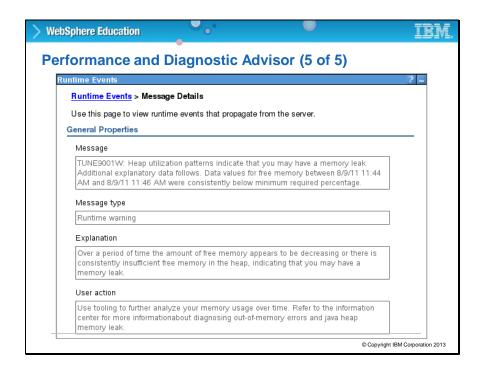
Tuning advice is provided as messages written to the runtime events page. The TUNExxxx messages are typically at the Warning level.

For examples,

TUNE0220W: The Java virtual machine is spending a considerable amount of time in garbage collection. Consider increasing the heap size.

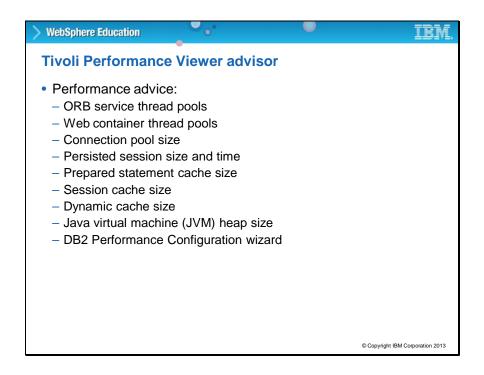
A complete list of TUNE messages is available in the WebSphere Application Server V8.5 information center.

Slide 38



Title: Performance and Diagnostic Advisor (5 of 5)

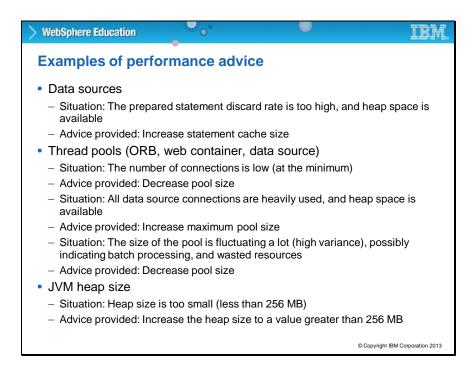
When you select a tune message from the runtime events panel, it displays the detailed information. Notice that you get the message, an explanation, and an action that the user can take. The advisor does not itself make the recommended change; the administrator must apply the advice.



Title: Tivoli Performance Viewer advisor

The performance advisor in Tivoli Performance Viewer provides advice to help tune systems for optimal performance and provide recommendations on inefficient settings by using collected Performance Monitoring Infrastructure (PMI) data. Obtain the advice by selecting the performance advisor in Tivoli Performance Viewer.

In a Network Deployment environment, the performance advisor in Tivoli Performance Viewer runs within the JVM of the node agent and can provide advice on resources that are more expensive to monitor and analyze. In a stand-alone application server environment, the performance advisor in Tivoli Performance Viewer runs within the application server JVM. The Tivoli Performance Viewer advisor requires that you enable performance modules, counters, or both.



Title: Examples of performance advice

Here are some examples of performance advice.

The data source example is using PMI statistics about the rate of prepared statements that are discarded from the cache, which is high. The setting implies that some connections are creating new prepared statements instead of retrieving ones from the cache, which is not a good practice as the application server is creating and removing prepared statements. Also, the PMI statistics are showing that there is enough available memory in the heap. The advice is to increase the size of the prepared statement cache.

Slide 41

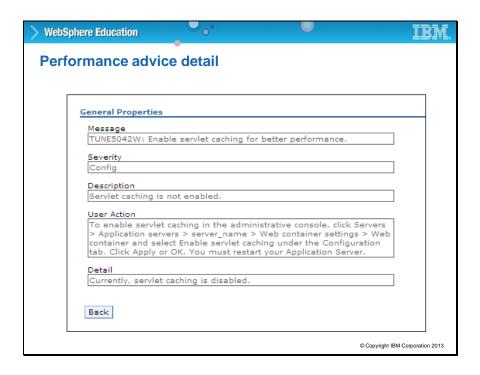


Title: Viewing performance advice

To view advice messages in Tivoli Performance Viewer, click the Advisor link.

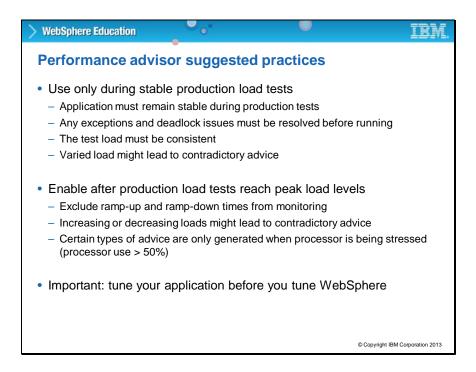
From the list of messages, click a link to see more detail.

Slide 42



Title: Performance advice detail

In this example, the message suggests enabling servlet caching for better performance. Servlet caching is a web container setting that is disabled by default. The User Action section in the advice details provides instructions for enabling servlet caching.



Title: Performance advisor suggested practices

When using the performance advisors, processor usage must rise above 50% before advice is generated. Typically when running your production level load, you push the processor usage to 80–100% before turning on one of the performance advisors.

Consider the following when using a performance advisor for tuning:

If the load changes on the system under test, contradictory advice is generated. This behavior is because the collected PMI data shows a different type of environment, causing the advice to shift. To avoid this situation, always run the advisors while simulating the load WebSphere experiences during deployment (peak load).

If the pool size minimum and maximum values are the same, the performance advisor rules are much more likely to give contradictory advice when load fluctuates.

The amount of processor usage determines the amount of system activity. The advisors do not consider disk activity, network activity, memory usage, or other factors to get a more realistic view of system load.

Recommendations are only generated when processor load reaches 50% and higher.

Performance advisors from different application servers might give contradictory advice on the same node resources. This behavior is because the application servers take into account *only* how they are individually employing the resource. In this situation, if the advice from the different advisors varies greatly, consider the generated advice and decide what changes to make.

ZA855 Unit 20 Transcript

However, if all advisors are giving the same recommendations, then you must seriously consider the suggested changes.

If the performance advisor suggests setting a pool size to X, you must set the minimum value to X/2 and the maximum value to X.

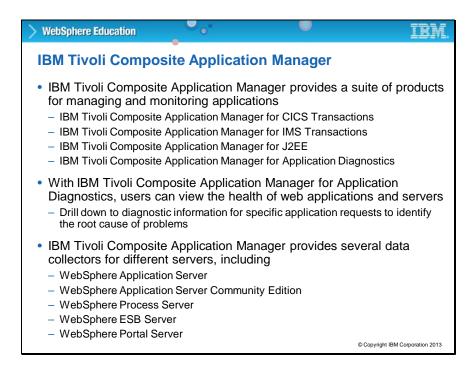
If the performance advisor suggests setting the *prepared statement cache* value to a certain setting, check the amount of memory that is available before using this setting. The advisors do not take into account the amount of actual physical memory available on the system.

Slide 44



Topic: ITCAM for WebSphere Application Server

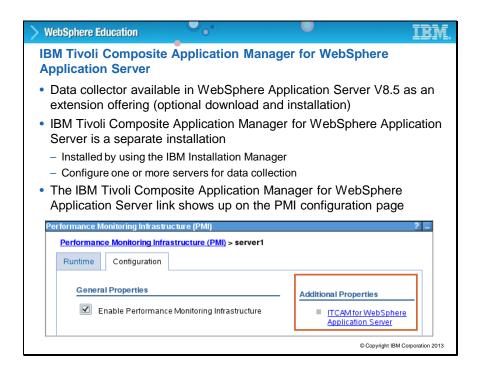
This topic describes ITCAM for WebSphere and how to configure and use it.



Title: IBM Tivoli Composite Application Manager (ITCAM)

IBM Tivoli Composite Application Manager is a suite of products that are used to monitor and manage applications. ITCAM enables users to view the health of applications and servers. ITCAM has several data collectors for different servers.

Slide 46



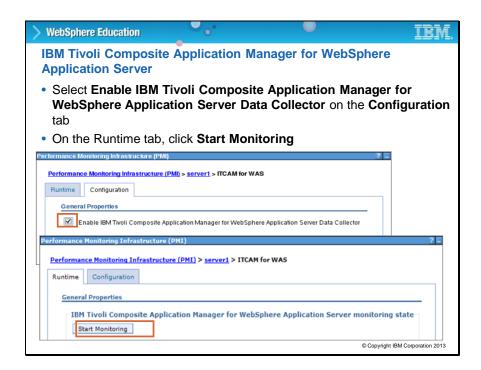
Title: ITCAM for WebSphere Application Server

IBM Tivoli Composite Application Manager (ITCAM) for WebSphere Application Server is enhanced in version 8.0 and can be installed together with the application server. This integrated monitoring tool allows you to view the health of web applications and servers, and drill down to diagnostic information for specific application requests to identify the root cause of problems. ITCAM for WebSphere Application Server can be configured per server by selecting **Monitoring and Tuning > Performance Monitoring Infrastructure > server_name**.

The server must be configured with the ITCAM interface before you can see the ITCAM for WebSphere Application Server link under Additional Properties on the PMI configuration tab of the server.

Use this page to enable or disable the ITCAM for WebSphere Application Server Data Collector. Changes take effect after the server is restarted.

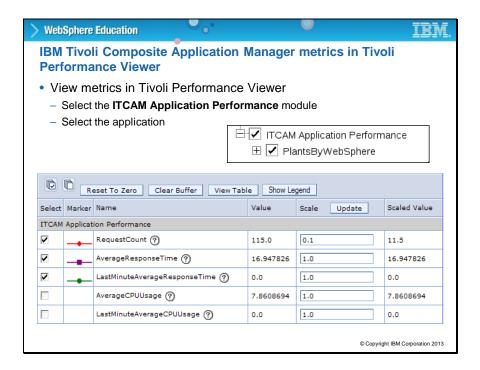
Slide 47



Title: ITCAM for WebSphere Application Server

To see the ITCAM modules and metrics in the Tivoli Performance Viewer, you are required to enable it by selecting the check box on the **Configuration** tab. The next step is to click the **Start Monitoring** button on the runtime tab.

Slide 48



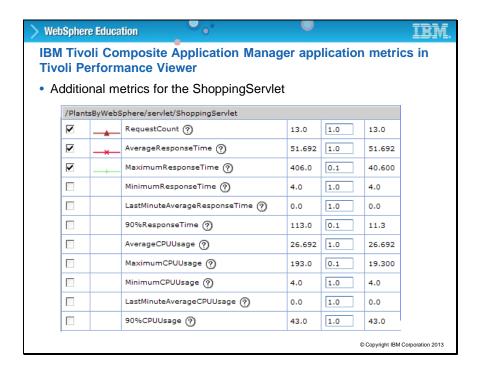
Title: ITCAM metrics in Tivoli Performance Viewer

To view ITCAM monitoring information in Tivoli Performance Viewer, you must enable the counters.

- 1. In the navigation pane, click **Monitoring and Tuning > Performance Viewer > Current Activity**.
- 2. Expand Performance Modules and click ITCAM Application Performance.
- 3. To refresh the view, click **Tivoli Performance Viewer** and select the *application_server* instance for which you want to view performance data.

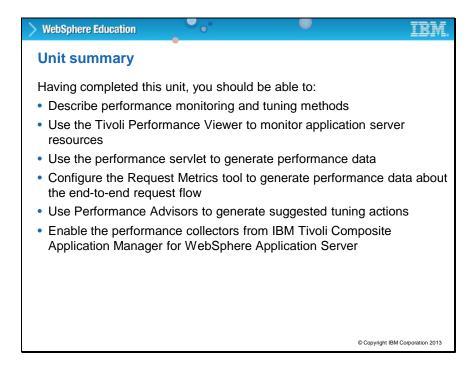
ITCAM for WebSphere Application Server provides more request-based response time and processor metrics.

Slide 49



Title: ITCAM application metrics in Tivoli Performance Viewer

Several performance metrics are collected and displayed for each component of an application. This screen shows the metrics for the PlantsByWebSphere shopping servlet after a load test. Clicking or hovering over the question mark (?) for each metric displays a description of the metric.



Title: Unit summary

Having completed this unit, you should be able to:

- · Describe performance monitoring and tuning methods
- Use the Tivoli Performance Viewer to monitor application server resources
- Use the performance servlet to generate performance data
- Configure the Request Metrics tool to generate performance data about the end-to-end request flow
- Use Performance Advisors to generate tuning actions
- Enable the performance collectors from IBM Tivoli Composite Application Manager for WebSphere Application Server