

Note

Before using this information, be sure to read the general information under “Notices” on page 125.

Compilation date: June 30, 2005

© Copyright International Business Machines Corporation 2005. All rights reserved.

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

| | | | |
|--|-----------|--|-----|
| How to send your comments | v | Tuning TCP/IP buffer sizes | 28 |
| Chapter 1. How do I tune performance? | 1 | Tuning Java virtual machines | 29 |
| Chapter 2. Planning ahead for performance. | 5 | Tuning transport channel services | 35 |
| Application design consideration | 5 | Checking hardware configuration and settings. | 38 |
| Chapter 3. Taking advantage of performance functions | 9 | Tuning operating systems | 40 |
| Chapter 4. Obtaining performance advice from the performance advisors | 11 | Tuning Windows systems. | 40 |
| Why use the performance advisors? | 11 | Tuning Linux systems | 43 |
| Choosing the right performance advisor for the purpose | 12 | Tuning AIX systems | 45 |
| Runtime Performance Advisor | 12 | Tuning Solaris systems | 48 |
| Tivoli Performance Viewer advisor | 13 | Tuning HP-UX systems | 51 |
| Using the Runtime Performance Advisor | 14 | Tuning Web servers. | 54 |
| Runtime Performance Advisor configuration settings. | 16 | Tuning WebSphere applications. | 56 |
| Restart | 16 | Web applications | 58 |
| Enable automatic heap dump collection | 16 | Tuning session management. | 58 |
| Enable Runtime Performance Advisor | 16 | Web services | 65 |
| Calculation Interval. | 17 | Tuning Web services security for version 6.0.x applications | 65 |
| Maximum warning sequence | 17 | Tuning Web services security for version 5.x applications | 66 |
| Number of processors | 17 | Monitoring the performance of Web services applications | 66 |
| Advice configuration settings | 17 | Service integration | 69 |
| Advice name | 17 | Setting tuning properties of a messaging engine | 69 |
| Advice applied to component | 17 | Tuning and problem solving for messaging engine data stores | 69 |
| Advice status. | 17 | Setting tuning properties for a mediation | 73 |
| Advice status. | 17 | Enabling CMP entity beans and messaging engine data stores to share database connections | 74 |
| Viewing the Runtime Performance Advisor recommendations | 17 | Tuning service integration messaging. | 76 |
| Understanding the lightweight memory leak detection | 18 | Tuning the SIBWS | 79 |
| Enabling the lightweight memory leak detection | 19 | Data access resources | 81 |
| Getting the automated heap dump generation support. | 19 | EJB Container Cache tuning | 81 |
| Enabling automated heap dump generation | 20 | Database performance tuning | 84 |
| Generating heap dumps manually. | 20 | DB2 tuning parameters | 85 |
| Locating and analyzing heap dumps | 21 | Tuning parameters for data access resources | 88 |
| Using the performance advisor in Tivoli Performance Viewer | 22 | Messaging. | 90 |
| Performance advisor report in Tivoli Performance Viewer | 24 | Setting tuning properties of a messaging engine | 90 |
| Message | 24 | Tuning and problem solving for messaging engine data stores | 91 |
| Performance data in the upper panel | 24 | Setting tuning properties for a mediation | 94 |
| Chapter 5. Tuning the application serving environment. | 25 | Enabling CMP entity beans and messaging engine data stores to share database connections | 95 |
| Tuning parameter hot list. | 26 | Tuning service integration messaging. | 97 |
| | | Tuning the SIBWS | 100 |
| | | Security | 102 |
| | | Tuning security configurations. | 102 |
| | | Object Request Broker | 109 |
| | | Object Request Broker tuning guidelines | 109 |
| | | Learn about WebSphere programming extensions | 114 |
| | | Dynamic cache | 114 |
| | | Work area | 118 |

**Chapter 6. Troubleshooting
performance 121**

Notices 125

Trademarks and service marks 127

How to send your comments

Your feedback is important in helping to provide the most accurate and highest quality information.

- To send comments on articles in the WebSphere Application Server Information Center
 1. Display the article in your Web browser and scroll to the end of the article.
 2. Click on the **Feedback** link at the bottom of the article, and a separate window containing an e-mail form appears.
 3. Fill out the e-mail form as instructed, and click on **Submit feedback** .
- To send comments on PDF books, you can e-mail your comments to: **wasdoc@us.ibm.com** or fax them to 919-254-0206.






Be sure to include the document name and number, the WebSphere Application Server version you are using, and, if applicable, the specific page, table, or figure number on which you are commenting.

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Chapter 1. How do I tune performance?



Legend for "How do I?..." links

| Documentation  | Show me  | Tell me  | Guide me  | Teach me  |
|--|--|--|---|---|
| Refer to the detailed steps and reference | Watch a brief multimedia demonstration | View the presentation for an overview | Be led through the console pages | Perform the tutorial with sample code |
| Approximate time: Varies | Approximate time: 3 to 5 minutes | Approximate time: 10 minutes+ | Approximate time: 1/2 hour+ | Approximate time: 1 hour+ |



Tune the application serving environment

Each WebSphere Application Server process has several parameters influencing application performance.

Documentation



Obtain tuning advice

The Performance Advisor in Tivoli Performance Viewer (TPV) provides advice to help tune systems for optimal performance and gives recommendations on inefficient settings by using collected Performance Monitoring Infrastructure (PMI) data.

Documentation



Show me

- Performance advisor
- TPV advisor

Tell me



Tune WebSphere applications

Expand **Tuning performance > Tuning WebSphere applications** in the information center table of contents for information about tuning specific application components and their administrative configurations (such as EJB container settings).



Pool database connections

As described in Connection pooling, when accessing any database, establishing connections is an expensive operation. Connection pooling enables administrators to establish a pool of database connections that applications can share on an application server. When connection pooling capabilities are used, performance improvements up to 20 times the normal results are realized.

Documentation



Configure caching

The dynamic cache service improves performance by caching the output of servlets, commands, and JSP files.

Documentation



Show me



Tell me



Related concepts

Overview and new features for tuning performance

Product overview and quick start

This topic provides shortcuts to information for obtaining a high level understanding of the product, then getting started quickly. Links include various conceptual overviews, descriptions of what is new, and how the documentation is organized. "How do I?" links lead to task descriptions, including ShowMe demonstrations, tutorials, and presentations when they are available.

Chapter 2. Planning ahead for performance

How well a Web site performs while receiving heavy user traffic is an essential factor in the overall success of an organization. This section provides online resources that you can consult to ensure that your site performs well under pressure.

- Consult the following Web resources for learning.

IBM Patterns for e-Business

IBM Patterns for e-business is a group of reusable assets that can help speed the process of developing Web-based applications. The patterns leverage the experience of IBM architects to create solutions quickly, whether for a small local business or a large multinational enterprise.

Planning for availability in the enterprise

Availability is an achievable service-level characteristic that every enterprise struggles with. The worst case scenario is realized when load is underestimated or bandwidth is swamped because availability planning was not carefully conducted. Applying the information in this article and the accompanying spreadsheet to your planning exercises can help you avoid such a scenario.

A concise guide to WebSphere capacity management

How much capacity is enough for your Web site? It is a difficult question, and often customers have either too much or too little capacity to meet the demands of their site visitors. When developing a plan for your Web site, you must first define its service levels. Next, you must determine the health of your existing site and apply what you know to planning for your future site. Finally, you must validate your capacity estimates through testing and then eventually deploy the site. Here, developing and implementing a capacity plan for your site is explored.

Hardware configurations for WebSphere Application Server production environments

This article describes the most common production hardware configurations, and provides the reasons for choosing each one. It begins with a single machine configuration, and then proceeds with additional configurations that have higher fault tolerance, horizontal scaling, and a separation of Web and enterprise bean servers.

- See the documentation for the product functionality to improve performance .

Application design consideration

This topic describes the architectural suggestions in design and how to tune applications.

Consult the Designing applications topic, which highlights Web sites and other ideas for finding best practices for designing WebSphere applications, particularly in the realm of WebSphere extensions to the Java 2 Platform, Enterprise Edition (J2EE) specification.

The Designing applications topic contains the architectural suggestions in design and the implementation of applications. For existing applications, the suggestions

might require changing the existing implementations. Tuning the application server and resource parameters can have the greatest effect on performance of the applications that are well designed.

Use the following information as an architectural guide when implementing applications:

- Persistence
- Model-view-controller pattern
- Statelessness
- Caching
- Asynchronous considerations
- Third-party libraries

Persistence

Java 2 Platform, Enterprise Edition (J2EE) applications load, store, create, and remove data from relational databases, a process commonly referred to as *persistence*. Most enterprise applications have significant database access. The architecture and performance of the persistence layer is critical to the performance of an application. Therefore, persistence is a very important area to consider when making architectural choices that require trade-offs related to performance. This guide recommends first focusing on a solution that has clean architecture. The clean architecture considers data consistency, security, maintenance, portability, and the performance of that solution. Although this approach might not yield the absolute peak performance obtainable from manual coding a solution that ignores the mentioned qualities of service, this approach can achieve the appropriate balance of data consistency, maintainability, portability, security, and performance.

Multiple options are available in J2EE for persistence: Session beans using entity beans including container-managed persistence (CMP) or bean-managed persistence (BMP), session beans using Java Database Connectivity (JDBC), and Java beans using JDBC. For the reasons previously mentioned, consider CMP entity persistence because it provides maximum security, maintenance, and portability. CMP is also recommended for good performance. Refer to the Tune the EJB container section of the Tuning application servers topic on tuning enterprise beans and more specifically, CMP.

If an application requires using enterprise beans not using EJB entities, the persistence mechanism usually involves the JDBC API. Because JDBC requires manual coding, the Structured Query Language (SQL) that runs against a database instance, it is critical to optimize the SQL statements that are used within the application. Also, configure the database server to support the optimal performance of these SQL statements. Finally, usage of specific JDBC APIs must be considered including prepared statements and batching.

Regardless of which persistence mechanism is considered, use container-managed transactions where the bean delegates management of transactions to the container. For applications that use JDBC, this is easily achieved by using the session façade pattern, which wraps all JDBC functions with a stateless session bean.

Finally, information about tuning the connection over which the EJB entity beans or JDBC communicates can be found in the Tune the data sources section of the Tuning application servers topic.

Model-view-controller pattern

One of the standard J2EE programming architectures is the model-view-controller (MVC) architecture, where a call to a controller servlet might include one or more child JavaServer Pages (JSP) files to construct the view. The MVC pattern is a recommended pattern for application architecture. This pattern requires distinct separation of the view (JSP files or presentation logic), the controller (servlets), and the model (business logic). Using the MVC pattern enables optimization of the performance and scalability of each layer separately.

Statelessness

Implementations that avoid storing the client user state scale and perform the best. Design implementations to avoid storing state. If state storage is needed, ensure that the size of the state data and the time that the state is stored are kept to the smallest possible values. Also, if state storage is needed, consider the possibility of reconstructing the state if a failure occurs, instead of guaranteeing state failover through replication.

Specific tuning of state affects HTTP session state, dynamic caching, and enterprise beans. Refer to the follow tuning guides for tuning the size, replication, and timing of the state storage:

- “Tuning session management” on page 58
- “EJB Container Cache tuning” on page 81
- “Tuning dynamic cache with the cache monitor” on page 116

Caching

Most J2EE application workloads have more read operations than write operations. Read operations require passing a request through several topology levels that consist of a front-end Web server, the Web container of an application server, the EJB container of an application server, and a database. WebSphere Application Server provides the ability to cache results at all levels of the network topology and J2EE programming model that include Web services.

Application designers must consider caching when the application architecture is designed because caching integrates at most levels of the programming model. Caching is another reason to enforce the MVC pattern in applications. Combining caching and MVC can provide caching independent of the presentation technology and in cases where there is no presentation to the clients of the application.

Network designers must consider caching when network planning is performed because caching also integrates at most levels of the network topology. For applications that are available on the public Internet, network designers might want to consider Edge Side Include (ESI) caching when WebSphere Application Server caching extends into the public Internet. Network caching services are available in the proxy server for WebSphere Application Server, WebSphere Edge Component Caching Proxy, and the WebSphere plug-in.

Asynchronous considerations

J2EE workloads typically consist of two types of operations. You must perform the first type of operation to respond to a system request. You can perform the second type of operation asynchronously after the user request that initiated the operation is fulfilled.

An example of this difference is an application that enables you to submit a purchase order, enables you to continue while the system validates the order, queries remote systems, and in the future informs you of the purchase order status. This example can be implemented synchronously with the client waiting for the response. The synchronous implementation requires application server resources and you wait until the entire operations complete. If the process enables you to continue, while the result is computed asynchronously, the application server can schedule the processing to occur when it is optimal in relation to other requests. The notification to you can be triggered through e-mail or some other interface within the application.

Because the asynchronous approach supports optimal scheduling of workloads and minimal server resource, consider asynchronous architectures. WebSphere Application Server supports asynchronous programming through J2EE Java Message Service (JMS) and message-driven beans (MDB) as well as asynchronous beans that are explained in the Tuning Java Message Service and Tuning MDB topics.

Third-party libraries

Verify that all the libraries that applications use are also designed for server-side performance. Some libraries are designed to work well within a client application and fail to consider server-side performance concerns, for example, memory utilization, synchronization, and pooling. It is suggested that all libraries that are not developed as part of an application undergo performance testing using the same test methodologies as used for the application.

Additional reference:

IBM WebSphere Developer Technical Journal: The top 10 (more or less) J2EE best practices

Improve performance in your XML applications, Part 2

Related tasks

Chapter 5, "Tuning the application serving environment," on page 25
This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

Chapter 3. Taking advantage of performance functions

This topic highlights a few main ways you can improve performance through a combination of product features and application development considerations.

- Use this product functionality to improve performance.

Balancing workloads with clusters

Clusters are sets of servers that are managed together and participate in workload management. The servers that are members of a cluster can be on different host machines, as opposed to the servers that are part of the same node and must be located on the same host machine. A cell can have no clusters, one cluster, or multiple clusters.

Task overview: Using the dynamic cache service to improve performance

The dynamic cache service improves performance by caching the output of servlets, commands, and JavaServer Pages (JSP) files. Dynamic caching features include cache replication among clusters, cache disk offload, Edge-side include caching, and external caching, which is the ability to control caches outside of the application server, such as that of your Web server.

- Ensure your applications perform well.
Details are available in the following topics:
 - “Application design consideration” on page 5 (architectural suggestions)
 - Designing applications (coding best practices)

Chapter 4. Obtaining performance advice from the performance advisors

Tuning WebSphere Application Server is a critical part of getting the best performance from your Web site. 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, analyze the performance data and provide configuration recommendations to improve the application server performance. Therefore, the performance advisors provide a starting point to the application server tuning process and help you without requiring that you become an expert.

Why use the performance advisors?

The performance advisors analyze the Performance Monitoring Infrastructure (PMI) data of WebSphere Application Server using general performance principles, best practices, and WebSphere-specific rules for tuning. The performance advisors that are based on this information provide advice on how to set some of your configuration parameters to better tune WebSphere Application Server.

The performance advisors provide a variety of advice on the following application server resources:

- Object Request Broker service thread pools
- Web container thread pools
- Connection pool size
- Persisted session size and time
- Data source statement cache size
- Session cache size
- Dynamic cache size
- Java virtual machine heap size
- DB2 Performance Configuration Wizard

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 being used in an active connection. (Both 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 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.

Using another example with pools in the application server, the idea behind pooling is to use an existing thread or connection from the pool instead of creating a new instance for each request. Because each thread or connection in the pool consumes memory and increases the context-switching cost, the pool size is an important configuration parameter. A pool that is too large can hurt performance

as much as a pool that is too small. The performance advisors use PMI information about current pool usage, minimum or maximum pool size, and the application server CPU utilization to recommend efficient values for the pool sizes.

Related tasks

Chapter 4, “Obtaining performance advice from the performance advisors,” on page 11

Choosing the right performance advisor for the purpose

Two performance advisors are available: the Runtime Performance Advisor and the performance advisor in Tivoli Performance Viewer (TPV advisor). The Runtime Performance Advisor runs in the Java virtual machine (JVM) process of Application Server; therefore, it does not provide expensive advice. In a Network Deployment environment, the performance advisor in TPV runs on the node agent and can provide advice on resources which are more expensive to monitor and analyze. The TPV advisor requires that you enable performance modules, counters, or both.

The following chart shows the differences between the Runtime Performance Advisor and the TPV advisor:

| | Runtime Performance Advisor | TPV advisor |
|------------------------|---|--|
| Start location | Application server | TPV client |
| Invocation of tool | Administrative console | TPV |
| Output | The SystemOut.log file and the administrative console | TPV in the administrative console |
| Frequency of operation | Configurable | When you select refresh in the TPV administrative console |
| Types of advice | <ul style="list-style-type: none"> • Object Request Broker (ORB) service thread pools • Web container thread pools • Connection pool size • Persisted session size and time • Prepared statement cache size • Session cache size • Memory leak detection | <ul style="list-style-type: none"> • ORB service thread pools • Web container thread pools • Connection pool size • Persisted session size and time • Prepared statement cache size • Session cache size • Dynamic cache size • Java virtual machine (JVM) heap size • DB2 Performance Configuration wizard |

Related tasks

Chapter 4, “Obtaining performance advice from the performance advisors,” on page 11

Runtime Performance Advisor

The Runtime Performance Advisor provides advice to help tune systems for optimal performance and is configured using the WebSphere Application Server administrative console or the wsadmin tool. Running in the Java virtual machine (JVM) of the application server, the Runtime Performance Advisor periodically checks for inefficient settings and issues recommendations as standard product warning messages. These recommendations are displayed both as warnings in the

administrative console under WebSphere Runtime Messages in the WebSphere Status panel and as text in the application server SystemOut.log file. Enabling the Runtime Performance Advisor has minimal system performance impact.

In WebSphere Application Server, Version 6.0.2, you can use the Runtime Performance Advisor to enable the lightweight memory leak detection, which is designed to provide early detection of memory problems in test and production environments.

The advice that the Runtime Performance Advisor gives is all on the server level. The only difference when running in a Network Deployment environment is that you might receive contradictory advice on resources that are declared at the node or cell level and used at the server level.

For example, two sets of advice are given if a data source is declared at the node level to have a connection pool size of {10,50} and is used by two servers (server1 and server2). If server1 uses only two connections and server2 uses all fifty connections during peak load, the optimal connection pool size is different for the two servers. Therefore, the Runtime Performance Advisor gives two sets of advice (one for server1 and another for server2). The data source is declared at the node level and you must make your decisions appropriately by setting one size that works for both, or by declaring two different data sources for each server with the appropriate level.

Related concepts

“Choosing the right performance advisor for the purpose” on page 12

Two performance advisors are available: the Runtime Performance Advisor and the performance advisor in Tivoli Performance Viewer (TPV advisor). The Runtime Performance Advisor runs in the Java virtual machine (JVM) process of Application Server; therefore, it does not provide expensive advice. In a Network Deployment environment, the performance advisor in TPV runs on the node agent and can provide advice on resources which are more expensive to monitor and analyze. The TPV advisor requires that you enable performance modules, counters, or both.

“Tivoli Performance Viewer advisor”

Tivoli Performance Viewer advisor

The performance advisor in Tivoli Performance Viewer (TPV) 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 TPV.

Related concepts

“Choosing the right performance advisor for the purpose” on page 12

Two performance advisors are available: the Runtime Performance Advisor and the performance advisor in Tivoli Performance Viewer (TPV advisor). The Runtime Performance Advisor runs in the Java virtual machine (JVM) process of Application Server; therefore, it does not provide expensive advice. In a Network Deployment environment, the performance advisor in TPV runs on the node agent and can provide advice on resources which are more expensive to monitor and analyze. The TPV advisor requires that you enable performance modules, counters, or both.

“Runtime Performance Advisor” on page 12

The Runtime Performance Advisor provides advice to help tune systems for optimal performance and is configured using the WebSphere Application Server

administrative console or the wsadmin tool. Running in the Java virtual machine (JVM) of the application server, the Runtime Performance Advisor periodically checks for inefficient settings and issues recommendations as standard product warning messages. These recommendations are displayed both as warnings in the administrative console under WebSphere Runtime Messages in the WebSphere Status panel and as text in the application server SystemOut.log file. Enabling the Runtime Performance Advisor has minimal system performance impact.

Using the Runtime Performance Advisor

The Runtime Performance Advisor provides advice to help tune systems for optimal performance and is configured using the WebSphere Application Server administrative console or the wsadmin tool . The Runtime Performance Advisor uses Performance Monitoring Infrastructure (PMI) data to provide recommendations for performance tuning. Running in the Java Virtual Machine (JVM) of the application server, this advisor periodically checks for inefficient settings, and issues recommendations as standard product warning messages. View these recommendations by clicking **Troubleshooting > Runtime Messages > Warning** in the administrative console. Enabling the Runtime Performance Advisor has minimal system performance impact.

1. Enable PMI in the application server administrative console as described in the Enabling PMI using the administrative console article. To obtain advice, you must first enable PMI through the administrative console and restart the server. The Runtime Performance Advisor enables the appropriate monitoring counter levels for all enabled advice when PMI is enabled. If specific counters exist that are not wanted, disable the corresponding advice in the Runtime Performance Advisor panel, and disable unwanted counters. If specific counters exist that are not wanted, or when disabling the Runtime Performance Advisor, you might want to disable PMI or the counters that the Runtime Performance Advisor enabled.
2. If running Network Deployment, you must enable PMI on both the server and the node agent, and restart the server and the node agent.
3. Click **Servers > Application servers** in the console navigation tree.
4. Click *server_name* > **Runtime Performance Advisor Configuration**.
5. Under the Configuration tab, specify the number of processors on the server. This setting is critical to ensure accurate advice for the specific configuration of the system.
6. Select the **Calculation Interval**. PMI data is taken over an interval of time and averaged to provide advice. The calculation interval specifies the length of the time over which data is taken for this advice. Therefore, details within the advice messages display as averages over this interval.
7. Select the **Maximum Warning Sequence**. The maximum warning sequence refers to the number of consecutive warnings that are issued before the threshold is updated. For example, if the maximum warning sequence is set to 3, then the advisor only sends three warnings to indicate that the prepared statement cache is overflowing. After three warnings, a new alert is only issued if the rate of discards exceeds the new threshold setting.
8. Click **Apply**.
9. Click **Save**.
10. Click the **Runtime** tab.

11. Click **Restart**. Select **Restart** on the Runtime tab to re-initialize the Runtime Performance Advisor using the last configuration information that is saved to disk.

Note: This action also resets the state of the Runtime Performance Advisor. For example, the current warning count is reset to zero (0) for each message.

12. Simulate a production level load. If you use the Runtime Performance Advisor in a test environment, do any other tuning for performance, or simulate a realistic production load for your application. The application must run this load without errors. This simulation includes numbers of concurrent users typical of peak periods, and drives system resources, for example, CPU and memory, to the levels that are expected in production. The Runtime Performance Advisor only provides advice when CPU utilization exceeds a sufficiently high level. For a list of IBM business partners providing tools to drive this type of load, see the article, Performance: Resources for learning in the subsection of Monitoring performance with third party tools.
13. Select the check box to enable the Runtime Performance Advisor.

Note: To achieve the best results for performance tuning, enable the Runtime Performance Advisor when a stable production level load is being applied.

14. Click **OK**.
15. Select **Warnings** in the administrative console under the WebSphere Runtime Messages in the WebSphere Status panel or look in the `SystemOut.log` file, which is located in the `install_root/profiles/profile_name/logs/server_name` directory to view tuning advice. Some messages are not issued immediately.
16. Update the product configuration for improved performance, based on advice. Although the performance advisors attempt to distinguish between loaded and idle conditions, misleading advice might be issued if the advisor is enabled while the system is ramping up or down. This result is especially likely when running short tests. Although the advice helps in most configurations, there might be situations where the advice hinders performance. Because of these conditions, advice is not guaranteed. Therefore, test the environment with the updated configuration to ensure that it functions and performs better than the previous configuration.

Over time, the advisor might issue differing advice. The differing advice is due to load fluctuations and the run-time state. When differing advice is received, you need to look at all advice and the time period over which it is issued. Advice is taken during the time that most closely represents the peak production load.

Performance tuning is an iterative process. After applying advice, simulate a production load, update the configuration that is based on the advice, and retest for improved performance. This procedure is continued until optimal performance is achieved.

You can enable and disable advice in the Advice Configuration panel. Some advice applies only to certain configurations, and can only be enabled for those configurations. For example, unbounded Object Request Broker (ORB) service thread pool advice is only relevant when the ORB service thread pool is unbounded, and can only be enabled when the ORB thread pool is unbounded. For more information on Advice configuration, see the article, "Advice configuration settings" on page 17.

Related tasks

Chapter 4, “Obtaining performance advice from the performance advisors,” on page 11

Runtime Performance Advisor configuration settings

Use this page to specify settings for the Runtime Performance Advisor.

For more information on how to use the Runtime Performance Advisor, see the article, “Using the Runtime Performance Advisor” on page 14.

To view this administrative page, click **Servers > Application Servers > *server_name* > Runtime Performance Advisor Configuration** under the Performance section.

Restart

Selecting Restart on the Runtime tab reinitializes the Runtime Performance Advisor using the last information saved to disk. Note that this action also resets the state of the Runtime Performance Advisor. For example, the current warning sequence is reset to zero for each recommendation or advice.

Enable automatic heap dump collection

Specifies whether the Runtime Performance Advisor automatically generates heap dumps for post analysis when suspicious memory activity is detected.

Enable Runtime Performance Advisor

Specifies whether the Runtime Performance Advisor runs on the server startup.

The Runtime Performance Advisor requires that the Performance Monitoring Infrastructure (PMI) be enabled. It does not require that individual counters be enabled. When a counter that is needed by the Runtime Performance Advisor is not enabled, the Runtime Performance Advisor enables it automatically. When disabling the Runtime Performance Advisor, you might want to disable Performance Monitoring Infrastructure (PMI) or the counters that Runtime Performance Advisor enabled. The following counters might be enabled by the Runtime Performance Advisor:

- ThreadPools (module)
 - Web Container (module)
 - Pool Size
 - Active Threads
 - Object Request Broker (module)
 - Pool Size
 - Active Threads
- JDBC Connection Pools (module)
 - Pool Size
 - Percent used
 - Prepared Statement Discards
- Servlet Session Manager (module)
 - External Read Size
 - External Write Size
 - External Read Time
 - External Write Time
 - No Room For New Session
- System Data (module)
 - CPU Utilization
 - Free Memory

Calculation Interval

Specifies the length of time over which data is taken for this advice.

PMI data is taken over an interval of time and averaged to provide advice. The calculation interval specifies the length of time over which data is taken for this advice. Details within the advice messages display as averages over this interval.

Maximum warning sequence

The maximum warning sequence refers to the number of consecutive warnings that are issued before the threshold is relaxed.

For example, if the maximum warning sequence is set to 3, the advisor only sends three warnings to indicate that the prepared statement cache is overflowing. After three warnings, a new alert is only issued if the rate of discards exceeds the new threshold setting.

Number of processors

Specifies the number of processors on the server.

This setting is critical to ensure accurate advice for the specific configuration of the system.

Advice configuration settings

Use this page to select the advice you wish to enable or disable.

To view this administrative page, click **Servers > Application Servers > *server_name* > Runtime Performance Advisor Configuration** under the Performance section > **Advice Configuration**.

Advice name

Specifies the advice that you can enable or disable.

Advice applied to component

Specifies the WebSphere Application Server component to which the run-time performance advice applies.

Advice status

Specifies whether the advice is stopped or started.

Specify one of two values: **Started** and **Stopped**. **Started** means that the advice runs if the advice applies. **Stopped** means that the advice does not run.

Advice status

Specifies whether the advice is stopped, started or unavailable.

The advice status has one of three values: **Started**, **Stopped** or **Unavailable**. **Started** means that the advice is applied. **Stopped** means that the advice is not applied. **Unavailable** means that the advice does not apply to the current configuration, for example, persisted session size advice in a configuration without persistent sessions.

Viewing the Runtime Performance Advisor recommendations

The Runtime Performance Advisor recommendations are displayed at the following locations:

- The WebSphere Application Server log file SystemOut.log. The following log file is a sample output of advice on the SystemOut.log file:

```
[4/2/04 15:50:26:406 EST] 6a83e321 TraceResponse W TUNE0202W:  
Increasing the Web Container thread pool's Maximum Size to 48  
may improve performance.
```

Additional explanatory data follows.

Average number of threads: 48.

Configured maximum pool size: 2.

This alert has been issued 1 time(s) in a row.
The threshold will be updated to reduce the
overhead of the analysis.

- The Runtime Messages panel in the administrative console. To view this administrative page, click **Troubleshooting > Runtime Messages > Warning**.

Related tasks

“Using the Runtime Performance Advisor” on page 14

Understanding the lightweight memory leak detection

This topic describes memory leaks in Java applications and introduces a new function, lightweight memory leak detection, in WebSphere Application Server, Version 6.0.2.

Memory leaks in Java applications

Although a Java application has a built-in garbage collection mechanism, which frees the programmer from any explicit object deallocation responsibilities, memory leaks are still common in Java applications. Memory leaks occur in Java applications when unintentional references are made to unused objects. This occurrence prevents Java garbage collection from freeing memory.

The term *memory leak* is overused; a memory leak refers to a memory misuse or mismanagement. Old unused data structures might have outstanding references but are never garbage collected. A data structure might have unbounded growth or there might not be enough memory that is allocated to efficiently run a set of applications.

Lightweight memory leak detection in WebSphere Application Server

Most existing memory leak technologies are based upon the idea that you know that you have a memory leak and want to find it. Because of these analysis requirements, these technologies have significant performance burdens and are not designed for use as a detection mechanism in production. This limitation means that memory leaks are generally not detected until the problem is critical; the application passes all system tests and is put in production, but it crashes and nobody knows why.

WebSphere Application Server has implemented a lightweight memory leak detection mechanism that runs within the WebSphere Runtime Performance Advisor framework. This mechanism is designed to provide early detection of memory problems in test and production environments. This framework is not designed to provide analysis of the source of the problem, but rather to provide notification and help generating the information that is required to use analysis tools. The mechanism only detects memory leaks in the Java heap and does not detect native leaks.

The lightweight memory leak detection in WebSphere Application Server does not require any additional agents. The detection relies on algorithms that are based on information that is available from the Performance Monitoring Infrastructure service and has minimal performance overhead.

Enabling the lightweight memory leak detection

Use this task to enable the lightweight memory leak detection using the Runtime Performance Advisor.

If you have a memory leak and want to confirm the leak, or you want to automatically generate heap dumps on Java virtual machines (JVM) in WebSphere Application Server, consider changing your minimum and maximum heap sizes to be equal. This change provides the memory leak detection more time for reliable diagnosis.

To enable the lightweight memory leak detection using the Runtime Performance Advisor, perform the following steps in the administrative console:

1. Click **Servers > Application servers** in the administrative console navigation tree.
2. Click *server_name* > **Runtime Performance Advisor Configuration**.
3. Click the **Runtime** tab.
4. Select the check box to enable the Runtime Performance Advisor.
5. Click **OK**.
6. From the Runtime or Configuration tab of Runtime Performance Advisor, click **Advice Configuration**.
7. Enable the memory leak detection advice and disable any other unwanted advice.

The memory leak detection advice is enabled.

Important: To achieve the best results for performance tuning, enable the Runtime Performance Advisor when a stable production level load is running.

You can monitor any notifications of memory leaks by checking the SystemOut.log file or Runtime Messages. For more information, see the Viewing the Runtime Performance Advisor recommendations topic.

Getting the automated heap dump generation support

This topic explains the automated heap dump generation support, which is only available on IBM Software Development Kit and analyzes memory leak problems.

Instructions from support involve setting environment variables, restarting the application server, and using the wsadmin tool. To help you analyze memory leak problems when memory leak detection occurs, some automated heap dump generation support is available. This functionality is only available on IBM Software Development Kit.

Most memory leak analysis tools perform some forms of difference evaluation on two heap dumps. Upon detection of a suspicious memory situation, two heap dumps are automatically generated at appropriate times. The general idea is to take an initial heap dump as soon as problem detection occurs. Monitor the memory usage and take another heap dump when you determine that enough memory is leaked, so that you can compare the heap dumps to find the source of the leak.

Enabling automated heap dump generation

Use this task to enable automated heap dump generation on AIX, Linux, and Windows operating systems.

Although heap dumps are only generated in response to a detected memory leak, you must understand that generating heap dumps can have a severe performance impact on the Application Server for several minutes.

To help you analyze memory leak problems when memory leak detection occurs, some automated heap dump generation support is available.

To enable automated heap dump generation support, perform the following steps in the administrative console:

1. Click **Servers > Application servers** in the console navigation tree.
2. Click *server_name* > **Runtime Performance Advisor Configuration**.
3. Click the **Runtime** tab.
4. Select the **Enable automatic heap dump collection** check box.
5. Click **OK**.

The automated heap dump generation support is enabled.

Important: To preserve disk space, the Runtime Performance Advisor does not take heap dumps if more than 10 heap dumps already exist in the WebSphere Application Server home directory. Depending on the size of the heap and the workload on the application server, taking a heap dump might be quite expensive and might temporarily affect system performance.

You can monitor any notifications of memory leaks by checking the `SystemOut.log` file or Runtime Messages. For more information, see the Viewing the Runtime Performance Advisor recommendations topic. If a memory leak is detected and you want to find the heap dump, refer to the Locating and analyzing heap dumps topic.

Related tasks

“Generating heap dumps manually”

Use this task to manually generate heap dumps.

“Locating and analyzing heap dumps” on page 21

Use this task to locate and analyze heap dumps.

Generating heap dumps manually

Use this task to manually generate heap dumps.

Although heap dumps are only generated in response to a detected memory leak, you must understand that generating heap dump can have a severe performance impact on the Application Server for several minutes. When generating multiple heap dumps manually for memory leak analysis, make sure that significant objects are leaked in between the two heap dumps. This approach enables problem determination tools to identify the source of the memory leak.

You might want to manually generate heap dumps for analysis of memory leaks. On Java virtual machines (JVM) in WebSphere Application Server, you cannot enable automated heap dump generation. You might want to designate certain times to take heap dumps because of the overhead involved. On JVM in WebSphere Application Serve, you can manually produce heap dumps by using

the generateHeapDump operation on WebSphere Application Server managed beans (MBeans) that are special Java beans.

The WebSphere Application Server wsadmin tool provides the ability to run scripts. You can use the wsadmin tool to manage a WebSphere Application Server installation, as well as configuration, application deployment, and server run-time operations. The WebSphere Application Server only supports the Jacl and Jython scripting languages. To learn more about the wsadmin tool, see the Starting the wsadmin scripting client topic.

1. Start the wsadmin scripting client. There are several options for you to run scripting commands, ranging from running them interactively to running them in a profile.
2. Invoke the generateHeapDump operation on JVM MBean, for example,
 - Finding JVM objectName:


```
<wsadmin> set objectName [$AdminControl queryNames
WebSphere:type=JVM,process=<serverName>,node=<nodeName>,*]
```
 - Invoking the generateHeapDump operation on JVM MBean:


```
<wsadmin> $AdminControl invoke $objectName generateHeapDump
where,
```

| | |
|------------------|--|
| \$ | is a Jacl operator for substituting a variable name with its value |
| invoke | is the command |
| generateHeapDump | is the operation you are invoking |
| <serverName> | is the name of the server on which you want to generate a HeapDump |
| <nodeName> | is the node to which <serverName> belongs |

After running the wsadmin command, the file name of the heap dump is returned. For more information on finding heap dumps, refer to the Locating and analyzing heap dumps topic. When you have couple of heap dumps, use a number of memory leak problem determination tools to analyze your problem. For example, you can download a technology preview from WebSphere Serviceability.

Related tasks

“Enabling automated heap dump generation” on page 20

Use this task to enable automated heap dump generation on AIX, Linux, and Windows operating systems.

“Locating and analyzing heap dumps”

Use this task to locate and analyze heap dumps.

Locating and analyzing heap dumps

Use this task to locate and analyze heap dumps.

Do not analyze heap dumps on the Application Server machine because analysis is very expensive. For analysis, transfer heap dumps to a dedicated problem determination machine.

When a memory leak is detected and heap dumps are generated, you must analyze heap dumps on a problem determination machine and not on the application server because the analysis is very CPU and disk I/O intensive.

Perform the following steps to locate the heap dump files:

1. On the physical application server where a memory leak is detected, go to the WebSphere Application Server home directory. For example, on the Windows operating system, the directory is:
`C:\WebSphere\AppServer\profiles\myProfile`
2. IBM heap dump files are usually named in the following way:
`heapdump.<date>..<timestamp><pid>.phd`
3. Gather all the .phd files and transfer them to your problem determination machine for analysis.
4. Many tools are available to analyze heap dumps that include Rational Application Developer 6.0 (RAD). WebSphere Application Server serviceability released a technology preview called Memory Dump Diagnostic For Java. You can download this preview from the product download Web site.

When you have couple of heap dumps, use a number of memory leak problem determination tools to analyze your problem. For example, you can download a technology preview from WebSphere Serviceability.

Related tasks

“Enabling automated heap dump generation” on page 20

Use this task to enable automated heap dump generation on AIX, Linux, and Windows operating systems.

“Generating heap dumps manually” on page 20

Use this task to manually generate heap dumps.

Using the performance advisor in Tivoli Performance Viewer

The performance advisor in Tivoli Performance Viewer (TPV) provides advice to help tune systems for optimal performance and provides recommendations on inefficient settings by using the collected Performance Monitoring Infrastructure (PMI) data. Obtain advice by clicking **Performance advisor** in TPV. The performance advisor in TPV provides more extensive advice than the Runtime Performance Advisor. For example, TPV provides advice on setting the dynamic cache size, setting the Java virtual machine (JVM) heap size and using the DB2 Performance Configuration wizard.

1. Enable PMI in the application server as described in the Enabling PMI using the administrative console article. To monitor performance data through the PMI interfaces, you must first enable PMI through the administrative console before restarting the server. If running in a Network Deployment environment, you must enable PMI on both the server and on the node agent before restarting the server and the node agent.
2. Enable data collection and set the PMI monitoring level to Extended. The monitoring levels that determine which data counters are enabled can be set dynamically, without restarting the server. These monitoring levels and the data selected determine the type of advice you obtain. The performance advisor in TPV uses the extended monitoring level; however, the performance advisor in TPV can use a few of the more expensive counters (to provide additional advice) and provide advice on which counters can be enabled.

For example, the advice pertaining to session size needs the PMI statistic set to All. Or, you can use the PMI Custom Monitoring Level to enable the Servlet Session Manager SessionObjectSize counter. The monitoring of the SessionSize PMI counter is expensive, and is not in the Extended PMI statistic set.

Complete this action in one of the following ways:

- a. Performance Monitoring Infrastructure settings.

- b. Enabling Performance Monitoring Infrastructure using the wsadmin tool.
3. In the administrative console, click **Monitoring and Tuning > Performance Viewer > Current activity**.
4. Simulate a production level load. Simulate a realistic production load for your application, if you use the performance advisor in a test environment, or do any other performance tuning. The application must run this load without errors. This simulation includes numbers of concurrent users typical of peak periods, and drives system resources, for example, CPU and memory to the levels that are expected in production. The performance advisor only provides advice when CPU utilization exceeds a sufficiently high level. For a list of IBM business partners providing tools to drive this type of load, see the article, Performance: Resources for learning in the subsection of Monitoring performance with third party tools.
5. Log performance data with TPV.
6. Clicking **Refresh** on top of the table of advice causes the advisor to recalculate the advice based on the current data in the buffer.
7. Tuning advice displays when the Advisor icon is chosen in the TPV Performance Advisor. Double-click an individual message for details. Because PMI data is taken over an interval of time and averaged to provide advice, details within the advice message display as averages.

Note: If the Refresh Rate is adjusted, the Buffer Size must also be adjusted to enable sufficient data to be collected for performing average calculations. Currently 5 minutes of data is required. Hence, the following guidelines intend to help you use the Tivoli Performance Advisor:

- a. You cannot have a Refresh Rate of more than 300 seconds.
- b. $\text{RefreshRate} * \text{BufferSize} > 300$ seconds. Buffer Size * Refresh Rate is the amount of PMI data available in memory and it must be greater than 300 seconds.
- c. For the Tivoli Performance Advisor to work properly with TPV logs, the logs must be at least 300 seconds of duration.

For more information about configuring user and logging settings of TPV, refer to the Configuring TPV settings article.

8. Update the product configuration for improved performance, based on advice. Because Tivoli Performance Viewer refreshes advice at a single instant in time, take the advice from the peak load time. Although the performance advisors attempt to distinguish between loaded and idle conditions, misleading advice might be issued if the advisor is enabled while the system is ramping up or down. This result is especially likely when running short tests. Although the advice helps in most configurations, there might be situations where the advice hinders performance. Because of these conditions, advice is not guaranteed. Therefore, test the environment with the updated configuration to ensure it functions and performs well.

Over a period of time the advisor might issue differing advice. The differing advice is due to load fluctuations and run-time state. When differing advice is received, you need to look at all advice and the time period over which it was issued. You must take advice during the time that most closely represents the peak production load.

Performance tuning is an iterative process. After applying advice, simulate a production load, update the configuration that is based on the advice, and retest for improved performance. This procedure is continued until optimal performance is achieved.

Related tasks

Chapter 4, “Obtaining performance advice from the performance advisors,” on page 11

Performance advisor report in Tivoli Performance Viewer

View recommendations and data from the performance advisor in Tivoli Performance Viewer (TPV) by clicking the Advisor link in TPV for a server.

For more information on how to use the performance advisor in TPV, see the article, *Using the performance advisor in Tivoli Performance Viewer*.

Message

Specifies recommendations for performance tuning.

Click the message to obtain more details.

Performance data in the upper panel

Displays a summary of performance data for WebSphere Application Server. Data here corresponds to the same period that recommendations were provided for. However, recommendations might use a different set of data points during analysis than the set that is displayed by the summary page.

The first table represents the number of requests per second and the response time in milliseconds for both the Web and Enterprise JavaBeans containers.

The pie graph displays the CPU activity as percentage busy and idle.

The second table displays the average thread activity for the Web container and Object Request Broker (ORB) thread pools, and the average database connection activity for connection pools. The activity is expressed as the number of threads or connections busy and idle.

Chapter 5. Tuning the application serving environment

This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

WebSphere Application Server provides tunable settings for its major components to enable you to make adjustments to better match the run-time environment to the characteristics of your application. Many applications can run successfully without any changes to the default values for these tuning parameters. Other applications might need changes, for example, a larger heap size, to achieve optimal performance.

Performance tuning can yield significant gains in performance even if an application is not optimized for performance. However, correcting shortcomings of an application typically results in larger performance gains than are possible with just altering tuning parameters. Many factors contribute to a high performing application.

To optimize your WebSphere Application Servers to their fullest extent, use the performance advisors, the suggested procedures or parameters in the tuning parameter hot list, and the troubleshooting performance checklist.

Performance advisors The performance advisors use the Performance Monitoring Infrastructure (PMI) data to suggest configuration changes to Object Request Broker (ORB) service thread pools, Web container thread pools, connection pool size, persisted session size and time, prepared statement cache size, and session cache size. The Runtime Performance Advisor runs in the application server process, while the other advisor runs in the Tivoli Performance Viewer (TPV). For more information, see the articles, “Using the Runtime Performance Advisor” on page 14 and “Using the performance advisor in Tivoli Performance Viewer” on page 22.

Tuning parameter hot list Review the tuning parameter hot list. These parameters have an important impact on performance. Because these parameters are application dependent, the parameter settings for specific applications and environments can vary.

Tuning parameter index for z/OS Performance tuning for WebSphere Application Server for z/OS becomes a complex exercise because the nature of the runtime involves many different components of the operating system and middleware. Use the Tuning index for WebSphere Application Server for z/OS to find information and parameters for tuning the z/OS operating system, subsystems, the WebSphere Application Server for z/OS run-time environment, and some Java 2 Platform, Enterprise Edition (J2EE) application tuning tips.

Recommendation: Before you read a description of WebSphere Application Server for z/OS tuning guidelines, it is important to note that, no matter how well the middleware is tuned, it cannot make up for poorly designed and coded applications. Focusing on the application code can help improve performance. Often, poorly written or designed application code changes will make the most dramatic improvements to overall performance.

The tuning guide focuses on server tuning. If you want to tune your applications, see the Performance: Resources for learning article for more information about application tuning.

For your convenience, procedures for tuning parameters in other products, such as DB2, Web servers and operating systems are included. Because these products might change, consider these descriptions as suggestions.

Each WebSphere Application Server process has several parameters that influence application performance. You can use the WebSphere Application Server administrative console to configure and tune applications, Web containers, Enterprise JavaBeans (EJB) containers, application servers and nodes in the administrative domain.

If you are a WebSphere Application Server administrator or Systems programmer on WebSphere Application Server for z/OS, refer to Tuning index for WebSphere Application Server for z/OS for z/OS specific tuning tips.

Each parameter description: explains the parameter; provides reasons to adjust the parameter; discusses how to view or set the parameter; as well as indicates default and recommended values.

The Troubleshooting performance topic can save you time detecting problems and help you troubleshoot performance problems.

Additional references:

- WebSphere Application Server - Performance Web site

Related tasks

“Using the Runtime Performance Advisor” on page 14

“Using the performance advisor in Tivoli Performance Viewer” on page 22

“Tuning security configurations” on page 102

Chapter 6, “Troubleshooting performance,” on page 121

This topic illustrates that solving a performance problem is an iterative process and shows how to troubleshoot performance problems.

Tuning application servers

Related reference

“DB2 tuning parameters” on page 85

“Secure Sockets Layer performance tips” on page 107

EJB method Invocation Queuing

Java memory tuning tips

Example: Choosing a topology for better performance

Performance: Resources for learning

HTTP transport custom properties

Use this page to set custom properties for an HTTP transport.

Tuning parameter hot list

WebSphere Application Server provides several tunable parameters and options to match the application server environment to the requirements of your application. The following hot list contains recommendations that have improved performance or scalability, or both, for many applications.

- **Review the hardware and software requirements**

Review the hardware and software requirements on the IBM WebSphere Application Server supported hardware, software, and APIs Web site to get started.

- **Install the most current refresh pack, fix pack, and the recommended interim fixes**

The list of recommended updates is maintained on the Support site.

- **Install the latest WebSphere Application Server PTFs**

The list of recommended updates is maintained on the Support site

- **Check hardware configuration and settings**

Check network connections to make sure that they are running at their highest speed. Sometimes transient errors can cause Ethernet adapters to shift down to a lower speed. Verify that the system has adequate memory and that the number and position of memory dual inline memory module (DIMM) are optimal. With some systems, there are some memory DIMM configurations that permit higher performance than other DIMM configurations. Verify that the hardware that is used is the hardware that is supposed to be used.

- **Review your application design**

You can track many performance problems back to the application design. Review the design to determine if it causes performance problems.

- **Tune the operating systems**

Operating system configuration plays a key role in performance. In many cases, adjustments to some TCP/IP parameters might be necessary for your application.

- **Set the minimum and maximum Java virtual machine (JVM) heap sizes**

Many applications need a larger heap size for best performance.

- **Use a type 4 (or pure Java) JDBC driver**

The type 4 JDBC drivers generally perform faster than type 2 JDBC drivers. Use the immediately preceding link to view a list of database vendor-specific requirements, which can tell you if a type 4 JDBC driver is supported for your database.

- **Tune WebSphere Application Server JDBC data sources and associated connection pools**

The JDBC data source configuration might have a significant performance impact. For example, the connection pool size and prepared statement cache need to be sized based on the number of concurrent requests being processed and the design of the application.

See the Connection pooling topic for more information.

- **Enable the pass by reference option**

Use applications that can take advantage of the pass by reference option to avoid the cost of copying parameters to the stack.

- **Ensure that the transaction log is assigned to a fast disk**

Some applications generate a high rate of writes to the WebSphere Application Server transaction log. Locating the transaction log on a fast disk or disk array can improve response time

- **Tune related components, for example, Web servers and database**

In many cases, some other component, for example, a Web server or database needs adjustments to achieve higher throughput for your entire configuration.

For more information, see the Queuing network topic.

Related tasks

Chapter 5, "Tuning the application serving environment," on page 25

This topic describes the benefits of tuning for optimal performance, highlights

the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

“Tuning Web servers” on page 54

Task overview: Using the dynamic cache service to improve performance

Use the dynamic cache service to improve application performance by caching the output of servlets, commands, and JavaServer Pages (JSP) files.

Related information

IBM eServer Workload Estimator

Tuning TCP/IP buffer sizes

WebSphere Application Server uses the TCP/IP sockets communication mechanism extensively. For a TCP/IP socket connection, the send and receive buffer sizes define the receive window. The receive window specifies the amount of data that can be sent and not received before the send is interrupted. If too much data is sent, it overruns the buffer and interrupts the transfer. The mechanism that controls data transfer interruptions is referred to as flow control. If the receive window size for TCP/IP buffers is too small, the receive window buffer is frequently overrun, and the flow control mechanism stops the data transfer until the receive buffer is empty.

Flow control can consume a significant amount of CPU time and result in additional network latency as a result of data transfer interruptions. It is recommended that you increase buffer sizes avoid flow control under normal operating conditions. A larger buffer size reduces the potential for flow control to occur, and results in improved CPU utilization. However, a large buffer size can have a negative effect on performance in some cases. If the TCP/IP buffers are too large and applications are not processing data fast enough, paging can increase. The goal is to specify a value large enough to avoid flow control, but not so large that the buffer accumulates more data than the system can process.

The default buffer size is 8 KB. The maximum size is 8 MB. The optimal buffer size depends on several network environment factors including types of switches and systems, acknowledgment timing, error rates and network topology, memory size, and data transfer size. When data transfer size is extremely large, you might want to set the buffer sizes up to the maximum value to improve throughput, reduce the occurrence of flow control, and reduce CPU cost.

Buffer sizes for the socket connections between the Web server and WebSphere Application Server are set at 64KB. In most cases this value is adequate.

Flow control can be an issue when an application uses either the IBM Developer Kit for Java(TM) JDBC driver or the IBM Toolbox for Java JDBC driver to access a remote database. If the data transfers are large, flow control can consume a large amount of CPU time. If you use the IBM Toolbox for Java JDBC driver, you can use custom properties to configure the buffer sizes for each data source. It is recommended that you specify large buffer sizes, for example, 1 MB.

Some system-wide settings can override the default 8 KB buffer size for sockets. With some applications, for example, WebSphere Commerce Suite, a buffer size of 180 KB reduces flow control and typically does not adversely affect paging. The optimal value is dependent on specific system characteristics. You might need to try several values before you determine the ideal buffer size for your system. To change the system wide value, perform the following steps:

- Run the Change TCP/IP Attribute, **CHGTCPA** command.

- On the Change TCP/IP Attributes window, press **F4**. The buffer sizes are displayed as the TCP receive and send buffer sizes.
- Type new values.
- Save your changes.
- Recycle TCP/IP.
- Monitor CPU and paging rates to determine if they are within recommended system guidelines.

Repeat this process until you determine the ideal buffer size.

The TCP/IP buffer sizes are changed. Repeat this process until you determine the ideal buffer size.

For more information about TCP/IP performance, see Chapter 5 of the Performance Capabilities Reference. Links to several editions of the Performance Capabilities Reference are in the Performance Management Resource Library.

Tuning Java virtual machines

The application server, being a Java process, requires a Java virtual machine (JVM) to run, and to support the Java applications running on it. As part of configuring an application server, you can fine-tune settings that enhance system use of the JVM. In addition to the following tuning parameters, see also Java memory tuning tips.

A JVM provides the runtime execution environment for Java based applications. WebSphere Application Server is a combination of a JVM runtime environment and a Java based server runtime. It can run on JVMs from different JVM providers. To determine the JVM provider on which your Application Server is running, issue the `java -fullversion` command from within your WebSphere Application Server `install_root/java/bin` directory. You can also check the `SystemOut.log` from one of your servers. When an application server starts, Websphere Application Server writes information about the JVM, including the JVM provider information, into this log file.

From a JVM tuning perspective, there are two main types of JVMs:

- IBM JVMs
- Sun HotSpot based JVMs, including Sun HotSpot JVM on Solaris and HP's JVM for HP-UX

Even though JVM tuning is dependent on the JVM provider general tuning concepts apply to all JVMs. These general concepts include:

- **Compiler tuning.** All JVMs use Just In Time (JIT) compilers to compile Java byte codes into native instructions during server run-time.
- **Java memory or heap tuning.** The JVM memory management function, or garbage collection provides one of the biggest opportunities for improving JVM performance.
- **Class loading tuning.**
- **Optimize the startup performance and the runtime performance** In some environments, it is more important to optimize the startup performance of your WebSphere Application Server rather than the runtime performance. In other environments, it is more important to optimize the runtime performance. By

default, IBM JVMs are optimized for runtime performance while HotSpot based JVMs are optimized for startup performance.

The behavior of the Java JIT compiler has the biggest influence over whether startup or runtime performance are optimized. The initial optimization level used by the compiler influences the length of time it takes to compile a class method and the length of time it takes to start the server. For faster startups, you can reduce the initial optimization level that the compiler uses. This means that the runtime performance of your applications may be degraded because the class methods are now compiled at a lower optimization level.

It is hard to provide a specific runtime performance impact statement because the compilers might recompile class methods during runtime execution based upon the compiler's determination that recompiling might provide better performance. Ultimately, the size of the application will have a major influence on the amount of runtime degradation. Smaller applications have a higher probability of having their methods recompiled. Bigger applications are less likely to have their methods recompiled. The default settings for IBM JVMs use a high optimization level for the initial compiles. You can use the following IBM JVM option if you need to change this behavior:

-Xquickstart

This setting influences how the IBM JVM uses a lower optimization level for class method compiles, which provides for faster server startups, at the expense of runtime performance. If this parameter is not specified, the IBM JVM defaults to starting with a high initial optimization level for compiles. This setting provides faster runtime performance at the expense of slower server starts.

| | |
|---------------------|---|
| Default: | High initial compiler optimizations level |
| Recommended: | High initial compiler optimizations level |
| Usage: | -Xquickstart can provide faster server startup times. |

JVMs based on Sun's Hotspot technology initially compile class methods with a low optimization level. Use the following JVM option to change this behavior:

-server

JVMs based on Sun's Hotspot technology initially compile class methods with a low optimization level. These JVMs use a simple compiler and an optimizing JIT compiler. Normally the simple JIT compiler is used. However you can use this option to make the optimizing compiler the one that is used. This change will significantly increase the performance of the server but the server takes longer to warm up when the optimizing compiler is used.

| | |
|---------------------|--|
| Default: | Simple compiler |
| Recommended: | Optimizing compiler |
| Usage: | -server enables the optimizing compiler. |

- **Set the heap size** The following command line parameters are useful for setting the heap size.

– **-Xms**

This setting controls the initial size of the Java heap. Properly tuning this parameter reduces the overhead of garbage collection, improving server response time and throughput. The default setting for this option is typically too low, resulting in a high number of minor garbage collections

| | |
|-----------------|------|
| Default: | 50MB |
|-----------------|------|

| | |
|---------------------|--|
| Recommended: | Workload specific, but higher than the default. |
| Usage: | -Xms256m sets the initial heap size to 256 megabytes |

– **-Xmx**

This setting controls the maximum size of the Java heap. Properly tuning this parameter can reduce the overhead of garbage collection, improving server response time and throughput. The default setting for this is typically too low, resulting in a high number of minor garbage collections.

| | |
|---------------------|--|
| Default: | 256MB |
| Recommended: | Workload specific, but higher than the default. |
| Usage: | -Xmx512m sets the maximum heap size to 512 megabytes |

– **-Xlp**

This setting can be used with the IBM JVM to allocate the heap using large pages. However, if you use this setting your operating system must be configured to support large pages. Using large pages can reduce the CPU overhead needed to keep track of heap memory and might also allow the creation of a larger heap.

See “Tuning operating systems” on page 40 for more information about tuning your operating system.

The size you should specify for the heap depends on your heap usage over time. In cases where the heap size changes frequently, you might improve performance if you specify the same value for the Xms and Xmx parameters.

• **Tune the IBM JVM’s garbage collector**

Use the Java -X option to see the list of memory options.

– **-Xgcpolicy**

Setting gcpolicy to optthruput disables concurrent mark. If you do not have pause time problems, denoted by erratic application response times, you should get the best throughput using this option. Setting gcpolicy to optavgpause enables concurrent mark with its default values. This setting alleviates erratic application response times caused by normal garbage collection. However, this option might decrease overall throughput.

| | |
|---------------------|----------------------|
| Default: | optthruput |
| Recommended: | optthruput |
| Usage: | Xgcpolicy:optthruput |

– **-Xnoclassgc**

By default the JVM unloads a class from memory when there are no live instances of that class left, but this can degrade performance. Turning off class garbage collection eliminates the overhead of loading and unloading the same class multiple times.

If a class is no longer needed, the space that it occupies on the heap is normally used for the creation of new objects. However, if you have an application that handles requests by creating a new instance of a class and if requests for that application come in at random times, it is possible that when the previous requester is finished, the normal class garbage collection will

clean up this class by freeing the heap space it occupied, only to have to re-instantiate the class when the next request comes along. In this situation you might want to use this option to disable the garbage collection of classes.

| | |
|---------------------|--|
| Default: | class garbage collection enabled |
| Recommended: | class garbage collection disabled |
| Usage: | Xnoclassgc disables class garbage collection |

For additional information, see the following DeveloperWorks article:

<http://www.ibm.com/developerworks/java/>

- **Tune the Sun JVM's garbage collector**

On the Solaris platform, the WebSphere Application Server runs on the Sun Hotspot JVM rather than the IBM JVM. It is important to use the correct tuning parameters with the Sun JVM in order to utilize its performance optimizing features.

The Sun HotSpot JVM relies on generational garbage collection to achieve optimum performance. The following command line parameters are useful for tuning garbage collection.

- **-XX:SurvivorRatio**

The Java heap is divided into a section for old (long lived) objects and a section for young objects. The section for young objects is further subdivided into the section where new objects are allocated (eden) and the section where new objects that are still in use survive their first few garbage collections before being promoted to old objects (survivor space). Survivor Ratio is the ratio of eden to survivor space in the young object section of the heap. Increasing this setting optimizes the JVM for applications with high object creation and low object preservation. Since WebSphere Application Server generates more medium and long lived objects than other applications, this setting should be lowered from the default.

| | |
|---------------------|----------------------|
| Default: | 32 |
| Recommended: | 16 |
| Usage: | -XX:SurvivorRatio=16 |

- **-XX:PermSize**

The section of the heap reserved for the permanent generation holds all of the reflective data for the JVM. This size should be increased to optimize the performance of applications that dynamically load and unload a lot of classes. Setting this to a value of 128MB eliminates the overhead of increasing this part of the heap.

| | |
|---------------------|---|
| Recommended: | 128 MB |
| Usage: | XX:PermSize=128m sets perm size to 128 megabytes. |

- **-Xmn**

This setting controls how much space the young generation is allowed to consume on the heap. Properly tuning this parameter can reduce the overhead of garbage collection, improving server response time and throughput. The default setting for this is typically too low, resulting in a high number of minor garbage collections. Setting this setting too high can cause the JVM to only perform major (or full) garbage collections. These usually take several seconds and are extremely detrimental to the overall

performance of your server. You must keep this setting below half of the overall heap size to avoid this situation.

| | |
|---------------------|--|
| Default: | 2228224 bytes |
| Recommended: | Approximately 1/4 of the total heap size |
| Usage: | -Xmn256m sets the size to 256 megabytes. |

– **-Xnoclassgc**

By default the JVM unloads a class from memory when there are no live instances of that class left, but this can degrade performance. Turning off class garbage collection eliminates the overhead of loading and unloading the same class multiple times.

If a class is no longer needed, the space that it occupies on the heap is normally used for the creation of new objects. However, if you have an application that handles requests by creating a new instance of a class and if requests for that application come in at random times, it is possible that when the previous requester is finished, the normal class garbage collection will clean up this class by freeing the heap space it occupied, only to have to re-instantiate the class when the next request comes along. In this situation you might want to use this option to disable the garbage collection of classes.

| | |
|---------------------|--|
| Default: | class garbage collection enabled |
| Recommended: | class garbage collection disabled |
| Usage: | Xnoclassgc disables class garbage collection |

For additional information on tuning the Sun JVM, see Performance Documentation for the Java HotSpot VM.

• **Tune the HP JVM's garbage collector**

The HP JVM relies on generational garbage collection to achieve optimum performance. The following command line parameters are useful for tuning garbage collection.

– **-Xoptgc**

This setting optimizes the JVM for applications with many short-lived objects. If this parameter is not specified, the JVM usually does a major (full) garbage collection. Full garbage collections can take several seconds and can significantly degrade server performance.

| | |
|---------------------|---|
| Default: | off |
| Recommended: | on |
| Usage: | -Xoptgc enables optimized garbage collection. |

– **-XX:SurvivorRatio**

The Java heap is divided into a section for old (long lived) objects and a section for young objects. The section for young objects is further subdivided into the section where new objects are allocated (eden) and the section where new objects that are still in use survive their first few garbage collections before being promoted to old objects (survivor space). Survivor Ratio is the ratio of eden to survivor space in the young object section of the heap. Increasing this setting optimizes the JVM for applications with high object creation and low object preservation. Since WebSphere Application Server generates more medium and long lived objects than other applications, this setting should be lowered from the default.

| | |
|---------------------|----------------------|
| Default: | 32 |
| Recommended: | 16 |
| Usage: | -XX:SurvivorRatio=16 |

– **-XX:PermSize**

The section of the heap reserved for the permanent generation holds all of the reflective data for the JVM. This size should be increased to optimize the performance of applications which dynamically load and unload a lot of classes. Specifying a value of 128 megabytes eliminates the overhead of increasing this part of the heap.

| | |
|---------------------|--|
| Default: | 0 |
| Recommended: | 128 megabytes |
| Usage: | -XX:PermSize=128m sets PermSize to 128 megabytes |

– **-XX:+ForceMmapReserved**

By default the Java heap is allocated "lazy swap." This saves swap space by allocating pages of memory as needed, but this also forces the use of 4KB pages. This allocation of memory can spread the heap across hundreds of thousands of pages in large heap systems. This command disables "lazy swap" and allows the operating system to use larger memory pages, thereby optimizing access to the memory making up the Java heap.

| | |
|---------------------|--|
| Default: | off |
| Recommended: | on |
| Usage: | -XX:+ForceMmapReserved will disable "lazy swap". |

– **-Xmn**

This setting controls how much space the young generation is allowed to consume on the heap. Properly tuning this parameter can reduce the overhead of garbage collection, improving server response time and throughput. The default setting for this is typically too low, resulting in a high number of minor garbage collections.

| | |
|---------------------|--|
| Default: | No default |
| Recommended: | Approximately 3/4 of the total heap size |
| Usage: | -Xmn768m sets the size to 768 megabytes |

– **Virtual Page Size**

Setting the Java virtual machine instruction and data page sizes to 64MB can improve performance.

| | |
|---------------------|------|
| Default: | 4MB |
| Recommended: | 64MB |

| | |
|---------------|---|
| Usage: | Use the following command. The command output provides the current operating system characteristics of the process executable: <pre>chatr +pi64M +pd64M /opt/WebSphere/AppServer/java/bin/PA_RISC2.0/native_threads/java</pre> |
|---------------|---|

– **-Xnoclassgc**

By default the JVM unloads a class from memory when there are no live instances of that class left, but this can degrade performance. Turning off class garbage collection eliminates the overhead of loading and unloading the same class multiple times.

If a class is no longer needed, the space that it occupies on the heap is normally used for the creation of new objects. However, if you have an application that handles requests by creating a new instance of a class and if requests for that application come in at random times, it is possible that when the previous requester is finished, the normal class garbage collection will clean up this class by freeing the heap space it occupied, only to have to re-instantiate the class when the next request comes along. In this situation you might want to use this option to disable the garbage collection of classes.

| | |
|---------------------|--|
| Default: | class garbage collection enabled |
| Recommended: | class garbage collection disabled |
| Usage: | Xnoclassgc disables class garbage collection |

For additional information on tuning the HP virtual machine, see Java technology software HP-UX 11i.

- **Tune HP's JVM for HP-UX** Set the following options to improve application performance:

```
-XX:SchedulerPriorityRange=SCHED_NOAGE
-Djava.nio.channels.spi.SelectorProvider=sun.nio.ch.DevPollSelectorProvider
-XX:-ExtraPollBeforeRead
```

Tuning transport channel services

The transport channel services manage client connections and I/O processing for HTTP and JMS requests. These new I/O services are based on new non-blocking I/O features available in Java™. These services provide a highly scalable foundation to WebSphere Application Server request processing.

Key features of the new transport channel services include:

- Scalability, which enables the WebSphere Application Server to handle many concurrent requests.
- Asynchronous request processing, which provides a many-to-one mapping of client requests to Web container threads
- Resource sharing and segregation, which enables thread pools to be shared between the Web container and a messaging service.
- Improved usability and
- Incorporation of autonomic tuning and configuration functions.

Changing the default values for settings on one or more of the TCP, HTTP, or Web container transport channels associated with a transport chain can improve the performance of that chain.

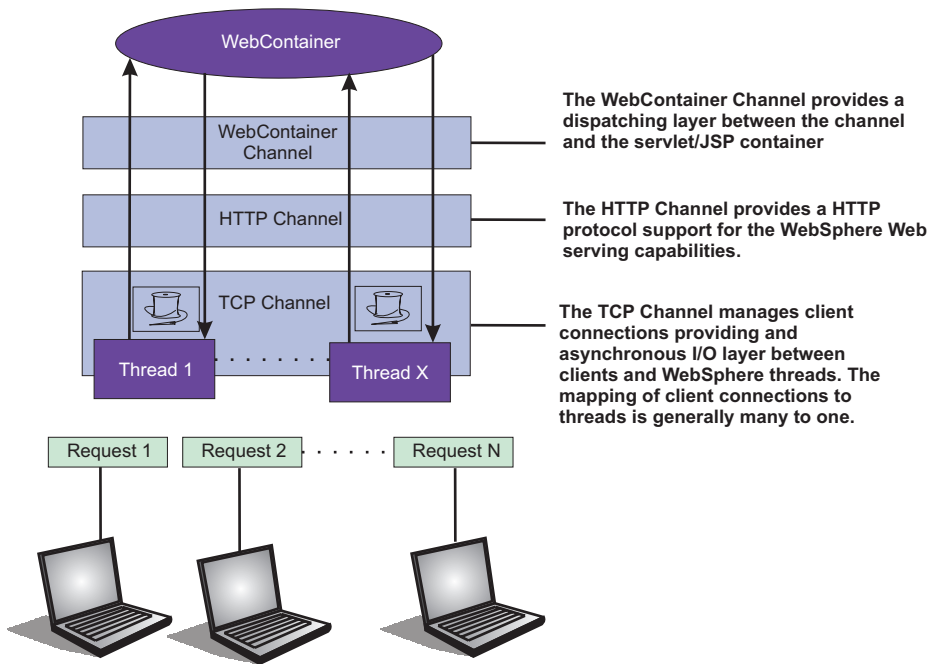


Figure 1. Transport Channel Service

- **Adjust TCP transport channel settings.** In the administration console, click **Servers > Application servers > server_name > Ports**. Then click **View associated transports** for the appropriate port.
 1. Select the transport chain whose properties you are changing.
 2. Click on the TCP transport channel defined for that chain.
 3. **Leave the Maximum open connections parameter set to the default value.** This parameter controls the maximum number of connections that are available for a server's use. It should be left at the default value of 20000, which is the maximum number of connections allowed. The transport channel service by default manages high client connection counts and requires no tuning.
 4. **If client connections are being closed without data being written back to the client, change the value specified for the Inactivity timeout parameter.** This parameter controls the maximum number of connections available for a server's use. Upon receiving a new connection, the TCP transport channel waits for enough data to arrive to dispatch the connection to the protocol specific channels above the TCP transport channel. If not enough data is received during the time period specified for the Inactivity timeout parameter, the TCP transport channel closes the connection. The default value for this parameter is 60 seconds, which is adequate for most applications. You should increase the value specified for this parameter if your workload involves a lot of connections and all of these connections can not be serviced in 60 seconds.
 5. **Assign a thread pool to a specific HTTP port.** Each TCP transport channel is assigned to a particular thread pool. Thread pools can be shared between one or more TCP transport channels as well as with other components. The

default settings for a TCP transport channel is to have all HTTP based traffic assigned to the WebContainer thread pool and all other traffic assigned to the Default thread pool. Use the Thread pool pull-down to assign a particular thread pool to each TCP transport channel. The default settings for this parameter has all HTTP based traffic assigned to the WebContainer thread pool and all other traffic is assigned to the Default thread pool. (Thread pool collection describes how to create additional thread pools.)

6. **Tune the size of your thread pools.** By default, a thread pool can have a minimum of 10 threads and a maximum of 50 maximum threads. To adjust these values, click on **Thread pools** > *threadpool_name* and adjust the values specified for the Minimum Size and Maximum Size parameters for that thread pool.

Typical applications usually do not need more than 10 threads per processor. One exception is if there is some off server condition, such as a very slow backend request, that causes a server thread to wait for the backend request to complete. In such a case, CPU usage is usually low and increasing the workload does not increase CPU throughput. Thread dumps show nearly all threads in a call out to the backend resource. If this condition exists, and the backend is tuned correctly, try increasing the minimum number of threads in the pool until you see improvements in throughput and thread dumps show threads in other areas of the runtime besides the backend call.

The setting for the Grow as needed parameter should not be changed unless your backend is prone to hanging for long periods of time. This condition might indicate that all of your runtime threads are blocked waiting for the backend instead of processing other work that does not involve the hung backend.

- **Adjust HTTP transport channel settings.** In the administration console, click **Servers > Application servers > server_name > Ports**. Then click **View associated transports** for the appropriate port.

1. Select the transport chain whose properties you are changing.
2. Click on the HTTP transport channel defined for that chain.
3. **Tune HTTP keep-alive.** The Use persistent (keep-alive) connections setting controls whether or not connections are left open between requests. Leaving the connections open can save setup and teardown costs of sockets if your workload has clients that send multiple requests. The default value is true and is the optimal setting in most cases.

If your clients only send single requests over substantially long periods of time, it is probably better to disable this option and close the connections right away rather than to have the HTTP transport channel setup the timeouts to close the connection at some later time.

4. **Change the value specified for the Maximum persistent requests parameter to increase the number of requests that can flow over a connection before it is closed.** When the Use persistent (keep-alive) connections option is enabled, the Maximum persistent requests parameter controls the number of requests that can flow over a connection before it is closed. The default value is 100. This value should be set to a value such that most, if not all, clients always have an open connection when they make multiple requests during the same session. A proper setting for this parameter helps to eliminate unnecessary setting up and tearing down of sockets.

For test scenarios in which the client will never close a socket or where sockets are always proxy or Web servers in front of your application server, a value of -1 will disable the processing which limits the number of requests

over a single connection. The persistent timeout will still shutdown some idle sockets and protect your server from running out of open sockets.

5. **Change the value specified for the Persistent timeout parameter to increase the length of time that a connection is held open before being closed due to inactivity.** The Persistent timeout parameter controls the length of time that a connection is held open before being closed because there is no activity on that connection. The default value is 30 seconds. This parameter should be set to a value that keeps enough connections open so that most clients can obtain a connection available when they need to make a request.
 6. **If clients are having trouble completing a request because it takes them more than 60 seconds to send their data, change the value specified for the Read timeout parameter.** Some clients pause more than 60 seconds while sending data as part of a request. To ensure they are able to complete their requests, change the value specified for this parameter to a length of time in seconds that is sufficient for the clients to complete the transfer of data. Be careful when changing this value that you still protect the server from clients who send incomplete data and thereby utilize resources (sockets) for an excessive amount of time.
 7. **If some of your clients require more than 60 seconds to receive data being written to them, change the value specified for the Write timeout parameter.** Some clients are slow and require more than 60 seconds to receive data that is sent to them. To ensure they are able to obtain all of their data, change the value specified for this parameter to a length of time in seconds that is sufficient for all of the data to be received. Be careful when changing this value that you still protect the server from malicious clients.
- **Adjust Web container transport channel settings.** In the administration console, click **Servers > Application servers > *server_name* > Ports**. Then click **View associated transports** for the appropriate port.
 1. Select the transport chain whose properties you are changing.
 2. Click on the Web container transport channel defined for that chain.
 3. **If multiple writes are required to handle responses to the client, change the value specified for the Write buffer size parameter to a value that is more appropriate for your clients.** The Write buffer size parameter controls the maximum amount of data per thread that the Web container buffers before sending the request on for processing. The default value is 32768 bytes, which is sufficient for most applications. If the size of a response is greater than the size of the write buffer, the response will be chunked and written back in multiple TCP writes.

If you need to change the value specified for this parameter, make sure the new value enables most requests to be written out in a single write. To determine an appropriate value for this parameter, look at the size of the pages that are returned and add some additional bytes to account for the HTTP headers.
 - Click **Apply** and then **Save** to save these changes.

Checking hardware configuration and settings

An optimal hardware configuration enables applications to get the greatest benefit from performance tuning. The hardware speed impacts all types of applications and is critical to overall performance.

The following parameters include considerations for selecting and configuring the hardware on which the application servers run.

- **Optimize disk speed**

- **Description:** Disk speed and configuration have a dramatic effect on the performance of application servers running applications that are heavily dependent on the database support, using extensive messaging, or processing workflow. The disk input or output subsystems that are optimized for performance, for example Redundant Array of Independent Disks (RAID) array, high-speed drives, and dedicated caches, are essential components for optimum application server performance in these environments.

Application servers with fewer disk requirements can benefit from a mirrored disk drive configuration that improves reliability and has good performance.

- **Recommendation:** Spread the disk processing across as many disks as possible to avoid contention issues that typically occur with 1- or 2-disk systems. Placing the database tables on disks that are separate from the disks that are used for the database log files reduces disk contention and improve throughput.
- **Description:** Disk speed and the number of disk arms have a significant effect on application server performance in the following cases:
 - Your application is heavily dependent on database support .
 - Your application uses messaging extensively.
- **Recommendation:** Use disk I/O subsystems that are optimized for performance, for example, Redundant Array of Independent Disks (RAID). Distribute the disk processing across as many disks as possible to avoid contention issues that occur with 1 or 2 disk systems. For more information about disk arms and how they can affect performance, see the iSeries Disk Arm Requirements documentation.

- **Increase processor speed and processor cache**

- **Description:** In the absence of other bottlenecks, increasing the processor speed often helps throughput and response times. A processor with a larger L2 or L3 cache yields higher throughput, even if the processor speed is the same as a CPU with a smaller L2 or L3 cache.
- **Description:** In the absence of other bottlenecks, increasing the processing power can improve throughput, response times, or both. On WebSphere Application Server for i5/OS, processing power can be related to the Commercial Processing Workload (CPW) value of the system. For more information about CPW values, see the Performance Capabilities Reference Manual in the Performance Management Resource Library Web site.

- **Increase system memory**

- **Description:** Increase memory to prevent the system from paging memory to the disk to improve performance. Allow a minimum of 256 MB of memory for each processor and 512 MB per application server. Adjust the available memory when the system pages and the processor utilization is low because of the paging. The memory access speed might depend on the number and placement of the memory modules. Check the hardware manual to make sure that your configuration is optimal.
- **Recommendation:** Use 256 MB of memory for each processor and 512 MB per application server. Some applications might require more memory.
- **Description:** If a large number of page faults occur, performing the following tasks to improve performance:
 - Increase the memory available to WebSphere Application Server.
 - Move WebSphere Application Server to another memory pool.
 - Remove jobs from the WebSphere Application Server memory pool

- **Recommendation:** To determine the current page fault level, run the Work with System Status (WRKSYSSTS) command from an i5/OS command line. For information about the minimum memory requirements, see the IBM Support Web site.
- **Run network cards and network switches at full duplex**
 - **Description:** Run network cards and network switches at full duplex and use the highest supported speed. Full duplex is much faster than half duplex. Verify that the network speed of adapters, cables, switches, and other devices can accommodate the required throughput. Some Web sites might require multiple gigabit links.
 - **Recommendation** Make sure that the highest speed is in use on 10/100/1000 Ethernet networks.
- **Verify that the activity levels for storage pools are sufficient**
 - **Description:** Verify that the activity levels for storage pools are sufficient. Increasing these values can prevent threads from transitioning into the ineligible condition.
 - **Recommendation**
 - To modify the activity level for the storage pool in which you are running WebSphere Application Server, run the following **WRKSYSSTS** command from the command line:
WRKSYSSTS ASTLVL(*INTERMED)
 - Perform the following steps to set the QMAXACTLVL system value to a value equal to or greater than the total activity level for all pools, or *NOMAX:
 - Run the following **WRKSYSSTS** command from the command line:
WRKSYSSTS ASTLVL(*INTERMED)
 - Adjust the value in the **Max Active** column.

Related tasks

“Tuning operating systems”

Tuning the z/OS operating system

Related information

<http://www-1.ibm.com/servers/eserver/series/perfmgmt/resource.htm>

Tuning operating systems

The following tuning parameters are specific to operating systems. Because these operating systems are not WebSphere Application Server products, be aware that the products can change and results can vary.

Use the following links to go to the appropriate operating system to tune:

Note: Check your operating system documentation to determine how to make the tuning parameters changes permanent and if a reboot is required.

Tuning Windows systems

This topic describes how to tune Windows 2000, Windows XP, and Windows 2003 operating systems to optimize the performance of WebSphere Application Server. Because Windows operating systems are not WebSphere Application Server products, be aware that the products can change and results can vary.

When you have a performance concern, check the operating system settings to determine if they are appropriate for your application.

- **TcpTimedWaitDelay**
 - **Description:** Determines the time that must elapse before TCP/IP can release a closed connection and reuse its resources. This interval between closure and release is known as the TIME_WAIT state or twice the maximum segment lifetime (2MSL) state. During this time, reopening the connection to the client and server costs less than establishing a new connection. By reducing the value of this entry, TCP/IP can release closed connections faster and provide more resources for new connections. Adjust this parameter if the running application requires rapid release, the creation of new connections, or an adjustment because of a low throughput caused by multiple connections in the TIME_WAIT state.
 - **How to view or set:**
 1. Use the **regedit** command, access the HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\TCPIP\Parameters registry subkey, and create a new REG_DWORD value named TcpTimedWaitDelay.
 2. Set the value to decimal 30, which is Hex 0x0000001e. This value sets the wait time to 30 seconds.
 3. Stop and restart the system.
 - **Default value:** 0xF0, which sets the wait time to 240 seconds (4 minutes).
 - **Recommended value:** A minimum value of 0x1E, which sets the wait time to 30 seconds.
- **MaxUserPort**
 - **Description:** Determines the highest port number that TCP/IP can assign when an application requests an available user port from the system.
 - **How to view or set:**
 1. Use the **regedit** command, access the HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\TCPIP\Parameters registry subkey, and create a new REG_DWORD value named MaxUserPort.
 2. Set this value to at least decimal 32768.
 3. Stop and restart the system.
 - **Default value:** None
 - **Recommended value:** At least decimal 32768.
- **MaxConnect Backlog**
 - **Description:** If many connection attempts are received simultaneously, increase the default number of pending connections that are supported by the operating system.
 - **How to view or set:**
 1. Use the **regedit** command and access the HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\AFD\Parameters registry subkey
 2. Create and set (and create if necessary) the following values:
 - "EnableDynamicBacklog"=dword:00000001
 - "MinimumDynamicBacklog"=dword:00000020
 - "MaximumDynamicBacklog"=dword:00001000
 - "DynamicBacklogGrowthDelta"=dword:00000010
 3. These values request a minimum of 20 and a maximum of 1000 available connections. The number of available connections is increased by 10 each time that there are fewer than the minimum number of available connections.
 4. Stop and restart the system.
- **KeepAliveInterval**

- **Description:** Determines how often TCP repeats keep-alive transmissions when no response is received.
- **How to view or set:**
 1. Use the **regedit** command, access the HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\AFD\Parameters registry subkey, and create a new REG_DWORD value named KeepAliveInterval.
 2. Set this value to 1 second.
 3. Stop and restart the system.
- **Default value:** 1 second
- **Recommended value:** None
- **TcpMaxDataRetranmission**
 - **Description:** Determines how many times TCP retransmits an unacknowledged data segment on an existing connection.
 - **How to view or set:**
 1. Use the **regedit** command, access the HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters registry subkey, and edit or create a new REG_DWORD value named TcpMaxDataRetranmissions.
 2. Set this value to 5 seconds
 3. Stop and restart the system.
 - **Default value:** 5 second
 - **Recommended value:** None
- **TPC/IP acknowledgements**
 - TCP/IP can be the source of some significant remote method delays. You can increase TCP performance by immediately acknowledging incoming TCP segments, in all situations.
Complete the following steps to immediately acknowledge incoming TCP segments on a server that runs a Microsoft Windows 2000 operating system:
 1. Start the Registry Editor (regedit.exe).
 2. Locate and click the following registry subkey:
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\Interfaces\
 3. On the Edit menu, click **Add Value**, and create the following registry value:
Value name: TcpDelAckTicks
Data type: REG_DWORD
Value data: 0
Quit Registry Editor.
 4. Restart your Windows operating system.
Similarly, to immediately acknowledge incoming TCP segments on a server that runs a Microsoft Windows XP or Windows Server 2003 operating system:
 1. Start the Registry Editor (regedit.exe).
 2. Locate and then click the following registry subkey:
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\Interfaces\
 3. On the Edit menu, click **New > DWORD Value**.
 4. Name the new value, TcpAckFrequency, and assign it a value of 1.
 5. Close the Registry Editor.
 6. Restart your Windows operating system.
- **Large page support**
 - **Description:** Using large pages can reduce the CPU overhead of managing a large JVM heap.
 - **How to view or set:** The Windows operating system provides large page support by default. Use the -Xlp JVM option to make use of this support.

This tuning procedure improves performance of WebSphere Application Server on Windows 2000, Windows XP, and Windows 2003 operating systems.

After tuning your operating system for performance, consult other tuning topics for various tuning tips.

Related tasks

“Tuning Linux systems”

This topic describes how to tune the Linux operating system to optimize the performance of your WebSphere Application Server. Because the Linux operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning AIX systems” on page 45

This topic describes how to tune the AIX operating system to optimize the performance of your WebSphere Application Server. Because the AIX operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning Solaris systems” on page 48

The following tuning parameters are specific to the Solaris operating system. Because the Solaris operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

“Tuning HP-UX systems” on page 51

This topic describes how to tune the HP-UX operating system to optimize the performance of your WebSphere Application Server. Because the HP-UX operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

Tuning Linux systems

This topic describes how to tune the Linux operating system to optimize the performance of your WebSphere Application Server. Because the Linux operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

When you have a performance concern, check the operating system settings to determine if these settings are appropriate for your application.

• timeout_timewait parameter

- **Description:** Determines the time that must elapse before TCP/IP can release a closed connection and reuse its resources. This interval between closure and release is known as the TIME_WAIT state or twice the maximum segment lifetime (2MSL) state. During this time, reopening the connection to the client and server cost less than establishing a new connection. By reducing the value of this entry, TCP/IP can release closed connections faster, providing more resources for new connections. Adjust this parameter if the running application requires rapid release, the creation of new connections, and a low throughput due to many connections sitting in the TIME_WAIT state.

– **How to view or set:**

Issue the following command to set the timeout_timewait parameter to 30 seconds:

```
echo 30 > /proc/sys/net/ipv4/tcp_fin_timeout
```

• SUSE Linux Enterprise Server 8 (SLES 8) SP2A - sched_yield_scale tuning

- **Description:** The Linux scheduler is very sensitive to excessive context switching, so fixes are integrated into the SLES 8 kernel distribution to introduce delay when a thread yields processing. This fix is automatically enabled in SLES 8 SP3, but must be enabled explicitly in SLES 8 SP2A or later.

- **How to view or set:**
 1. Upgrade your SLES 8 service pack to SP2A.
 2. Issue the `sysctl -w sched_yield_scale=1` command .
- **Default value:** 0
- **Recommended value:** 1
- **RedHat Advanced Server 2.1 kernel update**
 - **Description:** Kernel updates for RedHat Advanced Server 2.1 implemented changes that affect WebSphere Application Server performance, especially memory-to-memory HTTP session replication.
 - **How to view or set:**
 1. Issue the `uname -a` command
 2. If you are running any kernel prior to 2.4.9-e.23, upgrade at least to the RedHat Advanced Server 2.1 kernel, but preferably to the latest supported.
 - **Default value:** 2.4.9-e.3
 - **Recommended value:** 2.4.9-e.23
- **Linux file descriptors (ulimit)**
 - **Description:** Specifies the number of open files that are supported. The default setting is typically sufficient for most applications. If the value set for this parameter is too low, a file open error, memory allocation failure, or connection establishment error might be displayed.
 - **How to view or set:** Check the UNIX reference pages on the `ulimit` command for the syntax of different shells. To set the `ulimit` command to 8000 for the KornShell shell (ksh), issue the `ulimit -n 8000` command. Use the `ulimit -a` command to display the current values for all limitations on system resources.
 - **Default value:** For SUSE Linux Enterprise Server 9 (SLES 9), the default is 1024.
 - **Recommended value:** 8000
- **Connection backlog**
 - **Description:** Change the following parameters when a high rate of incoming connection requests result in connection failures:

```
echo 3000 > /proc/sys/net/core/netdev_max_backlog
echo 3000 > /proc/sys/net/core/somaxconn
```
- **TCP_KEEPALIVE_INTERVAL**
 - **Description:** Determines the wait time between isAlive interval probes.
 - **How to view or set:** Issue the following command to set the value:

```
cd /proc/sys/net/ipv4 edit tcp_keepalive_intv
```
 - **Default value:** 75 seconds
 - **Recommended value:** 15 seconds
- **TCP_KEEPALIVE_PROBES**
 - **Description:** Determines the number of probes before timing out.
 - **How to view or set:** Issue the following command to set the value:

```
edit tcp_keepalive_probes
```
 - **Default value:** 9 seconds
 - **Recommended value:** 5 seconds
- **Allocating large pages for Java virtual machine (JVM) heap (tested with SLES 9)**

Some applications require a very large heap for optimal performance. The CPU overhead of managing a large heap can be reduced by using the "large page" support that is provided by the CPU and operating system. The following example assumes a large page size of 4MB and a desired heap size of 2300MB.

 1. Set the following three settings by a `sysctl.conf` file.
 - a. Set the number of large pages (2300MB = 575 * 4MB) by issuing the following command:

- ```
vm.nr_hugepages = 575
```
- b. Set the maximum shared segment size to 2300MB plus a little more (about 95MB) ( $2511724800 = 2300 * 1048576 + 100000000$ ) by issuing the following command:
 

```
kernal.shmmax = 2511724800
```
  - c. Set the total amount of memory to be shared by issuing the following command:
 

```
kernal.shmall = 2511724800
```
2. Specify the `-Xlp` JVM option so the JVM heap can utilize large pages.
  3. Set the `Xmx` JVM option to 2300MB.
  4. Relocate the program text to a lower virtual memory address to provide more address space for a larger heap. Run the following command to relocate the text in the script that invokes the JVM or in a `.profile` file:
 

```
echo "0x10000000" > /proc/self/mapped_base
```

This tuning procedure improves performance of WebSphere Application Server on the Linux operating system.

After tuning your operating system for performance, consult other tuning topics for various tuning tips.

#### Related tasks

“Tuning Windows systems” on page 40

This topic describes how to tune Windows 2000, Windows XP, and Windows 2003 operating systems to optimize the performance of WebSphere Application Server. Because Windows operating systems are not WebSphere Application Server products, be aware that the products can change and results can vary.

“Tuning AIX systems”

This topic describes how to tune the AIX operating system to optimize the performance of your WebSphere Application Server. Because the AIX operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning Solaris systems” on page 48

The following tuning parameters are specific to the Solaris operating system. Because the Solaris operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

“Tuning HP-UX systems” on page 51

This topic describes how to tune the HP-UX operating system to optimize the performance of your WebSphere Application Server. Because the HP-UX operating system is not a WebSphere Application Server product, be aware that it can change and results vary

## Tuning AIX systems

This topic describes how to tune the AIX operating system to optimize the performance of your WebSphere Application Server. Because the AIX operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

- **TCP\_TIMEWAIT**

- **Description:** Determines the time that must elapse before TCP/IP can release a closed connection and reuse its resources. This interval between closure and release is known as the `TIME_WAIT` state or twice the maximum segment lifetime (2MSL) state. During this time, reopening the connection to the client and server costs less than establishing a new connection. By reducing the

value of this entry, TCP/IP can release closed connections faster, providing more resources for new connections. Adjust this parameter, if the running application requires rapid release or the creation of new connections, or if a low throughput occurs due to many connections sitting in the TIME\_WAIT state.

– **How to view or set:**

Issue the following command to set TCP\_TIMEWAIT state to 15 seconds:

```
/usr/sbin/no -o tcp_timewait =1
```

• **AIX operating systems with DB2**

– **Description:** Separating your DB2 log files from the physical database files can boost performance. You can also separate the log and the database files from the drive that contains the Journaled File System (JFS) service. AIX uses specific volume groups and file systems for the JFS logging.

– **How to view or set:** Use the AIX filemon utility to view all the file system input and output and to strategically select the file system for the DB2 log files. Set the DB2 log location according to the “DB2 tuning parameters” on page 85 topic.

– **Default value:** The default location for the DB2 log files is often the same disk drive where the database tables are stored.

– **Recommended value:** Move the files to a disk that is separate from the DB2 data and has the minimum input or output activity.

• **AIX file descriptors (ulimit)**

– **Description:** Specifies the number of open files that are permitted. The default setting is typically sufficient for most applications. If the value set for this parameter is too low, errors might occur when opening files or establishing connections. A Memory allocation error might be displayed. To prevent WebSphere Application Server from running short on resources, remove the upper limits (ulimit) for resources on the user account that the WebSphere Application Server process runs on.

– **How to view or set:** Perform the following steps to change the ulimit settings:

1. Open the command window.
2. Type **smitty users** to open the AIX configuration program.
3. Select **Change** or **Show Characteristics** of a user.
4. Type the name of the user account that WebSphere Application Server runs on.
5. Press **Enter**.
6. Change the following settings to the indicated value:

```
Soft FILE Size -1
Soft CPU Time -1
Soft STACK Size -1
Soft CORE File Size -1
Hard FILE Size -1
Hard CPU Time -1
Hard STACK Size -1
Hard CORE File Size -1
```

7. Press **Enter** to save changes.
8. Log out and log in your account.
9. Restart WebSphere Application Server.

– **Default value:** For the AIX operating system, the default setting is 2000.

– **Recommended value:** Unlimited

• **AIX TCP\_KEEPIDLE**

– **Description:** The keepAlive packet ensures that a connection stays in an active/ESTABLISHED state.

- **How to view or set:** Use the **no** command to determine the current value or to set the value. The change is effective until the next time you restart the machine. To permanently change the value, add the **no** command to the `/etc/rc.net` directory. For example:

```
no -o tcp_keepidle=600
```

- **Default value:** 14400 milliseconds.
  - **Recommended value:** 600 milliseconds.
- **TCP\_KEEPINTVL**
    - **Description:** Specifies the interval between packets that are sent to validate the connection.
    - **How to view or set:** Use the following command to set the value to 5 seconds:

```
no -o tcp_keepintvl=10
```

- **Default value:** 150(1/2 seconds)
- **Recommended value:** 10(1/2 seconds)

- **TCP\_KEEPINIT**

- **Description:** Specifies the initial timeout value for TCP connection.
- **How to view or set:** Use the following command to set the value to 20 seconds:

```
no -o tcp_keepinit=40
```

- **Default value:** 150(1/2 seconds)
- **Recommended value:** 40(1/2 seconds)

- **Allocating large pages for Java virtual machines heap**

Some applications require a very large heap for optimal performance. Reduce the CPU overhead of managing a large heap by using large page support that is provided by the CPU and the operating system. The following steps allocate 4 GB of RAM as large pages:

1. As root user, run the following commands to reserve 4 GB of large page:

```
vmo -r -o lpgg_regions=256 -o lpgg_size=16777216
bosboot -ad /dev/ipldevice
reboot -q
```

2. After reboot, run the following command to enable large page support on the AIX operating system:

```
vmo -p -o v_pinshm=1
```

3. As root user, add the following capabilities for the user:

```
chuser capabilities=CAP_BYPASS_RAC_VMM,CAP_PROPAGATE $USER
```

4. Add the `-Xlp` Java options to the Java command.

5. Validate large page support is used with the following command:

```
vmstat -l 1
```

**Note:** The "alp" column is non-zero when the application is running.

- **Other AIX information**

- Consider the other AIX operating system settings that are not within the scope of this document. You can adjust the following additional settings:
  - Adapter transmit and receive queue
  - TCP/IP socket buffer
  - IP protocol mbuf pool performance
  - Update file descriptors
  - Update the scheduler

For more information about AIX operating systems, see Performance: Resources for learning.

This tuning procedure improves performance of WebSphere Application Server on the AIX operating system.

After tuning your operating system for performance, consult the other tuning topics for various tuning tips.

#### **Related tasks**

“Tuning Windows systems” on page 40

This topic describes how to tune Windows 2000, Windows XP, and Windows 2003 operating systems to optimize the performance of WebSphere Application Server. Because Windows operating systems are not WebSphere Application Server products, be aware that the products can change and results can vary.

“Tuning Linux systems” on page 43

This topic describes how to tune the Linux operating system to optimize the performance of your WebSphere Application Server. Because the Linux operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning Solaris systems”

The following tuning parameters are specific to the Solaris operating system. Because the Solaris operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

“Tuning HP-UX systems” on page 51

This topic describes how to tune the HP-UX operating system to optimize the performance of your WebSphere Application Server. Because the HP-UX operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

## **Tuning Solaris systems**

The following tuning parameters are specific to the Solaris operating system. Because the Solaris operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

On the Solaris operating system, WebSphere Application Server runs on the Sun Hotspot Java virtual machine (JVM). It is important to use the correct tuning parameters with the Sun JVM to utilize its performance optimizing features. Refer to the Tuning Java virtual machines topic for more information about tuning the JVM. Also, consider the following parameters that are specific to the Solaris operating system to ensure that WebSphere Application Server has enough resources.

- **Solaris file descriptors (ulimit)**
  - **Description:** Specifies the maximum number of open files supported. If the value of this parameter is too low, a Too many files open error is displayed in the WebSphere Application Server stderr.log file.
  - **How to view or set:** Check the UNIX reference pages on the **ulimit** command for the syntax of different shells. For the KornShell (ksh) shell use the **ulimit -n 1024** command. Use the **ulimit -a** command to display the current settings. Use the **ulimit -n 2000** command to set the values.
  - **Default value:** None
  - **Recommended value:** 8000
- **Solaris TCP\_TIME\_WAIT\_INTERVAL**
  - **Description:** Notifies TCP/IP on how long to keep the connection control blocks closed. After the applications complete the TCP/IP connection, the control blocks are kept for the specified time. When high connection rates occur, a large backlog of the TCP/IP connections accumulates and can slow server performance. The server can stall during certain peak periods. If the

server stalls, the **netstat** command shows that many of the sockets that are opened to the HTTP server are in the CLOSE\_WAIT or FIN\_WAIT\_2 state. Visible delays can occur for up to four minutes, during which time the server does not send any responses, but CPU utilization stays high, with all of the activities in system processes.

- **How to view or set:** Use the **get** command to determine the current interval and the **set** command to specify an interval of 30 seconds. For example:

```
ndd -get /dev/tcp tcp_time_wait_interval
ndd -set /dev/tcp tcp_time_wait_interval 30000
```

- **Default value:** The default time wait interval for a Solaris operating system is 240000 milliseconds, which is equal to 4 minutes.
- **Recommended value:** 60000 milliseconds

- **Solaris TCP\_FIN\_WAIT\_2\_FLUSH\_INTERVAL**

- **Description:** Specifies the timer interval prohibiting a connection in the FIN\_WAIT\_2 state to remain in that state. When high connection rates occur, a large backlog of TCP/IP connections accumulates and can slow server performance. The server can stall during peak periods. If the server stalls, using the **netstat** command shows that many of the sockets opened to the HTTP server are in the CLOSE\_WAIT or FIN\_WAIT\_2 state. Visible delays can occur for up to four minutes, during which time the server does not send any responses, but CPU utilization stays high, with all of the activity in system processes.

- **How to view and set:** Use the **get** command to determine the current interval and the **set** command to specify an interval of 67.5 seconds. For example,

```
ndd -get /dev/tcp tcp_fin_wait_2_flush_interval
ndd -set /dev/tcp tcp_fin_wait_2_flush_interval 67500
```

- **Default value:** 675000 milliseconds
- **Recommended value:** 67500 milliseconds

- **Solaris TCP\_KEEPALIVE\_INTERVAL**

- **Description:** The keepAlive packet ensures that a connection stays in an active and established state.

- **How to view or set:** Use the **ndd** command to determine the current value or to set the value. For example:

```
ndd -set /dev/tcp tcp_keepalive_interval 300000
```

- **Default value:** 72000 milliseconds
- **Recommended value:** 15000 milliseconds

- **Solaris TCP\_KEEPALIVE\_PROBES**

- **Description:** Specifies the kernel how many TCP keepalive probes to send out before it decides if a specific connection is broken.

- **How to view or set:** Use the **ndd** command to determine the current value or to set the value. For example:

```
ndd -set /dev/tcp tcp_keepalive_probes 5000
```

- **Default value:** 9000 milliseconds
- **Recommended value:** 5000 milliseconds

- **Solaris kernel semsys:seminfo\_semume**

- **Description:** Limits the maximum semaphore undo entries per process. Because this setting specifies a maximum value, the parameter does not cause the use of additional memory unless it is needed.

- **How to view or set:** This value is displayed as the semume parameter if the **/usr/sbin/sysdef** command is run. There can be an entry in the **/etc/system** file for this tuning parameter. Set this parameter through the **/etc/system** entry:

```
set semsys:seminfo_semume = 200
```

- **Default value:** 10

- **Recommended value:** 200
- **Solaris kernel semsys:seminfo\_semopm**
  - **Description:** Displays as the semume parameter if the `/usr/sbin/sysdef` command is run. An entry in the `/etc/system` file can exist for this tuning parameter. This number is the maximum value of System V semaphore operations per semop call. The default value for this option is too low for highly concurrent systems.
  - **How to view or set:** Set this parameter through the `/etc/system` entry:

```
semsys:seminfo_semopm = 16384
```
  - **Default value:** None
  - **Recommended value:** 16384
- **Connection backlog**
  - **Description:** Change the following parameter when a high rate of incoming connection requests result in connection failures:

```
ndd -get /dev/tcp tcp_conn_req_max_q
ndd -set /dev/tcp tcp_conn_req_max_q 8000
```
  - **Default value:** For Solaris 8, the default value is 128.
  - **Recommended value:** 8000
- :
- **Large page support**

Using large pages can reduce the CPU overhead of managing a large JVM heap. With Solaris 9, large page support is provided by default. No operating system or JVM parameters are necessary to make use of large pages for the JVM heap. With Solaris 8, the Intimate Shared Memory (ISM) feature must be utilized to provide large page support. See the <http://java.sun.com/docs/hotspot/ism.html> Web site for instructions to enable large page support.

This tuning procedure improves the performance of WebSphere Application Server on the Solaris operating system.

After tuning your operating system for performance, consult other tuning topics for various tuning tips.

#### **Related tasks**

“Tuning Windows systems” on page 40

This topic describes how to tune Windows 2000, Windows XP, and Windows 2003 operating systems to optimize the performance of WebSphere Application Server. Because Windows operating systems are not WebSphere Application Server products, be aware that the products can change and results can vary.

“Tuning Linux systems” on page 43

This topic describes how to tune the Linux operating system to optimize the performance of your WebSphere Application Server. Because the Linux operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning AIX systems” on page 45

This topic describes how to tune the AIX operating system to optimize the performance of your WebSphere Application Server. Because the AIX operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning HP-UX systems” on page 51

This topic describes how to tune the HP-UX operating system to optimize the performance of your WebSphere Application Server. Because the HP-UX operating system is not a WebSphere Application Server product, be aware that it can change and results vary.



## Tuning HP-UX systems

This topic describes how to tune the HP-UX operating system to optimize the performance of your WebSphere Application Server. Because the HP-UX operating system is not a WebSphere Application Server product, be aware that it can change and results vary

On the HP-UX operating system, WebSphere Application Server runs on the Java virtual machine (JVM), which is based on the technology of Sun HotSpot JVM. Properly tuning this JVM significantly affects WebSphere Application Server performance by fully utilizing its performance optimizing characteristics. See the “Tuning Java virtual machines” on page 29 topic for details on setting up the JVM on the HP-UX system. It is also important to change some parameters that are specific to the HP-UX operating system to prevent WebSphere Application Server from being deprived of resources.

When you have a performance concern, check the operating system settings to determine if they are appropriate for your application.

•

### Tuning the HP operating system with the DB2 type 2 JDBC driver

When using the type 2 Java Database Connectivity (JDBC) driver on the HP operating system with DB2, you can increase the performance of WebSphere Application Server by preallocating the DB2 trace segment. Perform the following steps:

1. Before starting Application Server, switch to the user that is associated with the DB2 instance.
2. Run the **db2trc alloc** command.
3. Start Application Server.

**Usage note:** Use the type 4 driver for best performance and compatibility.

Another issue with the type 2 JDBC driver on the HP operating system is code page conversion. Creating the database using the UTF-8 code set avoids this problem and significantly increases performance. See the database documentation for instructions on creating databases with a specific code set. Refer to “DB2 tuning parameters” on page 85 for information about DB2 tuning parameters.

### The HP performance tuning parameters

Modify HP-UX 11i settings to significantly improve WebSphere Application Server performance. For additional information about the HP performance tuning parameters, see Performance: Resources for learning.

#### – Java virtual machine (JVM) virtual page size

- **Description:** Sets the JVM instruction and data page sizes to 64 MB to improve performance.
- **How to view or set:** Use the **chattr +pi64M +pd64M /opt/WebSphere/AppServer/java/bin/PA\_RISC2.0/native\_threads/java** command. The command output provides the current operating system characteristics of the process executable.
- **Default value:** 4 MB, if not assigned
- **Recommended value:** 64 MB

#### – HP-UX 11i TCP\_CONN\_REQUEST\_MAX

- **Description:** Specifies the maximum number of connection requests that the operating system can queue when the server does not have available threads. When high connection rates occur, a large backlog of TCP/IP connection requests builds up and client connections are dropped. Adjust

this setting when clients start to time out after waiting to connect. Verify this situation by issuing the **netstat -p tcp** command. Look for the following value: *connect requests dropped due to full queue*

- **How to view or set:** Set this parameter by using the **ndd -set /dev/tcp tcp\_conn\_request\_max 8192** command.
- **Default value:** 4096
- **Recommended value:** In most cases the default is sufficient. Consider adjusting this value to 8192, if the default proves inadequate.
- **HP-UX 11i kernel parameter recommendations**
  - **Description:** One of the main functions of kernel parameters is to control the number of available resources. It is important to alter the HP-UX kernel parameters to ensure that WebSphere Application Server has enough resources available to run optimally. Use the kernel parameter settings in the following table. Set these parameters using graphical System Administration Manager. Type **sam** at the command line to launch the utility.

| Kernel parameter | WebSphere Application Server setting | DB2 setting | Oracle setting |
|------------------|--------------------------------------|-------------|----------------|
| maxuprc          | --                                   | 512         | --             |
| maxfiles         | 2,048                                | --          | --             |
| maxfiles_lim     | 2,048                                | --          | --             |
| nkthread         | 10,000                               | --          | --             |
| max_thread_proc  | 2,048                                | --          | --             |
| nproc            | --                                   | 1,028       | --             |
| nflocks          | --                                   | 8,192       | --             |
| ninode           | --                                   | 2,048       | --             |
| nfile            | --                                   | 8,192       | --             |
| msgseg           | --                                   | 32,767      | --             |
| msgmnb           | --                                   | 65,535      | --             |
| msgmax           | --                                   | 65,535      | --             |
| msgtql           | --                                   | 1,024       | --             |
| msgmap           | --                                   | 258         | --             |
| msgmni           | --                                   | 256         | --             |
| msgssz           | --                                   | 16          | --             |
| semnmi           | --                                   | 512         | 70             |
| semmap           | --                                   | 514         | --             |
| semnms           | --                                   | 1,024       | 200            |
| semnmu           | --                                   | 1,024       | --             |
| shmmax           | --                                   | 966,367,642 | 1 GB           |
| shmmseg          | --                                   | 16          | 10             |
| shmmni           | --                                   | 300         | 100            |

- **How to view or set:** Use the HP-UX SAM utility to set the kernel parameters. See the documentation for your operating system for directions.
- **Default value:** None
- **Recommended value:** See table

- **TCP\_KEEPALIVE\_INTERVAL**
  - **Description:** Determines the interval between probes.
  - **How to view or set:** Use the **ndd** command to determine the current value or to set the value. For example:  

```
ndd -set /dev/tcp tcp_keepalive_interval 15000
```
  - **Default value:** None
  - **Recommended value:** 15000 milliseconds
- **TCP\_KEEPALIVE\_KILL**
  - **Description:** Determines the maximum number of times to probe before dropping.
  - **How to view or set:** Use the **ndd** command to determine the current value or to set the value. For example:  

```
ndd -set /dev/tcp tcp_keepalive_kill 5000
```
  - **Default value:** None
  - **Recommended value:** 5000 milliseconds
- Keeping current with the operating system and Java patches is one of the most important things you can do to optimize the performance of a server. For the latest Java patches, visit the following Web site:  
[HP-UX Patch Information](#)  
 Also, for the latest operating system quality pack, visit the following Web site:  
[Support Plus: Quality Pack Bundles](#)

This tuning procedure improves performance of WebSphere Application Server on the HP-UX operating system.

After tuning your operating system for performance, consult the other tuning topics for various tuning tips.

#### **Related tasks**

“Tuning Windows systems” on page 40

This topic describes how to tune Windows 2000, Windows XP, and Windows 2003 operating systems to optimize the performance of WebSphere Application Server. Because Windows operating systems are not WebSphere Application Server products, be aware that the products can change and results can vary.

“Tuning Linux systems” on page 43

This topic describes how to tune the Linux operating system to optimize the performance of your WebSphere Application Server. Because the Linux operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning AIX systems” on page 45

This topic describes how to tune the AIX operating system to optimize the performance of your WebSphere Application Server. Because the AIX operating system is not a WebSphere Application Server product, be aware that it can change and results can vary.

“Tuning Solaris systems” on page 48

The following tuning parameters are specific to the Solaris operating system. Because the Solaris operating system is not a WebSphere Application Server product, be aware that it can change and results vary.

---

## Tuning Web servers

WebSphere Application Server provides plug-ins for several Web server brands and versions. Each Web server operating system combination has specific tuning parameters that affect the application performance. Following is a list of tuning parameters specific to Web servers. The listed parameters may not apply to all of the supported Web servers. Check your Web server documentation before using any of these parameters.

- **Tune the IBM HTTP Server 2.0.47.1, Apache 2.0.48, and IBM HTTP Server 6.0.** Monitoring the CPU utilization and checking the IBM HTTP Server `error_log` and `http_plugin.log` files can help you diagnose Web server performance problems.

You can also configure the IBM HTTP Server to show a status page:

- Edit the IBM HTTP Server `httpd.conf` file and remove the comment character (#) from the following lines in this file:

```
#LoadModule status_module, modules/ApacheModuleStatus.dll,
#<Location/server-status>
#SetHandler server-status
#</Location>
```

- Save the changes and restart the IBM HTTP Server.
- In a Web browser, go to: `http://yourhost/server-status`. Alternatively, click **Reload** to update status.

(Optional) If the browser supports refresh, go to `http://your_host/server-status?refresh=5` to refresh every five seconds.

- (Optional) If the browser supports refresh, go to `http://your_host/server-status?refresh=5` to refresh every five seconds.

All of these Web servers allocate a thread to handle each client connection. Ensuring that enough threads are available for the maximum number of concurrent client connections helps prevent this tier from being a bottleneck. The settings for these Web servers can be tuned by making changes to the `httpd.conf` file on the Web server system.

You can check the IBM HTTP Server `error_log` file to see if there are any warnings about having reached the maximum number of clients (`MaxClients`). There are several parameters, depending on the specific operating system platform, that determine the maximum number of clients the Web server supports. See [http://httpd.apache.org/docs-2.0/mod/mpm\\_common.html#maxclients](http://httpd.apache.org/docs-2.0/mod/mpm_common.html#maxclients) for a description of the `MaxClients` parameters.

- **Support thousands of concurrent clients.** It is not unusual for a single IBM HTTP Server system to support thousands of concurrent clients. If your requirements are to support more concurrent clients than the number of threads that are supported by the Web server operating system and hardware, consider using multiple Web servers.
- **Respond to a Connection Refused error message.** Some clients might receive a Connection Refused error message if there is a sudden increase in the number of clients. Increasing the `ListenBacklog` and `StartServer` parameters can reduce or eliminate this error.
  - The `ListenBacklog` parameter indicates to the operating system the maximum allowed number of pending connections. Although the IBM HTTP Server default is 511, the actual value can be much higher or lower depending on the corresponding operating system parameter. To handle large numbers of simultaneous connections, this parameter and the corresponding OS parameter might need to be set to the number (possibly thousands) of

expected simultaneous connections. (See “Tuning operating systems” on page 40 for additional information on how to tune your operating system.

- The StartServers parameter indicates the number of IBM HTTP Server processes to initially start. Pre-starting these IBM HTTP Server threads/processes reduces the chance of a user having to wait for a new process to start. You should set this parameter to a value equal to the MinSpareServers parameter so that the minimum number of IBM HTTP Server processes needed for this client load is started immediately.
- **Prevent the frequent creation and destruction of client threads/processes as the number of users change.** You can use the MinSpareServers and MaxSpareServers to specify the minimum and maximum number of servers (client threads/processes) that can exist in an idle state. To prevent frequent creation and destruction of client threads/processes as the number of users change, set this range large enough to include the maximum number of simultaneous users.
- **Change the setting on the Web server’s Access logging parameter to reduce the load on the Web server.** If you do not need to log every access to the Application Server, change the default value of the Web server’s Access logging parameter. This change will reduce the load on the Web server.
- **Modify the settings of the Load balancing option and Retry interval Web server plug-in properties to improve performance.** You can improve the performance of IBM HTTP Server (with the WebSphere Web server plug-in) by modifying the following Web server plug-in configuration properties:
  - Load balancing option, which specifies the load balancing option that the plug-in uses in sending requests to the various application servers associated with that Web server.

The goal of the default load balance option, Round Robin, is to provide an even distribution of work across cluster members. Round Robin works best with Web servers that have a single process sending requests to the Application Server. If the Web server is using multiple processes to send requests to the Application Server, the Random option can sometimes yield a more even distribution of work across the cluster.

- Retry interval. which specifies the length of time to wait before trying to connect to a server that has been marked temporarily unavailable.

The plug-in marks a server temporarily unavailable if the connection to the server fails. Although a default value is 60 seconds, you might have to lower this value in order to increase throughput under heavy load conditions. Lowering the RetryInterval might help when the IBM HTTP Server is configured to have fewer than 10 threads per process.

How can lowering the RetryInterval affect throughput? If the plug-in attempts to connect to a particular application server while the application server threads are busy handling other connections, which happens under heavy load conditions, the connection might time out, causing the plug-in to mark the server temporarily unavailable. If the same plug-in process has other connections open to the same server and a response is received on one of these connections, the server is marked again. If there are only a few threads per IBM HTTP Server process, there might not be an established connection to this application server. When this situation occurs, the plug-in must wait for the entire retry interval.

**Note:** Although lowering the RetryInterval can improve performance, if all the application servers are running, a low value can have an adverse affect when one of the application servers is down. In this case, each

IBM HTTP Server process attempts to connect and fail more frequently, resulting in increased latency and decreased overall throughput.

Making these changes can help the IBM HTTP Server to support more WebSphere Application Server users. To modify these properties, in the administrative console, click **Servers > Web Servers > *Web\_server\_name* > Plug-in properties > Request routing** . (See Web server plug-in tuning tips For more information on tuning the plug-in.)

For more information about tuning heavily loaded Web servers, see Performance: Resources for learning

#### **Related tasks**

Communicating with Web servers

The WebSphere Application Server works with a Web server to route requests for dynamic content, such as servlets, from Web applications. The Web servers are necessary for directing traffic from browsers to the applications that run in WebSphere Application Server. The Web server plug-in uses the XML configuration file to determine whether a request is from the Web server or the Application Server.

---

## **Tuning WebSphere applications**

This topic provides quick links to information about tuning specific WebSphere application types, and the services and containers that support them.

**Note:** The WebSphere Application Server documentation contains a finite set of tuning topics to which the following table provides links. Installing the documentation plug-ins for additional components, such as Service integration, might add new entries to the information table of contents. The new entries will not be shown in the table. To see the complete set of application tuning topics available in this information center installation, expand **Tuning performance > Tuning WebSphere applications** in the table of contents.



## Product architecture and programming model, at a glance

| Application serving environment -- See Tuning the application serving environment                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | WebSphere applications                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | WebSphere applications                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Servers</b></p> <ul style="list-style-type: none"> <li>• Application servers</li> <li>• Java virtual machines</li> <li>• Transport channels</li> <li>• Web servers</li> <li>• More server types</li> <li>• Core groups</li> <li>• Workload balancing</li> </ul> <p><b>Environment</b></p> <ul style="list-style-type: none"> <li>• Hardware</li> <li>• Operating system</li> <li>• Virtual hosts</li> <li>• Variable settings</li> <li>• Shared libraries</li> <li>• Replication domains</li> </ul> <p><b>System administration</b></p> <ul style="list-style-type: none"> <li>• Administrative clients</li> <li>• Configuration files</li> <li>• Domains (cells, nodes)</li> </ul> <p><b>Performance tools</b></p> <ul style="list-style-type: none"> <li>• Monitoring</li> <li>• Tuning performance</li> </ul> <p><b>Troubleshooting tools</b></p> <ul style="list-style-type: none"> <li>• Diagnostic tools</li> <li>• Support and self-help</li> </ul> <p>The product subsystems are discussed in the Product architecture. For the most part, they do not depend on the type of applications being deployed</p> | <p><b>Services</b></p> <ul style="list-style-type: none"> <li>• Security</li> <li>• Naming</li> <li>• ORB</li> <li>• ORB</li> <li>• Transactions</li> </ul> <p><b>J2EE applications</b></p> <ul style="list-style-type: none"> <li>• Web applications &gt; Sessions</li> <li>• EJB applications</li> </ul> <p><b>Clients</b></p> <ul style="list-style-type: none"> <li>• Client applications</li> <li>• Web clients</li> <li>• Web services clients</li> <li>• Administrative clients</li> </ul> <p><b>Web services</b></p> <ul style="list-style-type: none"> <li>• Web services and Service Oriented Architecture (SOA)</li> <li>• Web services security</li> <li>• Web services security</li> </ul> | <p><b>J2EE resources</b></p> <ul style="list-style-type: none"> <li>• Data access resources</li> <li>• Messaging resources</li> <li>• Mail, URLs, and more</li> </ul> <p><b>WebSphere extensions</b></p> <ul style="list-style-type: none"> <li>• ActivitySessions</li> <li>• Application profiling</li> <li>• Asynchronous beans</li> <li>• Dynamic caching</li> <li>• Dynamic and EJB query</li> <li>• Internationalization</li> <li>• Object pools</li> <li>• Scheduler</li> <li>• Startup beans</li> <li>• Work area</li> </ul> |

### Related concepts

Chapter 5, "Tuning the application serving environment," on page 25

This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

## Web applications

### Tuning session management

WebSphere Application Server session support has features for tuning session performance and operating characteristics, particularly when sessions are configured in a distributed environment. These options support the administrator flexibility in determining the performance and failover characteristics for their environment.

The table summarizes the features, including whether they apply to sessions tracked in memory, in a database, with memory-to-memory replication, or all. Click a feature for details about the feature. Some features are easily manipulated using administrative settings; others require code or database changes.

| Feature or option                | Goal                                                                                                                                                                                      | Applies to sessions in memory, database, or memory-to-memory |
|----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| Write frequency                  | Minimize database write operations.                                                                                                                                                       | Database and Memory-to-Memory                                |
| Session affinity                 | Access the session in the same application server instance.                                                                                                                               | All                                                          |
| Multirow schema                  | Fully utilize database capacities.                                                                                                                                                        | Database                                                     |
| Base in-memory session pool size | Fully utilize system capacity without overburdening system.                                                                                                                               | All                                                          |
| Write contents                   | Allow flexibility in determining what session data to write                                                                                                                               | Database and Memory-to-Memory                                |
| Scheduled invalidation           | Minimize contention between session requests and invalidation of sessions by the Session Management facility. Minimize write operations to database for updates to last access time only. | Database and Memory-to-Memory                                |
| Tablespace and row size          | Increase efficiency of write operations to database.                                                                                                                                      | Database (DB2 only)                                          |

#### Related concepts

##### Sessions

A session is a series of requests to a servlet, originating from the same user at the same browser.

#### Related tasks

Task overview: Managing HTTP sessions

Chapter 5, "Tuning the application serving environment," on page 25

This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

#### Configuring scheduled invalidation:



Instead of relying on the periodic invalidation timer that runs on an interval based on the session timeout parameter, you can set specific times for the session management facility to scan for invalidated sessions in a distributed environment. When used with distributed sessions, this feature has the following benefits:

- You can schedule the scan for invalidated sessions for times of low application server activity, avoiding contention between invalidation scans of database or another WebSphere Application Server instance and read and write operations to service HTTP session requests.
- Significantly fewer external write operations can occur when running with the End of Service Method Write mode because the last access time of the session does not need to be written out on each HTTP request. (Manual Update options and Time Based Write options already minimize the writing of the last access time.)

#### **Usage considerations**

- With scheduled invalidation configured, HttpSession timeouts are not strictly enforced. Instead, all invalidation processing is handled at the configured invalidation times.
- HttpSessionBindingListener processing is handled at the configured invalidation times unless the HttpSession.invalidate method is explicitly called.
- The HttpSession.invalidate method immediately invalidates the session from both the session cache and the external store.
- The periodic invalidation thread still runs with scheduled invalidation. If the current hour of the day does not match one of the configured hours, sessions that have exceeded the invalidation interval are removed from cache, but not from the external store. Another request for that session results in returning that session back into the cache.
- When the periodic invalidation thread runs during one of the configured hours, all sessions that have exceeded the invalidation interval are invalidated by removal from both the cache and the external store.
- The periodic invalidation thread can run more than once during an hour and does not necessarily run exactly at the top of the hour.
- If you specify the interval for the periodic invalidation thread using the HttpSessionReaperPollInterval custom property, do not specify a value of more than 3600 seconds (1 hour) to ensure that invalidation processing happens at least once during each hour.

#### **Related reference**

“Tuning session management” on page 58

#### **Configuring write contents:**

In session management, you can configure which session data is written to the database or to another WebSphere instance, depending on whether you are using database persistent sessions or memory to memory replication. This flexibility allows for fewer code changes for the JavaServer Pages (JSP) writer when the application will be operating in a clustered environment. The following options are available in Session Management for tuning what is to be written back:

- Write changed (the default) - Write only session data properties that have been updated through setAttribute method and removeAttribute method calls.
- Write all - Write all session data properties.

The **Write all** setting might benefit servlet and JSP writers who change Java objects' states that reside as attributes in HttpSession and do not call HttpSession.setAttribute method.

However, the use of **Write all** could result in more data being written back than is necessary. If this situation applies to you, consider combining the use of **Write all** with **Time-based write** to boost performance overall. As always, be sure to evaluate the advantages and disadvantages for your installation.

With either Write Contents setting, when a session is first created, complete session information is written, including all of the objects bound to the session. When using database session persistence, in subsequent session requests, what is written to the database depends on whether a single-row or multi-row schema has been set for the session database, as follows:

| Write Contents setting | Behavior with single-row schema                                                    | Behavior with multirow schema                                                                                                                       |
|------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Write changed          | If any session attribute is updated, all objects bound to the session are written. | Only the session data modified through setAttribute method or removeAttribute method calls is written.                                              |
| Write all              | All bound session attributes are written.                                          | All session attributes that currently reside in the cache are written. If the session has never left the cache, all session attributes are written. |

1. Go to the appropriate level of Session Management.
2. Click Distributed Environment Settings
3. Click Custom Tuning Parameters.
4. Select Custom Settings, and click Modify.
5. Select the appropriate write contents setting.

**Related reference**

“Tuning session management” on page 58

**Configuring write frequency:**

In the Session Management facility, you can configure the frequency for writing session data to the database or to a WebSphere instance, depending on whether you use database distributed sessions or memory-to-memory replication. This flexibility enables you to weigh session performance gains against varying degrees of failover support. The following options are available in the Session Management facility for tuning write frequency:

- **End of service servlet-** Write session data at the end of the servlet service method call.
- **Manual update-** Write session data only when the servlet calls the IBMSession.sync method.
- **Time based** (the default) - Write session data at periodic intervals, in seconds (called the *write interval*).

When a session is first created, session information is always written at the end of the service call.

Using the time based write or manual update options can result in loss of data in failover scenarios since the backup copy of the session in the persistent store (for example, a database or another JVM) may not be in sync with the session in the session cache.

### Related reference

“Tuning session management” on page 58

**Base in-memory session pool size:** The base in-memory session pool size number has different meanings, depending on session support configuration:

- With in-memory sessions, session access is optimized for up to this number of sessions.
- With distributed sessions (meaning, when sessions are stored in a database or in another WebSphere Application Server instance); it also specifies the cache size and the number of last access time updates saved in manual update mode.

For distributed sessions, when the session cache has reached its maximum size and a new session is requested, the Session Management facility removes the least recently used session from the cache to make room for the new one.

General memory requirements for the hardware system, and the usage characteristics of the e-business site, determines the optimum value.

Note that increasing the base in-memory session pool size can necessitate increasing the heap sizes of the Java processes for the corresponding WebSphere Application Servers.

### Overflow in non-distributed sessions

By default, the number of sessions maintained in memory is specified by base in-memory session pool size. If you do not wish to place a limit on the number of sessions maintained in memory and allow overflow, set `overflow` to `true`.

Allowing an unlimited amount of sessions can potentially exhaust system memory and even allow for system sabotage. Someone could write a malicious program that continually hits your site and creates sessions, but ignores any cookies or encoded URLs and never utilizes the same session from one HTTP request to the next.

When overflow is disallowed, the Session Management facility still returns a session with the `HttpServletRequest getSession(true)` method when the memory limit is reached, and this is an invalid session that is not saved.

With the WebSphere Application Server extension to `HttpSession`, `com.ibm.websphere.servlet.session.IBMSession`, an `isOverflow` method returns `true` if the session is such an invalid session. An application can check this status and react accordingly.

#### Related concepts

Sessions

A session is a series of requests to a servlet, originating from the same user at the same browser.

#### Related reference

“Tuning session management” on page 58

### Controlling write operations:

You can manually control when modified session data is written out to the database or to another WebSphere Application Server instance by using the `sync` method in the `com.ibm.websphere.servlet.session.IBMSession` interface. This interface extends the `javax.servlet.http.HttpSession` interface. By calling the `sync`

method from the service method of a servlet, you send any changes in the session to the external location. When manual update is selected as the write frequency mode, session data changes are written to an external location only if the application calls the sync method. If the sync method is not called, session data changes are lost when a session object leaves the server cache. When end of service servlet or time based is the write frequency mode, the session data changes are written out whenever the sync method is called. If the sync method is not called, changes are written out at the end of service method or on a time interval basis based on the write frequency mode that is selected.

```
IBMSession iSession = (IBMSession) request.getSession();
iSession.setAttribute("name", "Bob");

//force write to external store
iSession.sync()
```

If the database is down or is having difficulty connecting during an update to session values, the sync method always makes three attempts before it finally creates a `BackedHashtable.getConnectionError` error. For each connection attempt that fails, the `BackedHashtable.StaleConnectionException` is created and can be found in the sync method. If the database opens during any of these three attempts, the session data in the memory is then persisted and committed to the database.

However, if the database is still not up after the three attempts, then the session data in the memory is persisted only after the next check for session invalidation. Session invalidation is checked by a separate thread that is triggered every five minutes. The data in memory is consistent unless a request for session data is issued to the server between these events. For example, if the request for session data is issued within five minutes, then the previous persisted session data is sent.

Sessions are not transactional resources. Because the sync method is associated with a separate thread than the client, the exception that is created does not propagate to the client, which is running on the primary thread. Transactional integrity of data can be maintained through resources such as enterprise beans.

#### **Related reference**

“Tuning session management” on page 58

#### **Tuning parameter settings:**

Use this page to set tuning parameters for distributed sessions.

To view this administrative console page, click **Servers > Application servers > *server\_name* > Web container settings > Session management > Distributed environment settings > Custom tuning parameters**.

#### **Related tasks**

Chapter 5, “Tuning the application serving environment,” on page 25  
This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

#### **Related reference**

“Tuning session management” on page 58

*Tuning level:*

Specifies that the session management facility provides certain predefined settings that affect performance.

Select one of these predefined settings or customize a setting. To customize a setting, select one of the predefined settings that comes closest to the setting desired, click **Custom settings**, make your changes, and then click **OK**.

### Very high (optimize for performance)

|                            |                         |
|----------------------------|-------------------------|
| Write frequency            | Time based              |
| Write interval             | 300 seconds             |
| Write contents             | Only updated attributes |
| Schedule sessions cleanup  | true                    |
| First time of day default  | 0                       |
| Second time of day default | 2                       |

### High

|                           |                        |
|---------------------------|------------------------|
| Write frequency           | Time based             |
| Write interval            | 300 seconds            |
| Write contents            | All session attributes |
| Schedule sessions cleanup | false                  |

### Medium

|                           |                         |
|---------------------------|-------------------------|
| Write frequency           | End of servlet service  |
| Write contents            | Only updated attributes |
| Schedule sessions cleanup | false                   |

### Low (optimize for failover)

|                           |                        |
|---------------------------|------------------------|
| Write frequency           | End of servlet service |
| Write contents            | All session attributes |
| Schedule sessions cleanup | false                  |

### Custom settings

|                                   |                        |
|-----------------------------------|------------------------|
| Write frequency default           | Time based             |
| Write interval default            | 10 seconds             |
| Write contents default            | All session attributes |
| Schedule sessions cleanup default | false                  |

### Tuning parameter custom settings:

Use this page to customize tuning parameters for distributed sessions.

To view this administrative console page, click **Servers > Application servers > *server\_name*Web container settings > Session management > Distributed environment settings > Custom tuning parameters > Custom settings**.

#### Related tasks

Chapter 5, "Tuning the application serving environment," on page 25

This topic describes the benefits of tuning for optimal performance, highlights

the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

**Related reference**

“Tuning parameter settings” on page 62

Use this page to set tuning parameters for distributed sessions.

“Tuning session management” on page 58

*Write frequency:*

Specifies when the session is written to the persistent store.

**End of servlet service**

A session writes to a database or another WebSphere Application Server instance after the servlet completes execution.

**Manual update**

A programmatic sync on the IBMSession object is required to write the session data to the database or another WebSphere Application Server instance.

**Time based**

Session data writes to the database or another WebSphere Application Server instance based on the specified Write interval value. Default: 10 seconds

*Write contents:*

Specifies whether updated attributes are only written to the external location or all of the session attributes are written to the external location, regardless of whether or not they changed. The external location can be either a database or another application server instance.

**Only updated attributes**

Only updated attributes are written to the persistent store.

**All session attribute**

All attributes are written to the persistent store.

*Schedule sessions cleanup:*

Specifies when to clean the invalid sessions from a database or another application server instance.

**Specify distributed sessions cleanup schedule**

Enables the scheduled invalidation process for cleaning up the invalidated HTTP sessions from the external location. Enable this option to reduce the number of updates to a database or another application server instance required to keep the HTTP sessions alive. When this option is not enabled, the invalidator process runs every few minutes to remove invalidated HTTP sessions.

When this option is enabled, specify the two hours of a day for the process to clean up the invalidated sessions in the external location. Specify the times when there is the least activity in the application servers. An external location can be either a database or another application server instance.

**First Time of Day (0 - 23)**

Indicates the first hour during which the invalidated sessions are cleared from the external location. Specify this value as a positive integer between 0 and 23. This value is valid only when schedule invalidation is enabled.

**Second Time of Day (0 - 23)**

Indicates the second hour during which the invalidated sessions are cleared from the external location. Specify this value as a positive integer between 0 and 23. This value is valid only when schedule invalidation is enabled.

## Web services

### Tuning Web services security for version 6.0.x applications

The Java Cryptography Extension (JCE) is integrated into the software development kit (SDK) version 1.4.x and is no longer an optional package. However, due to export and import regulations, the default Java Cryptography Extension (JCE) jurisdiction policy file shipped with the SDK enables you to use strong, but limited, cryptography only. To enforce this default policy, WebSphere Application Server uses a JCE jurisdiction policy file that might introduce a performance impact. The default JCE jurisdiction policy might have a performance impact on the cryptographic functions that are supported by Web services security. If you have Web services applications that use transport level security for XML encryption or digital signatures, you might encounter performance degradation over previous releases of WebSphere Application Server. However, IBM and Sun Microsystems provide versions of these jurisdiction policy files that do not have restrictions on cryptographic strengths. If you are permitted by your governmental import and export regulations, download one of these jurisdiction policy files. After downloading one of these files, the performance of JCE and Web Services security might improve.

For WebSphere Application Server platforms using IBM Developer Kit, Java Technology Edition Version 1.4.2, including the AIX, Linux, and Windows platforms, you can obtain unlimited jurisdiction policy files by completing the following steps:

1. Go to the following Web site:  
<http://www.ibm.com/developerworks/java/jdk/security/index.html>
2. Click **Java 1.4.2**
3. Click **IBM SDK Policy files**.  
The Unrestricted JCE Policy files for SDK 1.4 Web site is displayed.
4. Enter your user ID and password or register with IBM to download the policy files. The policy files are downloaded onto your machine.

For WebSphere Application Server platforms using the Sun-based Java Development Kit (JDK) Version 1.4.2, including the Solaris environments and the HP-UX platform, you can obtain unlimited jurisdiction policy files by completing the following steps:

1. Go to the following Web site: <http://java.sun.com/j2se/1.4.2/download.html>
2. Click **Other Downloads**.

3. Locate the Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files 1.4.2 information and click **Download**. The policy files are downloaded onto your machine.

After following either of these sets of steps, two Java Archive (JAR) files are placed in the JVM `jre/lib/security/` directory.

### **Tuning Web services security for version 5.x applications**

The Java Cryptography Extension (JCE) is integrated into the software development kit (SDK) version 1.4.x and is no longer an optional package. However, due to export and import regulations, the default Java Cryptography Extension (JCE) jurisdiction policy file shipped with the SDK enables you to use strong, but limited, cryptography only. To enforce this default policy, WebSphere Application Server uses a JCE jurisdiction policy file that might introduce a performance impact. The default JCE jurisdiction policy might have a performance impact on the cryptographic functions that are supported by Web services security. If you have Web services applications that use transport level security for XML encryption or digital signatures, you might encounter performance degradation over previous releases of WebSphere Application Server. However, IBM and Sun Microsystems provide versions of these jurisdiction policy files that do not have restrictions on cryptographic strengths. If you are permitted by your governmental import and export regulations, download one of these jurisdiction policy files. After downloading one of these files, the performance of JCE and Web Services security might improve.

For WebSphere Application Server platforms using IBM Developer Kit, Java Technology Edition Version 1.4.2, including the AIX, Linux, and Windows platforms, you can obtain unlimited jurisdiction policy files by completing the following steps:

1. Go to the following Web site:  
<http://www.ibm.com/developerworks/java/jdk/security/index.html>
2. Click **Java 1.4.2**
3. Click **IBM SDK Policy files**.  
The Unrestricted JCE Policy files for SDK 1.4 Web site is displayed.
4. Enter your user ID and password or register with IBM to download the policy files. The policy files are downloaded onto your machine.

For WebSphere Application Server platforms using the Sun-based Java Development Kit (JDK) Version 1.4.2, including the Solaris environments and the HP-UX platform, you can obtain unlimited jurisdiction policy files by completing the following steps:

1. Go to the following Web site: <http://java.sun.com/j2se/1.4.2/download.html>
2. Click **Other Downloads**.
3. Locate the Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files 1.4.2 information and click **Download**. The policy files are downloaded onto your machine.

After following either of these sets of steps, two Java Archive (JAR) files are placed in the JVM `jre/lib/security/` directory.

### **Monitoring the performance of Web services applications**

This topic explains how to monitor the performance of a Web service that is implemented in the WebSphere Application Server.



You can use the Performance Monitoring Infrastructure (PMI) to measure the time required to process Web services requests. To monitor the performance of a Web services application, follow the steps in this task:

1. Enable PMI services in an application server through the administrative console. Select the Web module named `webServicesModule` in step 7.
2. Monitor performance with Tivoli Performance Monitor In the left pane of the performance view, expand the host and server. Select **Web Services**. Run the Web services client application.

PMI provides detailed statistics that can help you gain clear insight into the runtime behavior and performance of Web services. Performance counters enable you to see key performance data for each individual Web service including:

- The number of requests dispatched to an implementation bean
- The number of requests dispatched with successful replies
- The average time in milliseconds to process full requests
- The average time in milliseconds between receiving the request and dispatching it to the bean
- The average time in milliseconds between the dispatch and receipt of a reply from the bean. This represents the time spent in business logic.
- The average time in milliseconds between the receipt of a reply from a bean to the return of a result to the client
- The average size of the SOAP request
- The average size of the SOAP reply

To read about other Web services PMI counters, see PMI data organization.

If your Web services applications are using security, see the information about Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files in Tuning Web services security.

If you are having problems with your Web services applications, read about problems and solutions in Troubleshooting Web services.

#### **Related tasks**

Chapter 5, “Tuning the application serving environment,” on page 25

This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

Enabling PMI using the administrative console

Monitoring performance with Tivoli Performance Viewer (TPV)

The Tivoli Performance Viewer (TPV) enables administrators and programmers to monitor the overall health of WebSphere Application Server without leaving the administrative console.

#### **Related reference**

Web services client to Web container optimized communication

To improve performance, there is an optimized communication path between a Web services client application and a Web container that are located in the same application server process. Requests from the Web services client that are normally sent to the Web container using a network connection are delivered directly to the Web container using an optimized local path. The local path is available because the Web services client application and the Web container are running in the same process.

`wsdeploy` command

**Web services performance best practices:** Web services are developed and deployed based on standards provided by the Web Services for Java 2 Platform, Enterprise Edition (J2EE) specification, as is the mechanism used to access a Web service. This article explains performance considerations for Web services supported by this specification.

When you develop or deploy a Web service, several artifacts are required, including a Web Services Description Language (WSDL) file. The WSDL file describes the format and syntax of the Web service input and output SOAP messages. When a Web service is implemented in the WebSphere Application Service runtime, the SOAP message is translated based on the J2EE request. The J2EE-based response is then translated back to a SOAP message.

The most critical performance consideration is the translation between the XML-based SOAP message and the Java object. Performance is high for a Web service implementation in WebSphere Application Server, however, application design, deployment and tuning can be improved. See *Monitoring the performance of Web services applications* for more information about analyzing and tuning Web services.

If you are using Web service application that was developed for a WebSphere Application Server version prior to Version 6.0, you can achieve better performance by running the `wsdeploy` command. The `wsdeploy` command regenerates Web services artifact classes to increase the serialization and deserialization performance.

## **Basic considerations for a high-performance Web services application**

The following are basic considerations you should know when designing a Web services application:

- Reduce the Web services requests by using a few highly functional APIs, rather than several simple APIs.
- Design your WSDL file interface to limit the size and complexity of SOAP messages.
- Use the document/literal style argument when you generate the WSDL file.
- Leverage the caching capabilities offered for WebSphere Application Server.
- Test the performance of your Web service.

## **Additional Web services performance features that you can leverage**

- In-process optimizations for Web services to optimize communication path between a Web services client application and a Web container that are located in the same application server process. For details and enabling this feature see *Web services client to Web container optimized communication*
- Access to Web services over multiple transport protocols extends existing JAX-RPC capabilities to support non-SOAP bindings such as RMI/IIOP and JMS. These alternative transports can improve performance and quality of service aspects for Web services. For more detailed information see *RMI-IIOP using JAX-RPC*.
- The SOAP with Attachments API for Java (SAAJ) 1.2 provides a low-level programming model for Web services relative to JAX-RPC. The SAAJ 1.2 API provides features to create and process SOAP requests using a low-level XML API. SAAJ 1.2 supports just-in-time parsing and other internal algorithms. For information about SAAJ 1.2 or Web services programming see *SOAP with Attachments API for Java*.

- The Web services tooling generates higher performance custom deserializers for all JAX-RPC beans. Redeploying a V5.x application into the V6 runtime can decrease the processing time for large messages.
- Serialization and deserialization runtime is enhanced to cache frequently used serializers and deserializers. This can decrease the processing time for large messages.
- The performance of WS-security encryption and digital signature validation is improved because of the use of the SAAJ 1.2 implementation.

IBM provides considerable documentation and best practices for Web services application design and development that details these items and more. For a list of key Web sites that discuss performance best practices, see *Web services: Resources for learning*.

## Service integration

### Setting tuning properties of a messaging engine

Use this task to set the tuning properties for a messaging engine.

You can set the following property to improve the performance of a messaging engine:

| Name          | Value                                                                                                                                                                                                                                                                        |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib.trm.retry | The messaging engine to messaging engine connection retry interval in seconds. The retry interval is the time delay left between attempts to contact neighboring messaging engines with which communications contact should exist. The default retry interval is 30 seconds. |

To set the tuning properties for a messaging engine, use the administrative console to complete the following steps:

1. In the navigation pane, click **Service integration** → **Buses** [Content Pane] → **bus\_name** → [Additional Properties] **Messaging engines** → **engine\_name** → [Additional Properties] **Custom properties**.
2. Type the name of the property that you want to set.
3. Type the value that you want to set for that property.
4. Click **OK**.
5. Save your changes to the master configuration.
6. Restart the messaging engine for the changes to take effect.

### Tuning and problem solving for messaging engine data stores

Use this task to obtain an overview of improving the performance of messaging engine data stores and understanding problems that can occur with a data store.

For more information about tuning and problem solving for messaging engine data stores, see the following topics:

- “Tuning the JDBC data source of a messaging engine” on page 70
- “Controlling the memory buffers used by a messaging engine” on page 70
- Sharing connections to benefit from one-phase commit optimization
- Diagnosing problems with data store exclusive access locks
- “Diagnosing problems with your data store configuration” on page 72

- “Avoiding failover problems when you use DB2 v8.2 with HADR as your data store” on page 73

### **Tuning the JDBC data source of a messaging engine:**

Use this task to tune the JDBC data source of a messaging engine for messaging performance on a service integration bus.

Consider whether you need to configure the connection pool for the JDBC data source to achieve your requirements for messaging performance.

The messaging engine uses the connection pool to obtain its connections to the database. With a heavy workload, a messaging engine might require a large number of concurrent connections to avoid delays waiting for connections to become available in the pool. For example, a very heavily loaded messaging engine might need 50 or more connections. Perform the following steps to configure the connection pool to meet your performance requirements:

1. Ensure that the configuration of your relational database management system (RDBMS) permits the number of connections that you require. Refer to the documentation for your RDBMS for more information.
2. Use the WebSphere administrative console to set the connection pool parameters. Navigate to the **General properties** for your data source and click **Additional properties** → **Connection pool properties**
  - a. Set the **Maximum connections** to the number of connections you require, for example, at least 50. The default number of connections is 10.

**Tip:** If your messaging engine times out when requesting a database connection, check the error log. If the error log contains error message CWSIS1522E, increase the number of connections and ensure that the configuration of your RDBMS permits that number of connections.

- b. Set the **Purge policy** to *EntirePool*. This policy enables the connection pool to release all connections when the messaging engine stops.

**Important:** You must set this value if the messaging engine can failover in a cluster.

### **Controlling the memory buffers used by a messaging engine:**

Use this task to control the sizes of the memory buffers used by a messaging engine.

Every messaging engine manages two memory buffers that contain messages and message-related data. You can set the following properties to improve the interaction of a messaging engine with its data store.

| Name                                   | Value                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib.msgstore.discardableDataBufferSize | <p>The size in bytes of the data buffer used by the messaging engine to contain data for which the quality of service attribute is best effort nonpersistent. The messaging engine holds this data entirely within this memory buffer and never writes this data to the data store. When the messaging engine adds data to this buffer, for example when the messaging engine receives a best effort nonpersistent message from a client, the messaging engine might discard data already in the buffer to make space. This behavior enables the messaging engine to discard best effort nonpersistent messages.</p> <p>The discardable data buffer contains all data for which the quality of service attribute is best effort nonpersistent. That data comprises data both that is involved in active transactions, and any other best effort nonpersistent that the messaging engine has neither discarded nor consumed. The messaging engine can discard only data that is not involved in active transactions.</p> <p><b>Tip:</b> If the messaging engine attempts to add data to the discardable data buffer when insufficient space remains after discarding all the data that is not involved in active transaction, the messaging engine throws a <code>com.ibm.ws.sib.msgstore.OutOfCacheSpace</code> exception. Client applications can catch this exception, wrapped inside API-specific exceptions such as <code>javax.jms.JMSEException</code>.</p> <p>The <code>sib.msgstore.discardableDataBufferSize</code> property of the messaging engine controls the size of the discardable data buffer. You specify the value of this property in bytes. The default value is 320000, which is approximately 320 kilobytes.</p> |

| Name                              | Value                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib.msgstore.cachedDataBufferSize | <p>The size in bytes of the data buffer used by the messaging engine to contain data for which the quality of service attribute is <i>better than</i> best effort nonpersistent and which is held in the data store. The purpose of the cached data buffer is to optimize the performance of the messaging engine by caching in memory the data that the messaging engine might otherwise need to read from the data store. As it writes data to the data store and reads from the data store, the messaging engine attempts to add that data to the cached data buffer. The messaging engine might discard data already in the buffer to make space.</p> <p>The <code>sib.msgstore.cachedDataBufferSize</code> property of the messaging engine controls the size of the cached data buffer. You specify the value of this property in bytes. The default value is 320000, which is approximately 320 kilobytes.</p> |
| sib.msgstore.transactionSendLimit | <p>The maximum number of operations that the messaging engine includes in each transaction. For example, each JMS send or receive is an operation that counts towards the transaction send limit. The default value is 100.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

**Attention:** The messaging engine uses approximate calculations to manage the data it holds in the memory buffers. Neither of the `DataBufferSize` properties gives an accurate indication of the amount of memory that the messaging engine consumes in the JVM heap. The messaging engine can consume considerably more heap storage than the `DataBufferSize` properties indicate.

To set the properties of a messaging engine to improve its interaction with its data store, use the administrative console to complete the following steps:

1. In the navigation pane, click **Service integration** → **Buses** [Content Pane] → **bus\_name** → [Additional Properties] **Messaging engines** → **engine\_name** → [Additional Properties] **Custom properties**.
2. Type the name of the property that you want to set.
3. Type the value that you want to set for that property.
4. Click **OK**.
5. Save your changes to the master configuration.

**Remember:** When you change any of these properties, the new values do not take effect until you restart the messaging engine.

#### **Diagnosing problems with your data store configuration:**

Use this task to diagnose problems that are caused by your data store configuration.

The following problems depend on the database that you use with your data store configuration and the level of that database:

- Examine this section if your messaging engine uses an Oracle 9i database for its data store and your messaging engine fails to start. If the messaging engine fails with the following message, where XXXXXXXX is the schema for the table, ensure that your Oracle installation is at 9.2.0.4, or higher:  
CWSIS1530E: The data type, 1,111, was found instead of the expected type, 2,004, for column, LONG\_DATA, in table, XXXXXXXX.SIB000.
- Examine this section if your messaging engine uses a Sybase database for its data store. When you create your Sybase server:
  - Ensure that you create the database server with a page size of at least 4k.
  - Ensure that you set the **lock scheme** property on your server to the value *datarows*. This avoids the possibility of a deadlock on the data store tables.
- Examine this section if your messaging engine uses an Informix database for its data store and the messaging engine is unable to access its data store. When you configure your messaging engine to use an Informix database, ensure that you specify the schema name in lower case. For a full description of the configuring procedure, refer to Configuring a messaging engine to use a data source.

### **Avoiding failover problems when you use DB2 v8.2 with HADR as your data store:**

Use this task to avoid problems that can occur when a messaging engine that is configured to use DB2 v8.2 with HADR for its data store terminates if the DB2 database fails over.

If you configure a messaging engine to use DB2 v8.2 with HADR for its data store and a DB2 failover occurs, the application server on which the messaging engine is running might stop. The cause of this problem is that the messaging engine does not treat the DB2 failover as a transient communications error.

When you configure a messaging engine to use DB2 v8.2 with HADR for its data store ensure that the messaging engine can restart automatically following an application server failure. Choose the option that is appropriate for your configuration:

- If you are running a single server, WebSphere Application Server provides no failover support. Consider other high availability provisions.
- If you are running the Network Deployment version without clustering, the default configuration for the node agent ensures automatic restart.
- If you are running the Network Deployment version with clustering, peer recovery restarts the messaging engine. Ensure that you have configured the high availability policy to enable peer recovery.

### **Setting tuning properties for a mediation**

Use this task to tune a mediation for performance using the administrative console.

Before you begin this task, you should review the guidance on when it is appropriate to tune a mediation for performance in the topic Guidance for tuning mediations for performance.

You can set the following tuning properties to improve the performance of a mediation:

| Name                       | Value                                                                                                                                                                                                                                                                                                       |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib:SkipCloneBeforeMediate | Whether you want to omit the cloning of messages before they are processed by the mediation. Either true or false.                                                                                                                                                                                          |
| sib:SkipWellFormedCheck    | Whether you want to omit the well formed check that is performed on messages after they have been processed by the mediation. Either true or false.<br><b>Note:</b> This property is overridden for messages that have the delivery option assured persistent, and a well formed check is always performed. |

You can set the following tuning property to improve the performance of a mediation:

| Name                    | Value                                                                                                                                                                                                                                                                                                       |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib:SkipWellFormedCheck | Whether you want to omit the well formed check that is performed on messages after they have been processed by the mediation. Either true or false.<br><b>Note:</b> This property is overridden for messages that have the delivery option assured persistent, and a well formed check is always performed. |

To set, or unset, one or more tuning properties for a mediation, use the administrative console to complete the following steps:

1. Display the mediation context information:
  - a. In the navigation pane, click **Service integration** → **Buses**
  - b. In the content pane, click the name of the service integration bus.
  - c. Under **Additional Properties**, click **Mediations**.
  - d. In the content pane, select the name of the mediation for which you want to configure tuning information.
  - e. Under **Additional Properties**, click **Context information**.
2. In the content pane, click **New**.
3. Type the name of the property in the **Name** field.
4. Select the type `Boolean` in the list box.
5. Type **true** in the **Context Value** field to set the property, or type **false** to unset the property.
6. Click **OK**.
7. Save your changes to the master configuration.

### Enabling CMP entity beans and messaging engine data stores to share database connections

Use this task to enable container-managed persistence (CMP) entity beans to share the database connections used by the data store of a messaging engine. This has been estimated as a potential performance improvement of 15% for overall message throughput, but can only be used for entity beans connected to the application server that contains the messaging engine.



To enable CMP entity beans to share the database connections used by the data store of a messaging engine, complete the following steps:

1. Configure the data store to use a data source that is not XA-capable. For more information about configuring a data store, see *Configuring a JDBC data source*.
2. Select the Share data source with CMP option.

This option is provided on the JMS connection factory used to connect to the service integration bus that hosts the bus destination that is used to store and process messages for the CMP bean.

This option is provided on the JMS connection factory or JMS activation specification used to connect to the service integration bus that hosts the bus destination that is used to store and process messages for the CMP bean.

For example, to select the option on a unified JMS connection factory, complete the following steps:

- a. Display the default messaging provider. In the navigation pane, expand **Resources** → **JMS Providers** → **Default Messaging Provider**.
- b. Optional: Change the **Scope** check box to set the level at which the connection factory is to be visible, according to your needs.
- c. In the content pane, under Additional Properties, click **JMS connection factories**
- d. Optional: To create a new unified JMS connection factory, click **New**.

Specify the following properties for the connection factory:

**Name** Type the name by which the connection factory is known for administrative purposes.

**JNDI name**

Type the JNDI name that is used to bind the connection factory into the name space.

**Bus name**

Type the name of the service integration bus that the connection factory is to create connections to. This service integration bus hosts the destinations that the JMS queues and topics are assigned to.

- e. Optional: To change the properties of an existing connection factory, click one of the connection factories displayed. This displays the properties for the connection factory in the content pane.
- f. Select the check box for the Share data source with CMP field
- g. Click **OK**.
- h. Save your changes to the master configuration.

The JMS connection factory can only be used to connect to a “local” messaging engine that is in the application server on which the CMP beans are deployed.

3. Deploy the CMP beans onto the application server that contains the messaging engine, and specify the same data source as used by the messaging engine. You can use the administrative consoles to complete the following steps:
  - a. Optional: To determine the data source used by the messaging engine, click **Servers** → **Application servers** → **server\_name** → **Messaging engines** → **engine\_name** → **Data store** The **Data source name** field displays the name of the data source; by default:  
`jdbc/com.ibm.ws.sib/engine_name`
  - b. Click **Applications** → **Install New Application**

- c. On the first Preparing for application install page, specify the full path name of the source application file (.ear file otherwise known as an EAR file), then click **Next**
- d. On the second Preparing for application install page, complete the following steps:
  - 1) Select the check box for the Generate Default Bindings property. Data source bindings (for EJB 1.1 JAR files) are generated based on the JNDI name, data source user name password options. This results in default data source settings for each EJB JAR file. No bean-level data source bindings are generated.
  - 2) Under Connection Factory Bindings, click the check box for the **Default connection factory bindings:** property, then type the JNDI name for the data source and optionally select a **Resource authorization** value.
  - 3) Click **Next**
4. If your application uses EJB modules that contain Container Managed Persistence (CMP) beans that are based on the EJB 1.x specification, for Step: Provide default datasource mapping for modules containing 1.x entity beans, specify a JNDI name for the default data source for the EJB modules. The default data source for the EJB modules is optional if data sources are specified for individual CMP beans.
5. If your application has CMP beans that are based on the EJB 1.x specification, for Step: Map datasources for all 1.x CMP, specify a JNDI name for data sources to be used for each of the 1.x CMP beans. The data source attribute is optional for individual CMP beans if a default data source is specified for the EJB module that contains CMP beans. If neither a default data source for the EJB module nor a data source for individual CMP beans are specified, then a validation error displays after you click Finish (step 13) and the installation is cancelled.
6. Complete other panels as needed.
7. On the Summary panel, verify the cell, node, and server onto which the application modules will install:
  - a. Beside Cell/Node/Server, click **Click here**.
  - b. Verify the settings on the Map modules to servers page displayed. Ensure that the application server that is specified contains the messaging engine and its data store.
  - c. Specify the Web servers as targets that will serve as routers for requests to this application. This information is used to generate the plug-in configuration file (plugin-cfg.xml) for each Web server.
  - d. Click **Finish**.

For more information about installing applications, see Installing application files with the console.

## Tuning service integration messaging

Use this task to set tuning properties that control the performance of message-driven beans and other messaging applications deployed to use service integration technologies.

To optimize the performance of messaging with service integration technologies, such as message-driven beans that use the default messaging provider, you can use the following parameters set through the WebSphere administrative console or command line interfaces.

On z/OS, the performance of messaging applications is affected by the number of servants, which can vary dynamically, and the distribution of work between servants. For more information about configuring and managing the number of servants and the distribution of work between servants, see Tuning the application serving environment.

- Providing sufficient threads to maximize the throughput of messages.

There is a trade-off between providing sufficient threads to maximize the throughput of messages and configuring excessive threads, which can lead to CPU starvation of the threads in the application server. If you notice that the throughput for express nonpersistent, reliable nonpersistent, or reliable persistent messaging has fallen as a result of increasing the size of the default thread pool, then you should decrease the size of the thread pool and reassess the message throughput.

1. To be able to view the available message count for a queue, you need to enable runtime AvailableMessageCount statistics for the queue. If you restart administrative server, you need to enable AvailableMessageCount statistics again because such runtime settings are not preserved when the server is restarted.

To enable AvailableMessageCount statistics using the administrative console, complete the following steps:

- a. In the navigation pane, click **Monitoring and Tuning** → **Performance Monitoring Infrastructure (PMI)**
  - b. In the content pane, click *server\_name*
  - c. Click the Runtime tab.
  - d. In the Currently monitored statistic set, click **Custom**
  - e. On the Custom monitoring level panel, click **SIB Service** → **SIB Messaging Engines** → *messageEngine\_name* → **Destinations** → **Queues** → *queue\_name*
  - f. Select the AvailableMessageCount option.
  - g. Click the **Enable** button at the top of the panel.
2. To view the available message count, you can use the administrative console to complete the following steps:

- a. In the navigation pane, click **Monitoring and Tuning** → **Performance Viewer** → **Current activity**
- b. In the content pane, click *server\_name*
- c. Click **Performance Modules** → **SIB Service** → **SIB Messaging Engines** → *messageEngine\_name* → **Destinations** → **Queues** → *queue\_name*
- d. Click the **View Module(s)** button at the top of the Resource Selection panel, located on the left side. This displays the AvailableMessageCount data in the Data Monitoring panel, located on the right side.

You can use the Data Monitoring panel to manage the collection of monitoring data; for example, you can use the buttons to start or stop logging, or to change the data displayed as either a table or graph.

3. To view or change the number of threads in the default thread pool for an application server, you can use the administrative console to complete the following steps:
  - a. In the navigation pane, click **Servers** → **Application servers**
  - b. In the content pane, click *server\_name*
  - c. Under Additional properties, click **Thread Pools** → **Default**. By default the Minimum size value is set to 5 and the Maximum size value is set to 20. The best performance is obtained by setting the Maximum size value to

the expected maximum concurrency for all message-driven beans. For high throughput using a single message-driven bean, 41 was found to be the optimal Maximum size value.

- d. To change the Maximum size value, type the new value in the Maximum size field then click **OK**. Finally, save your changes to the master configuration.
- Reducing the number of OutOfCacheSpace errors in the SystemOut.log file.  
OutOfCacheSpace errors in the SystemOut.log file indicate that the discardable data buffer used by the messaging engine is overflowing. For best effort nonpersistent messages, the messaging engine starts to discard messages when this buffer is full. You can increase the size of this data buffer to allow more best effort nonpersistent data to be handled before the messaging engine begins to discard the messages.

For more information about tuning the size of the discardable data buffer, set by the `sib.msgstore.discardableDataBufferSize` property of a messaging engine, see “Controlling the memory buffers used by a messaging engine” on page 70.

- Tuning the messaging engine data store (not required for Cloudscape).  
To improve the performance of messaging throughput of a messaging engine data store, you can tune the JDBC connection pool and statement cache size. In tests of high throughput MDB workloads, the following changes provided a 10% gain in throughput.

1. The messaging engine uses a connection pool for managing the JDBC connections to its data store. Tuning the size of the pool can improve the messaging throughput.

To view or change the size of the connection pool, you can use the administrative console to complete the following steps:

- a. In the navigation pane, click **Resources** → **JDBC Providers**
- b. In the content pane, click *jdbc\_provider\_name*
- c. Under Additional properties, click **Data sources** → *data\_source\_name*
- d. Under Additional properties, click **Connection pool properties**
- e. View the Maximum connections property and the Minimum connections property. By default, these properties are set to Maximum connections=10 and Minimum connections=1. Setting the value of both these properties to 50 is recommended. For especially high throughput workloads, setting the value of both these properties up to 100 can be beneficial. You may need to configure the underlying database to accept this many concurrent connections.
- f. To change the value of a property, type a new value in the property field then click **OK**. Finally, save your changes to the master configuration.

2. The statement cache contains recently used prepared statements to remove the costs associated with repeated preparation of statements. Tuning the size of the cache helps prevent useful entries from being discarded to make room for new entries.

To view or change the size of the statement cache, you can use the administrative console to complete the following steps:

- a. In the navigation pane, click **Resources** → **JDBC Providers**
- b. In the content pane, click *jdbc\_provider\_name*
- c. Under Additional properties, click **Data sources** → *data\_source\_name*
- d. Under Additional properties, click **WebSphere Application Server data source properties**

- e. View the Statement cache size property. By default, the value of this property is set to 10. For high throughput JMS messaging, a value of 40 is recommended.
  - f. To change the value of the property, type a new value in the property field then click **OK**. Finally, save your changes to the master configuration.
- Using a remote DB2 Universal Database.  
By default, messaging engines are configured to use Cloudscape as the underlying data store database. Changing to a DB2 database, especially a remote database, significantly improves performance. This is of particular benefit on systems with more than two CPUs. A remote DB2 database provides a 50% higher throughput than a local DB2 database, for the same total (application server plus database) CPU utilization.
  - Tuning reliability levels for messages.  
The reliability level chosen for the messages has a significant impact on performance. In order of decreasing performance (fastest first), the reliability levels are: Best-Effort Nonpersistent, Express Nonpersistent, Reliable Nonpersistent, Reliable Persistent, and Assured Persistent. For MDB point-to-point messaging, best-effort nonpersistent throughput is more than 6 times greater than assured persistent.  
For more information about reliability levels, see [Message reliability levels](#). For the relative performance of the different reliability levels, see the [WebSphere Application Server performance report](#).
  - Tuning MDB performance with the default messaging provider.  
For information about tuning the throttling of message-driven beans, including controlling the maximum number of instances of each message-driven bean and the message batch size for serial delivery, see [Configuring MDB throttling on the default messaging provider](#).

## Tuning the SIBWS

You can use the administrative console or a Jacl script to tune performance settings for the service integration bus Web services enablement (SIBWS).

To improve the performance of the SIBWS, you can tune the following parameters:

- The Java virtual machine heap size (to help ensure there is enough memory available to process large messages, or messages with large attachments).
- The maximum permitted number of concurrent clients. (That is, the maximum number of instances of a message-driven bean that are permitted by the activation specification for the service integration technologies resource adapter).
- The maximum batch size for batches of messages to be delivered to a client. (By default, only a single message is delivered to a message-driven bean instance at one time. You can improve performance by allowing messages to be sent in batches to a message-driven bean).
- The number of threads available to service requests for each client. (That is, the number of threads available in the default thread pool and the Web container thread pool for a given application server).

To tune the SIBWS, complete one of the following two main steps:

- **Optional:** To use the administrative console to tune the SIBWS, complete the following steps:
  1. Use the topic [Tuning Java virtual machines](#) to set the JVM heap size to a larger value than the default value (256 megabytes). The value should generally be as large as possible without incurring paging.

2. Use the topic Tuning service integration messaging to tune the maximum number of instances of a message-driven bean, the maximum batch size for batches of messages for a bean, and the number of threads available to service requests for a bean:
  - In the administrative console navigation pane, click **Resources** → **Resource adapters** → **SIB JMS Resource Adapter** → **J2C activation specification** → **SIBWS\_OUTBOUND\_MDB**, then set `maxConcurrency` to  $n$  and `maxBatchSize` to 5 (where  $n$  is the maximum permitted number of instances of a message-driven bean).
  - In the administrative console navigation pane, click **Servers** → **Application servers** → `server_name` → **Thread Pools**, then set the number of threads in the default and WebContainer thread pools to a maximum of  $n-1$  threads and a minimum of some figure approaching the maximum (to reduce lock contention).
- To use a Jacl script to tune the SIBWS, create and run a script based on the following example script:

```
puts "****Configuring server thread pools ****"

if {[llength $argv] !=1} {
 puts "ERROR: Incorrect usage."
 puts ""
 puts {Usage: wsadmin -f configure-soapservice1.jacl serverName}
 puts ""
 puts "Where:"
 puts ""
 puts " serverName = name of server"
 exit
}

#-----
#--- Setup the arguments
#-----

#--- Get args
set serverName [lindex $argv 0]

#-----
#--- Set the thread pool sizes and the JVM heap size
#-----
set serverId [$AdminConfig getid /Server:$serverName/]

#--- First the threadPools...
puts "Modifying Server ThreadPool Sizes for $serverName"

set inactivityTimeout [list inactivityTimeout 120000]
set isGrowable [list isGrowable false]
set maximumSize [list maximumSize 150]
set minimumSize [list minimumSize 145]

set threadPoolList [list $inactivityTimeout \
 $isGrowable \
 $maximumSize \
 $minimumSize \
]
set threadpools [$AdminConfig list ThreadPool $serverId]

foreach pool $threadpools {
 set modifyToggle 0
 #puts [$AdminConfig showAttribute $pool name]
 if {[string compare [$AdminConfig showAttribute $pool name] WebContainer] == 0} {
 set modifyToggle 1
 #puts "WebContainer ThreadPool found for server $serverName"
 }
}
```

```

 if {[string compare [$AdminConfig showAttribute $pool name] Default] == 0} {
set modifyToggle 1
#puts "Default ThreadPool found for server $serverName"
 }
 if {$modifyToggle == 1} {
puts "modifying threadpool [$AdminConfig showAttribute $pool name]"
$AdminConfig modify $pool $threadPoolList
 }
}
puts "ThreadPool modifications complete"

#--- Now the JVM heap size
set heapSize 1536
puts "Modifying JVM Heap size to $heapSize for server $serverName"

set initialHeapSize [list initialHeapSize $heapSize]
set maximumHeapSize [list maximumHeapSize $heapSize]

set jvmModifications [list $initialHeapSize $maximumHeapSize]

set processDef [$AdminConfig list ProcessDef $serverId]
set jvm [$AdminConfig list JavaVirtualMachine $processDef]

$AdminConfig modify $jvm $jvmModifications
puts "Heap size modification complete"

#-----
#--- Modify the J2CActivationSpec for the service integration technologies Resource adapter
#-----
puts "Modifying the J2CActivationSpec"

set actSpec [$AdminConfig getid /J2CActivationSpec:SIBWS_OUTBOUND_MDB/]
set propSet [$AdminConfig showAttribute $actSpec resourceProperties]

set propSet [lindex $propSet 0]

set maxConcurrency [list value 200]
set maxConcurrency [list $maxConcurrency]

set maxBatchSize [list value 5]
set maxBatchSize [list $maxBatchSize]

foreach propId $propSet {
 if { [string compare [$AdminConfig showAttribute $propId name] maxConcurrency] == 0} {
$AdminConfig modify $propId $maxConcurrency
puts "Custom property changed : [$AdminConfig showall $propId] "
 }
 if { [string compare [$AdminConfig showAttribute $propId name] maxBatchSize] == 0} {
$AdminConfig modify $propId $maxBatchSize
puts "Custom property changed : [$AdminConfig showall $propId] "
 }
}

puts "J2CActivationSpec modifications complete"

puts "Saving actions..."
$AdminConfig save

```

## Data access resources

### EJB Container Cache tuning

Monitoring Tivoli Performance Viewer (TPV) is a great way to diagnose if the EJB Container Cache size setting is tuned correctly for your application. If the

application has filled the cache causing evictions to occur, TPV will show a very high rate of `ejbStores()` being called and probably a lower than expected CPU utilization on the application server machine.

All applications using enterprise beans should have this setting adjusted from the default if the following formula works out to more than 2000.

$$\begin{aligned} \text{EJB\_Cache\_Size} = & (\text{Largest number of Option B or C Entity Beans enlisted in a} \\ & \text{transaction} * \text{maximum number of concurrent transactions}) + \\ & (\text{Largest number of unique Option A Entity Beans expected to be accessed during} \\ & \text{typical application workload}) + \\ & (\text{Number of stateful Session Beans active during typical workload}) + \\ & (\text{Number of stateless SessionBean types used during typical workload}) \end{aligned}$$

Where:

Option B and C Entity Beans are only held in the EJB cache during the lifetime of the transaction they are enlisted in. Therefore, the first term in the formula computes the average EJB cache requirements for these types of beans.

Option A Entity Beans are held in the EJB cache indefinitely, and are only removed from the cache if there start to become more beans in the cache than the cache size has been set to.

Stateful Session Beans are held in the EJB cache until they are removed by the application, or their session timeout value is reached.

Only a single stateless Session Bean instance for each EJB type is held in the cache during the time any methods are being executed on that stateless Session Bean. If two or more methods are being executed simultaneously on the same stateless Session Bean type, each method executes on its own bean instance, but only one cache location is used for all of these instances.

This calculates the upper bound on the maximum possible number of enterprise beans active at one time inside the application server. Because the EJB Containers cache is built to contain all these beans for performance optimizations, best performance can be achieved by setting this cache size to be larger than the number resulting from the calculation above.

<tuning parameter>

This setting can be found under Servers > Application Servers > serverName > EJB Container > EJB Cache Settings

Also while adjusting the EJB Cache Size, the EJB Container management thread parameter can be tuned to meet the needs of the application. The management thread is controlled through the Clean Up Interval setting. This setting controls how frequently a daemon thread inside of WebSphere Application Server wakes up and attempts to remove bean instances from the cache that have not been used recently, attempting to keep the number of bean instances at or below the cache size. This allows the EJB container to place and look up items in the cache as quickly as possible. It normally is best to leave this interval set to the default, however, in some cases, it may be worthwhile to see if there is a benefit to reducing this interval.

## EJB Container Pool Size

If the application is using the majority of the instances in the pool, TPV indicates this. When this occurs, then the size of those bean pools that are being exhausted should be increased. This can be done by adding the following parameter in the JVM's custom properties tag .

```
-Dcom.ibm.websphere.ejbcontainer.poolSize=<application_name>#<module_name>#<enterprisebean_name>=<minSize>,<maxSize>
```



where:

<application\_name> is the J2EE application name as defined in the application archive (.ear) file deployment descriptor, for the bean whose pool size is being set

<module\_name> is the .jar file name of the EJB module, for the bean whose pool size is being set,

<bean\_name> is the J2EE Enterprise Bean name as defined in the EJB module deployment descriptor, for the bean whose pool size is being set

<minSize> is the number of bean instances the container maintains in the pool, irrespective of how long the beans have been in the pool (beans greater than this number are cleared from the pool over time to optimize memory usage)

<maxSize> is the number of bean instances in the pool where no more bean instances are placed in the pool after they are used (that is, once the pool is at this size, any additional beans are discarded rather than added into the pool -- this ensures the number of beans in the pool has an upper limit so memory usage does not grow in an unbounded fashion).

To keep the number of instances in the pool at a fixed size, minSize and maxSize can be set to the same number. Note that there is a separate instance pool for every EJB type running in the application server, and that every pool starts out with no instances in it - that is, the number of instances grows as beans are used and then placed in the pool. When a bean instance is needed by the container and no beans are available in the pool, the container creates a new bean instance, uses it, then places that instance in the pool (unless there are already maxSize instances in the pool).

For example, the statement

```
-Dcom.ibm.websphere.ejbcontainer.poolSize=ivtApp#ivtEJB.jar#ivtEJBObject=125,1327
```

would set a minSize of 125 and a maxSize of 1327 on the bean named "ivtEJBObject" within the ivtEJB.jar file, in the application "ivtApp".

Where ivtApp is replaced by the actual application name, ivtEJB.jar is replaced by the jar containing the bean that needs to have its pool size increased, and ivtEJBObject is the bean name of the enterprise bean whose pool size should be increased. The 125,1327 is the minimum and maximum number of beans that will be held in the pool. These should be set so no more evictions occur from the pool and in most cases should be set equal if memory is plentiful because no growth and shrinkage of the pool will occur.

## EJB Container Primary Key Mutation

Application developers and administrators should have a good idea of how their application handles the creation of primary key objects for use by container-managed persistence (CMP) beans and bean-managed persistence (BMP) beans inside of WebSphere Application Server. The IBM EJB Container uses the primary key of an Entity bean as an identifier inside of many internal data structures to optimize performance. However, the EJB Container must copy these primary key objects upon the first access to the bean to ensure that the objects stored in the internal caches are separate from the ones used in an application, in case the application changes or mutates the primary key, to keep the internal structures consistent.

If the application does not mutate any of the primary keys used to create and access entity beans after they are created, then a special flag can be used that allows the EJB Container to skip the copy of the primary key object, thus saving CPU cycles and increasing performance. This mechanism can be enabled *at your own risk* by adding the following `-D` property to the JVM custom property field.

```
<tuning parameter>
-Dcom.ibm.websphere.ejbcontainer.noPrimaryKeyMutation=true
```

The performance benefit of this optimization depends on the application. If the application uses primitive types for enterprise beans' primary keys there will be no gain because these objects are already immutable and the copy mechanism takes this into account. If, however, the application uses many complex primary keys (that is, And object for a primary key or multiple fields) then this parameter can yield significant improvements.

### **Persistence Manager Deferred Insert on EJB Create**

The IBM Persistence manager is used by the EJB Container to persist data to the database from CMP entity beans. When creating entity beans by calling the `ejbCreate()` method, by default the Persistence manager immediately inserts the empty row with only the primary key in the database. In most cases applications, after creating the bean, modify fields in the bean created or in other beans inside of the same transaction. If the user wishes to postpone the insert into the database until the end of the transaction, so that it will eliminate one trip to the database, they may set this `-D` flag inside of the JVM custom properties field. The data will still be inserted into the database and consistency will be maintained.

```
<tuning parameter>
-Dcom.ibm.ws.pm.deferredcreate=true
```

The performance benefit of this optimization depends on the application. If the EJB applications transactions are very insert intensive the application could benefit largely from this optimization. If the application performs very few inserts then the benefit of this optimization will be much less.

### **Persistence Manager Database Batch Update on EJB Update**

When an EJB application accesses multiple CMP beans inside of a single transaction, depending on the operations performed on the beans (updates, inserts, reads), the number of operations issued to the database will correspond directly to the operations performed on the CMP beans. If the database system you are using supports batching of update statements you can enable this flag and gain a performance boost on all interactions with the database that involve more than two updates in a single transaction. This flag will let the persistence manager add all the update statements into one single batch statement which will then be issued to the database. This saves round trips to the database, thus increasing performance. If the user knows their application exhibits the behavior of updating multiple CMP beans in a single transaction and the database supports batch updates they may set this `-D` flag inside of the JVM custom properties field.

```
<tuning parameter>
-Dcom.ibm.ws.pm.batch=true
```

The performance benefit of this optimization depends on the application. If the application never or infrequently updates CMP beans or only updates a single bean per transaction there will be no performance gain. If the application updates multiple beans per transaction then this parameter will benefit your applications performance.

### **Database performance tuning**

Database performance tuning can dramatically affect the throughput of your application. For example, if your application requires high concurrency (multiple, simultaneous interactions with backend data), an improperly tuned database can

result in a bottleneck. Database access threads accumulate in a backlog when the database is not configured to accept a sufficient number of incoming requests.

Because WebSphere Application Server supports the integration of many different database products, each one with unique tuning configurations, consult your database vendor documentation for comprehensive tuning information. This information center provides introductory material on DB2 tuning parameters for your convenience.

Tuning parameters vary according to the type of database you are using. DB2 tuning tips for z/OS are provided for your convenience. Because DB2 is not a WebSphere Application Server product and can change, consider these descriptions as suggestions.

Tuning parameters vary according to the type of database you are using. If you run DB2 UDB for iSeries, consult the DB2 UDB tuning parameters article as a starting point reference.

### **DB2 tuning parameters**

DB2 has many parameters that you can configure to optimize database performance. For complete DB2 tuning information, refer to the *DB2 UDB Administration Guide: Performance* document.

#### **DB2 logging**

- **Description:** DB2 has corresponding log files for each database that provides services to administrators, including viewing database access and the number of connections. For systems with multiple hard disk drives, you can gain large performance improvements by setting the log files for each database on a different hard drive from the database files.
- **How to view or set:** At a DB2 command prompt, issue the command: `db2 update db cfg for [database_name] using newlogpath [fully_qualified_path]`.
- **Default value:** Logs reside on the same disk as the database.
- **Recommended value:** Use a separate high-speed drive, preferably performance enhanced through a redundant array of independent disk (RAID) configuration.

For more information about using AIX with DB2 see “Tuning AIX systems” on page 45.

#### **DB2 configuration advisor**

Located in the DB2 Control Center, this advisor calculates and displays recommended values for the DB2 buffer pool size, the database, and the database manager configuration parameters, with the option of applying these values. See more information about the advisor in the online help facility within the Control Center.

#### **Number of connections to DB2 - MaxAppls and MaxAgents**

When configuring the data source settings for the databases, confirm the DB2 MaxAppls setting is greater than the maximum number of connections for the data source. If you are planning to establish clones, set the MaxAppls value as the maximum number of connections multiplied by the number of clones. The same relationship applies to the session manager number of connections. The MaxAppls setting must be equal to or greater than the number of connections. If you are

using the same database for session and data sources, set the MaxAppls value as the sum of the number of connection settings for the session manager and the data sources.

For example, MaxAppls = (number of connections set for the data source + number of connections in the session manager) multiplied by the number of clones.

After calculating the MaxAppls settings for the WebSphere Application Server database and each of the application databases, verify that the MaxAgents setting for DB2 is equal to or greater than the sum of all of the MaxAppls values. For example, MaxAgents = sum of MaxAppls for all databases.

### DB2 buffpage

- **Description:** Improves database system performance. Buffpage is a database configuration parameter. A buffer pool is a memory storage area where database pages containing table rows or index entries are temporarily read and changed. Data is accessed much faster from memory than from disk.
- **How to view or set:** To view the current value of buffpage for database *x*, issue the DB2 command `get db cfg for x` and look for the value **BUFFPAGE**. To set **BUFFPAGE** to a value of *n*, issue the DB2 command `update db cfg for x using BUFFPAGE n` and set **NPAGES** to -1 as follows:

```
db2 <-- go to DB2 command mode, otherwise the following "select" does not work as is
connect to x <-- (where x is the particular DB2 database name)
select * from syscat.bufferpools
 (and note the name of the default, perhaps: IBMDEFAULTBP)
 (if NPAGES is already -1, there is no need to issue following command)
alter bufferpool IBMDEFAULTBP size -1
 (re-issue the above "select" and NPAGES now equals -1)
```

You can collect a snapshot of the database while the application is running and calculate the buffer pool hit ratio as follows:

1. Collect the snapshot:
    - a. Issue the **update monitor switches using bufferpool on** command.
    - b. Make sure that bufferpool monitoring is on by issuing the **get monitor switches** command.
    - c. Clear the monitor counters with the **reset monitor all** command.
  2. Run the application.
  3. Issue the **get snapshot for all databases** command before all applications disconnect from the database, otherwise statistics are lost.
  4. Issue the **update monitor switches using bufferpool off** command.
  5. Calculate the hit ratio by looking at the following database snapshot statistics:
    - Buffer pool data logical reads
    - Buffer pool data physical reads
    - Buffer pool index logical reads
    - Buffer pool index physical reads
- **Default value:** 250
  - **Recommended value:** Continue increasing the value until the snapshot shows a satisfactory hit rate.

The buffer pool hit ratio indicates the percentage of time that the database manager did not need to load a page from disk to service a page request. That is, the page is already in the buffer pool. The greater the buffer pool hit ratio, the lower the frequency of disk input and output. Calculate the buffer pool hit ratio as follows:

- $P = \text{buffer pool data physical reads} + \text{buffer pool index physical reads}$
- $L = \text{buffer pool data logical reads} + \text{buffer pool index logical reads}$

- Hit ratio =  $(1-(P/L)) * 100\%$

### DB2 query optimization level

- **Description:** Sets the amount of work and resources that DB2 puts into optimizing the access plan. When a database query runs in DB2, various methods are used to calculate the most efficient access plan. The range is from 0 to 9. An optimization level of 9 causes DB2 to devote a lot of time and all of its available statistics to optimizing the access plan.
- **How to view or set:** The optimization level is set on individual databases and can be set with either the command line or with the DB2 Control Center. Static SQL statements use the optimization level that is specified on the **prep** and **bind** commands. If the optimization level is not specified, DB2 uses the default optimization as specified by the `dft_queryopt` setting. Dynamic SQL statements use the optimization class that is specified by the current query optimization special register, which is set using the SQL Set statement. For example, the following statement sets the optimization class to 1:

```
Set current query optimization = 1
```

If the current query optimization register is not set, dynamic statements are bound using the default query optimization class.

- **Default value:** 5
- **Recommended value:** Set the optimization level for the needs of the application. Use high levels only when there are very complicated queries.

### DB2 reorgchk

- **Description:** Obtains the current statistics for data and rebinding. Use this parameter because SQL statement performance can deteriorate after many updates, deletes or inserts.
- **How to view or set:** Use the **DB2 reorgchk update statistics on table all** command to perform the **runstats** operation on all user and system tables for the database to which you are currently connected. Rebind packages using the **bind** command. If statistics are available, issue the **db2 -v "select tname, nleaf, nlevels, stats\_time from sysibm.sysindexes"** command on DB2 CLP. If no statistic updates exist, `nleaf` and `nlevels` are -1, and `stats_time` has an empty entry (for example: "-"). If the **runstats** command was previously run, the real-time stamp from completion of the **runstats** operation also displays under `stats_time`. If you think the time shown for the previous **runstats** operation is too old, run the **runstats** command again.
- **Default value:** None
- **Recommended value:** None

### DB2 locktimeout

- **Description:** Specifies the number of seconds that an application waits to obtain a lock. Setting this property helps avoid global deadlocks for applications.
- **How to view or set:** To view the current value of the lock timeout property for database `xxxxxx`, issue the **DB2 get db cfg for xxxxxx** command and look for the value, `LOCKTIMEOUT`. To set `LOCKTIMEOUT` to a value of `n`, issue the **DB2 update db cfg for xxxxxx** command using **LOCKTIMEOUT n**, where `xxxxxx` is the name of the application database and `n` is a value between 0 and 30 000 inclusive.
- **Default value:** -1, meaning lock timeout detection is turned off. In this situation, an application waits for a lock if one is not available at the time of the request, until either of the following events occurs:
  - The lock is granted
  - A deadlock occurs

- **Recommended value:** If your database access pattern tends toward a majority of writes, set this value so that it gives you early warning when a timeout occurs. A setting of 30 seconds suits this purpose. If your pattern tends toward a majority of reads, either accept the default lock timeout value, or set the property to a value greater than 30 seconds.

#### DB2 maxlocks

- **Description:** Specifies the percentage of the lock list that is reached when the database manager performs escalation, from row to table, for the locks held by the application. Although the escalation process does not take much time, locking entire tables versus individual rows decreases concurrency, and potentially decreases overall database performance for subsequent attempts to access the affected tables.
- **How to view or set:** To view the current value of the maxlocks property for database *xxxxxx*, issue the DB2 **get db cfg for xxxxxx** command and look for the MAXLOCKS value. To set MAXLOCKS to a value of *n*, issue the DB2 **update db cfg for xxxxxx** command using **MAXLOCKS n**, where *xxxxxx* is the name of the application database and *n* is a value between 1 and 100 inclusive.
- **Default value:** Refer to the current database information for property default values per operating system.
- **Recommended value:** If lock escalations are causing performance concerns, you might need to increase the value of this parameter or the locklist parameter, which is described in the following paragraph. You can use the database system monitor to determine if lock escalations are occurring.

#### DB2 locklist

- **Description:** Specifies the amount of storage that is allocated to the lock list.
- **How to view or set:** To view the current value of the locklist property for database *xxxxxx*, issue the DB2 **get db cfg for xxxxxx** command and look for the LOCKLIST value. To set LOCKLIST to a value of *n*, issue the DB2 **update db cfg for xxxxxx** command using **LOCKLIST n**, where *xxxxxx* is the name of the application database and *n* is a value between 4 and 60 000 inclusive.
- **Default value:** Refer to the current database information for property default values per operating system.
- **Recommended value:** If lock escalations are causing performance concerns, you might need to increase the value of this parameter or the maxlocks parameter, which is described in the previous paragraph. You can use the database system monitor to determine if lock escalations are occurring. Refer to the *DB2 Administration Guide: Performance* document for more details.

#### Related tasks

Chapter 5, “Tuning the application serving environment,” on page 25  
This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

#### Related information

<http://www-306.ibm.com/software/sw-library/literature/db2/>

<http://www-306.ibm.com/software/data/db2/udb/support/manualsv8/>

### Tuning parameters for data access resources

For better application performance, you can tune some data access resources through the WebSphere Application Server administrative console.

Tune these properties of data sources and connection pools to optimize the performance of transactions between your application and datastore.

## Data source tuning

To view the administrative console page where you configure the following properties, click **Resources > JDBC Providers > JDBC\_provider > Data sources > data\_source > WebSphere Application Server connection properties**.

### Enable JMS one phase optimization support

If your application does not use JMS messaging, **do not** select this option. Activating this support enables the Java Message Service (JMS) to get optimized connections from the data source. Activating this support also *prevents* JDBC applications from obtaining connections from the data source. For further explanation of JMS one phase support, refer to the article entitled "Sharing connections to benefit from one phase commit optimization" in this information center.

### Statement cache size

Specifies the number of statements that can be cached per connection.

The WebSphere Application Server data source optimizes the processing of *prepared statements* and *callable statements* by caching those statements that are not being used in an active connection. Both statement types help reduce overhead for transactions with backend data.

- A prepared statement is a precompiled SQL statement that is stored in a PreparedStatement object. Application Server uses this object to run the SQL statement multiple times, as required by your application run time, with values that are determined by the run time.
- A callable statement is an SQL statement that contains a call to a stored procedure, which is a series of precompiled statements that perform a task and return a result. The statement is stored in the CallableStatement object. Application Server uses this object to run a stored procedure multiple times, as required by your application run time, with values that are determined by the run time.

In general, the more statements your application has, the larger the cache should be. **Be aware**, however, that specifying a larger statement cache size than needed wastes application memory and *does not* improve performance.

Determine the value for your cache size by adding the number of uniquely prepared statements and callable statements (as determined by the SQL string, concurrency, and the scroll type) for each application that uses this data source on a particular server. This value is the maximum number of possible statements that can be cached on a given connection over the life of the server. Additional information about this setting...

Default: For most databases the default is 10. Zero means there is no cache statement.

## Connection pool tuning

To view the administrative console page where you configure the following properties, click **Resources > JDBC Providers > JDBC\_provider > Data sources > data\_source > Connection pool settings**.

### Maximum connections

Specifies the maximum number of physical connections that can be created in this pool. These are the physical connections to the backend datastore.

When this number is reached, no new physical connections are created; requestors must wait until a physical connection that is currently in use is returned to the pool.

For optimal performance, set the value for the connection pool lower than the value for the Web container threadpool size. Lower settings, such as 10 to 30 connections, might perform better than higher settings, such as 100. Additional information about this and related settings...

Default: 10

#### Minimum connections

Specifies the minimum number of physical connections to maintain. Until this number is exceeded, the pool maintenance thread does not discard physical connections.

If you set this property for a higher number of connections than your application ultimately uses at run time, you do not waste application resources. WebSphere Application Server does not create additional connections to achieve your minimum setting. Of course, if your application requires more connections than the value you set for this property, application performance diminishes as connection requests wait for fulfillment. Additional information about this and related settings...

Default: 1

## Messaging

### Setting tuning properties of a messaging engine

Use this task to set the tuning properties for a messaging engine.

You can set the following property to improve the performance of a messaging engine:

| Name          | Value                                                                                                                                                                                                                                                                        |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib.trm.retry | The messaging engine to messaging engine connection retry interval in seconds. The retry interval is the time delay left between attempts to contact neighboring messaging engines with which communications contact should exist. The default retry interval is 30 seconds. |

To set the tuning properties for a messaging engine, use the administrative console to complete the following steps:

1. In the navigation pane, click **Service integration** → **Buses** [Content Pane] → **bus\_name** → [Additional Properties] **Messaging engines** → **engine\_name** → [Additional Properties] **Custom properties**.
2. Type the name of the property that you want to set.
3. Type the value that you want to set for that property.
4. Click **OK**.
5. Save your changes to the master configuration.
6. Restart the messaging engine for the changes to take effect.



## Tuning and problem solving for messaging engine data stores

Use this task to obtain an overview of improving the performance of messaging engine data stores and understanding problems that can occur with a data store.

For more information about tuning and problem solving for messaging engine data stores, see the following topics:

- “Tuning the JDBC data source of a messaging engine” on page 70
- “Controlling the memory buffers used by a messaging engine” on page 70
- Sharing connections to benefit from one-phase commit optimization
- Diagnosing problems with data store exclusive access locks
- “Diagnosing problems with your data store configuration” on page 72
- “Avoiding failover problems when you use DB2 v8.2 with HADR as your data store” on page 73

### Tuning the JDBC data source of a messaging engine:

Use this task to tune the JDBC data source of a messaging engine for messaging performance on a service integration bus.

Consider whether you need to configure the connection pool for the JDBC data source to achieve your requirements for messaging performance.

The messaging engine uses the connection pool to obtain its connections to the database. With a heavy workload, a messaging engine might require a large number of concurrent connections to avoid delays waiting for connections to become available in the pool. For example, a very heavily loaded messaging engine might need 50 or more connections. Perform the following steps to configure the connection pool to meet your performance requirements:

1. Ensure that the configuration of your relational database management system (RDBMS) permits the number of connections that you require. Refer to the documentation for your RDBMS for more information.
2. Use the WebSphere administrative console to set the connection pool parameters. Navigate to the **General properties** for your data source and click **Additional properties** → **Connection pool properties**
  - a. Set the **Maximum connections** to the number of connections you require, for example, at least 50. The default number of connections is 10.

**Tip:** If your messaging engine times out when requesting a database connection, check the error log. If the error log contains error message CWSIS1522E, increase the number of connections and ensure that the configuration of your RDBMS permits that number of connections.

- b. Set the **Purge policy** to *EntirePool*. This policy enables the connection pool to release all connections when the messaging engine stops.

**Important:** You must set this value if the messaging engine can failover in a cluster.

### Controlling the memory buffers used by a messaging engine:

Use this task to control the sizes of the memory buffers used by a messaging engine.

Every messaging engine manages two memory buffers that contain messages and message-related data. You can set the following properties to improve the interaction of a messaging engine with its data store.

| Name                                   | Value                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib.msgstore.discardableDataBufferSize | <p>The size in bytes of the data buffer used by the messaging engine to contain data for which the quality of service attribute is best effort nonpersistent. The messaging engine holds this data entirely within this memory buffer and never writes this data to the data store. When the messaging engine adds data to this buffer, for example when the messaging engine receives a best effort nonpersistent message from a client, the messaging engine might discard data already in the buffer to make space. This behavior enables the messaging engine to discard best effort nonpersistent messages.</p> <p>The discardable data buffer contains all data for which the quality of service attribute is best effort nonpersistent. That data comprises data both that is involved in active transactions, and any other best effort nonpersistent that the messaging engine has neither discarded nor consumed. The messaging engine can discard only data that is not involved in active transactions.</p> <p><b>Tip:</b> If the messaging engine attempts to add data to the discardable data buffer when insufficient space remains after discarding all the data that is not involved in active transaction, the messaging engine throws a <code>com.ibm.ws.sib.msgstore.OutOfCacheSpace</code> exception. Client applications can catch this exception, wrapped inside API-specific exceptions such as <code>javax.jms.JMSEException</code>.</p> <p>The <code>sib.msgstore.discardableDataBufferSize</code> property of the messaging engine controls the size of the discardable data buffer. You specify the value of this property in bytes. The default value is 320000, which is approximately 320 kilobytes.</p> |

| Name                              | Value                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib.msgstore.cachedDataBufferSize | <p>The size in bytes of the data buffer used by the messaging engine to contain data for which the quality of service attribute is <i>better than</i> best effort nonpersistent and which is held in the data store. The purpose of the cached data buffer is to optimize the performance of the messaging engine by caching in memory the data that the messaging engine might otherwise need to read from the data store. As it writes data to the data store and reads from the data store, the messaging engine attempts to add that data to the cached data buffer. The messaging engine might discard data already in the buffer to make space.</p> <p>The <code>sib.msgstore.cachedDataBufferSize</code> property of the messaging engine controls the size of the cached data buffer. You specify the value of this property in bytes. The default value is 320000, which is approximately 320 kilobytes.</p> |
| sib.msgstore.transactionSendLimit | <p>The maximum number of operations that the messaging engine includes in each transaction. For example, each JMS send or receive is an operation that counts towards the transaction send limit. The default value is 100.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

**Attention:** The messaging engine uses approximate calculations to manage the data it holds in the memory buffers. Neither of the `DataBufferSize` properties gives an accurate indication of the amount of memory that the messaging engine consumes in the JVM heap. The messaging engine can consume considerably more heap storage than the `DataBufferSize` properties indicate.

To set the properties of a messaging engine to improve its interaction with its data store, use the administrative console to complete the following steps:

1. In the navigation pane, click **Service integration** → **Buses** [Content Pane] → **bus\_name** → [Additional Properties] **Messaging engines** → **engine\_name** → [Additional Properties] **Custom properties**.
2. Type the name of the property that you want to set.
3. Type the value that you want to set for that property.
4. Click **OK**.
5. Save your changes to the master configuration.

**Remember:** When you change any of these properties, the new values do not take effect until you restart the messaging engine.

#### Diagnosing problems with your data store configuration:

Use this task to diagnose problems that are caused by your data store configuration.

The following problems depend on the database that you use with your data store configuration and the level of that database:

- Examine this section if your messaging engine uses an Oracle 9i database for its data store and your messaging engine fails to start. If the messaging engine fails with the following message, where XXXXXXXX is the schema for the table, ensure that your Oracle installation is at 9.2.0.4, or higher:  
CWSIS1530E: The data type, 1,111, was found instead of the expected type, 2,004, for column, LONG\_DATA, in table, XXXXXXXX.SIB000.
- Examine this section if your messaging engine uses a Sybase database for its data store. When you create your Sybase server:
  - Ensure that you create the database server with a page size of at least 4k.
  - Ensure that you set the **lock scheme** property on your server to the value *datarows*. This avoids the possibility of a deadlock on the data store tables.
- Examine this section if your messaging engine uses an Informix database for its data store and the messaging engine is unable to access its data store. When you configure your messaging engine to use an Informix database, ensure that you specify the schema name in lower case. For a full description of the configuring procedure, refer to Configuring a messaging engine to use a data source.

### **Avoiding failover problems when you use DB2 v8.2 with HADR as your data store:**

Use this task to avoid problems that can occur when a messaging engine that is configured to use DB2 v8.2 with HADR for its data store terminates if the DB2 database fails over.

If you configure a messaging engine to use DB2 v8.2 with HADR for its data store and a DB2 failover occurs, the application server on which the messaging engine is running might stop. The cause of this problem is that the messaging engine does not treat the DB2 failover as a transient communications error.

When you configure a messaging engine to use DB2 v8.2 with HADR for its data store ensure that the messaging engine can restart automatically following an application server failure. Choose the option that is appropriate for your configuration:

- If you are running a single server, WebSphere Application Server provides no failover support. Consider other high availability provisions.
- If you are running the Network Deployment version without clustering, the default configuration for the node agent ensures automatic restart.
- If you are running the Network Deployment version with clustering, peer recovery restarts the messaging engine. Ensure that you have configured the high availability policy to enable peer recovery.

### **Setting tuning properties for a mediation**

Use this task to tune a mediation for performance using the administrative console.

Before you begin this task, you should review the guidance on when it is appropriate to tune a mediation for performance in the topic [Guidance for tuning mediations for performance](#).

You can set the following tuning properties to improve the performance of a mediation:

| Name                       | Value                                                                                                                                                                                                                                                                                                       |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib:SkipCloneBeforeMediate | Whether you want to omit the cloning of messages before they are processed by the mediation. Either true or false.                                                                                                                                                                                          |
| sib:SkipWellFormedCheck    | Whether you want to omit the well formed check that is performed on messages after they have been processed by the mediation. Either true or false.<br><b>Note:</b> This property is overridden for messages that have the delivery option assured persistent, and a well formed check is always performed. |

You can set the following tuning property to improve the performance of a mediation:

| Name                    | Value                                                                                                                                                                                                                                                                                                       |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sib:SkipWellFormedCheck | Whether you want to omit the well formed check that is performed on messages after they have been processed by the mediation. Either true or false.<br><b>Note:</b> This property is overridden for messages that have the delivery option assured persistent, and a well formed check is always performed. |

To set, or unset, one or more tuning properties for a mediation, use the administrative console to complete the following steps:

1. Display the mediation context information:
  - a. In the navigation pane, click **Service integration** → **Buses**
  - b. In the content pane, click the name of the service integration bus.
  - c. Under **Additional Properties**, click **Mediations**.
  - d. In the content pane, select the name of the mediation for which you want to configure tuning information.
  - e. Under **Additional Properties**, click **Context information**.
2. In the content pane, click **New**.
3. Type the name of the property in the **Name** field.
4. Select the type `Boolean` in the list box.
5. Type **true** in the **Context Value** field to set the property, or type **false** to unset the property.
6. Click **OK**.
7. Save your changes to the master configuration.

### Enabling CMP entity beans and messaging engine data stores to share database connections

Use this task to enable container-managed persistence (CMP) entity beans to share the database connections used by the data store of a messaging engine. This has been estimated as a potential performance improvement of 15% for overall message throughput, but can only be used for entity beans connected to the application server that contains the messaging engine.

To enable CMP entity beans to share the database connections used by the data store of a messaging engine, complete the following steps:

1. Configure the data store to use a data source that is not XA-capable. For more information about configuring a data store, see *Configuring a JDBC data source*.

2. Select the Share data source with CMP option.

This option is provided on the JMS connection factory used to connect to the service integration bus that hosts the bus destination that is used to store and process messages for the CMP bean.

This option is provided on the JMS connection factory or JMS activation specification used to connect to the service integration bus that hosts the bus destination that is used to store and process messages for the CMP bean.

For example, to select the option on a unified JMS connection factory, complete the following steps:

- a. Display the default messaging provider. In the navigation pane, expand **Resources** → **JMS Providers** → **Default Messaging Provider**.
- b. Optional: Change the **Scope** check box to set the level at which the connection factory is to be visible, according to your needs.
- c. In the content pane, under Additional Properties, click **JMS connection factories**
- d. Optional: To create a new unified JMS connection factory, click **New**.

Specify the following properties for the connection factory:

**Name** Type the name by which the connection factory is known for administrative purposes.

**JNDI name**

Type the JNDI name that is used to bind the connection factory into the name space.

**Bus name**

Type the name of the service integration bus that the connection factory is to create connections to. This service integration bus hosts the destinations that the JMS queues and topics are assigned to.

- e. Optional: To change the properties of an existing connection factory, click one of the connection factories displayed. This displays the properties for the connection factory in the content pane.
- f. Select the check box for the Share data source with CMP field
- g. Click **OK**.
- h. Save your changes to the master configuration.

The JMS connection factory can only be used to connect to a “local” messaging engine that is in the application server on which the CMP beans are deployed.

3. Deploy the CMP beans onto the application server that contains the messaging engine, and specify the same data source as used by the messaging engine. You can use the administrative consoles to complete the following steps:

- a. Optional: To determine the data source used by the messaging engine, click **Servers** → **Application servers** → **server\_name** → **Messaging engines** → **engine\_name** → **Data store** The **Data source name** field displays the name of the data source; by default:

`jdbc/com.ibm.ws.sib/engine_name`

- b. Click **Applications** → **Install New Application**

- c. On the first Preparing for application install page, specify the full path name of the source application file (.ear file otherwise known as an EAR file), then click **Next**
- d. On the second Preparing for application install page, complete the following steps:
  - 1) Select the check box for the Generate Default Bindings property. Data source bindings (for EJB 1.1 JAR files) are generated based on the JNDI name, data source user name password options. This results in default data source settings for each EJB JAR file. No bean-level data source bindings are generated.
  - 2) Under Connection Factory Bindings, click the check box for the **Default connection factory bindings:** property, then type the JNDI name for the data source and optionally select a **Resource authorization** value.
  - 3) Click **Next**
4. If your application uses EJB modules that contain Container Managed Persistence (CMP) beans that are based on the EJB 1.x specification, for Step: Provide default datasource mapping for modules containing 1.x entity beans, specify a JNDI name for the default data source for the EJB modules. The default data source for the EJB modules is optional if data sources are specified for individual CMP beans.
5. If your application has CMP beans that are based on the EJB 1.x specification, for Step: Map datasources for all 1.x CMP, specify a JNDI name for data sources to be used for each of the 1.x CMP beans. The data source attribute is optional for individual CMP beans if a default data source is specified for the EJB module that contains CMP beans. If neither a default data source for the EJB module nor a data source for individual CMP beans are specified, then a validation error displays after you click Finish (step 13) and the installation is cancelled.
6. Complete other panels as needed.
7. On the Summary panel, verify the cell, node, and server onto which the application modules will install:
  - a. Beside Cell/Node/Server, click **Click here**.
  - b. Verify the settings on the Map modules to servers page displayed. Ensure that the application server that is specified contains the messaging engine and its data store.
  - c. Specify the Web servers as targets that will serve as routers for requests to this application. This information is used to generate the plug-in configuration file (plugin-cfg.xml) for each Web server.
  - d. Click **Finish**.

For more information about installing applications, see Installing application files with the console.

## Tuning service integration messaging

Use this task to set tuning properties that control the performance of message-driven beans and other messaging applications deployed to use service integration technologies.

To optimize the performance of messaging with service integration technologies, such as message-driven beans that use the default messaging provider, you can use the following parameters set through the WebSphere administrative console or command line interfaces.

On z/OS, the performance of messaging applications is affected by the number of servants, which can vary dynamically, and the distribution of work between servants. For more information about configuring and managing the number of servants and the distribution of work between servants, see *Tuning the application serving environment*.

- Providing sufficient threads to maximize the throughput of messages.

There is a trade-off between providing sufficient threads to maximize the throughput of messages and configuring excessive threads, which can lead to CPU starvation of the threads in the application server. If you notice that the throughput for express nonpersistent, reliable nonpersistent, or reliable persistent messaging has fallen as a result of increasing the size of the default thread pool, then you should decrease the size of the thread pool and reassess the message throughput.

1. To be able to view the available message count for a queue, you need to enable runtime `AvailableMessageCount` statistics for the queue. If you restart administrative server, you need to enable `AvailableMessageCount` statistics again because such runtime settings are not preserved when the server is restarted.

To enable `AvailableMessageCount` statistics using the administrative console, complete the following steps:

- a. In the navigation pane, click **Monitoring and Tuning** → **Performance Monitoring Infrastructure (PMI)**
  - b. In the content pane, click *server\_name*
  - c. Click the Runtime tab.
  - d. In the Currently monitored statistic set, click **Custom**
  - e. On the Custom monitoring level panel, click **SIB Service** → **SIB Messaging Engines** → *messageEngine\_name* → **Destinations** → **Queues** → *queue\_name*
  - f. Select the `AvailableMessageCount` option.
  - g. Click the **Enable** button at the top of the panel.
2. To view the available message count, you can use the administrative console to complete the following steps:

- a. In the navigation pane, click **Monitoring and Tuning** → **Performance Viewer** → **Current activity**
- b. In the content pane, click *server\_name*
- c. Click **Performance Modules** → **SIB Service** → **SIB Messaging Engines** → *messageEngine\_name* → **Destinations** → **Queues** → *queue\_name*
- d. Click the **View Module(s)** button at the top of the Resource Selection panel, located on the left side. This displays the `AvailableMessageCount` data in the Data Monitoring panel, located on the right side.

You can use the Data Monitoring panel to manage the collection of monitoring data; for example, you can use the buttons to start or stop logging, or to change the data displayed as either a table or graph.

3. To view or change the number of threads in the default thread pool for an application server, you can use the administrative console to complete the following steps:
  - a. In the navigation pane, click **Servers** → **Application servers**
  - b. In the content pane, click *server\_name*
  - c. Under Additional properties, click **Thread Pools** → **Default**. By default the Minimum size value is set to 5 and the Maximum size value is set to 20. The best performance is obtained by setting the Maximum size value to



the expected maximum concurrency for all message-driven beans. For high throughput using a single message-driven bean, 41 was found to be the optimal Maximum size value.

- d. To change the Maximum size value, type the new value in the Maximum size field then click **OK**. Finally, save your changes to the master configuration.
- Reducing the number of OutOfCacheSpace errors in the SystemOut.log file. OutOfCacheSpace errors in the SystemOut.log file indicate that the discardable data buffer used by the messaging engine is overflowing. For best effort nonpersistent messages, the messaging engine starts to discard messages when this buffer is full. You can increase the size of this data buffer to allow more best effort nonpersistent data to be handled before the messaging engine begins to discard the messages.

For more information about tuning the size of the discardable data buffer, set by the `sib.msgstore.discardableDataBufferSize` property of a messaging engine, see “Controlling the memory buffers used by a messaging engine” on page 70.

- Tuning the messaging engine data store (not required for Cloudscape). To improve the performance of messaging throughput of a messaging engine data store, you can tune the JDBC connection pool and statement cache size. In tests of high throughput MDB workloads, the following changes provided a 10% gain in throughput.

1. The messaging engine uses a connection pool for managing the JDBC connections to its data store. Tuning the size of the pool can improve the messaging throughput.

To view or change the size of the connection pool, you can use the administrative console to complete the following steps:

- a. In the navigation pane, click **Resources** → **JDBC Providers**
  - b. In the content pane, click *jdbc\_provider\_name*
  - c. Under Additional properties, click **Data sources** → *data\_source\_name*
  - d. Under Additional properties, click **Connection pool properties**
  - e. View the Maximum connections property and the Minimum connections property. By default, these properties are set to Maximum connections=10 and Minimum connections=1. Setting the value of both these properties to 50 is recommended. For especially high throughput workloads, setting the value of both these properties up to 100 can be beneficial. You may need to configure the underlying database to accept this many concurrent connections.
  - f. To change the value of a property, type a new value in the property field then click **OK**. Finally, save your changes to the master configuration.
2. The statement cache contains recently used prepared statements to remove the costs associated with repeated preparation of statements. Tuning the size of the cache helps prevent useful entries from being discarded to make room for new entries.

To view or change the size of the statement cache, you can use the administrative console to complete the following steps:

- a. In the navigation pane, click **Resources** → **JDBC Providers**
- b. In the content pane, click *jdbc\_provider\_name*
- c. Under Additional properties, click **Data sources** → *data\_source\_name*
- d. Under Additional properties, click **WebSphere Application Server data source properties**

- e. View the Statement cache size property. By default, the value of this property is set to 10. For high throughput JMS messaging, a value of 40 is recommended.
  - f. To change the value of the property, type a new value in the property field then click **OK**. Finally, save your changes to the master configuration.
- Using a remote DB2 Universal Database.  
By default, messaging engines are configured to use Cloudscape as the underlying data store database. Changing to a DB2 database, especially a remote database, significantly improves performance. This is of particular benefit on systems with more than two CPUs. A remote DB2 database provides a 50% higher throughput than a local DB2 database, for the same total (application server plus database) CPU utilization.
  - Tuning reliability levels for messages.  
The reliability level chosen for the messages has a significant impact on performance. In order of decreasing performance (fastest first), the reliability levels are: Best-Effort Nonpersistent, Express Nonpersistent, Reliable Nonpersistent, Reliable Persistent, and Assured Persistent. For MDB point-to-point messaging, best-effort nonpersistent throughput is more than 6 times greater than assured persistent.  
For more information about reliability levels, see [Message reliability levels](#). For the relative performance of the different reliability levels, see the [WebSphere Application Server performance report](#).
  - Tuning MDB performance with the default messaging provider.  
For information about tuning the throttling of message-driven beans, including controlling the maximum number of instances of each message-driven bean and the message batch size for serial delivery, see [Configuring MDB throttling on the default messaging provider](#).

## Tuning the SIBWS

You can use the administrative console or a Jacl script to tune performance settings for the service integration bus Web services enablement (SIBWS).

To improve the performance of the SIBWS, you can tune the following parameters:

- The Java virtual machine heap size (to help ensure there is enough memory available to process large messages, or messages with large attachments).
- The maximum permitted number of concurrent clients. (That is, the maximum number of instances of a message-driven bean that are permitted by the activation specification for the service integration technologies resource adapter).
- The maximum batch size for batches of messages to be delivered to a client. (By default, only a single message is delivered to a message-driven bean instance at one time. You can improve performance by allowing messages to be sent in batches to a message-driven bean).
- The number of threads available to service requests for each client. (That is, the number of threads available in the default thread pool and the Web container thread pool for a given application server).

To tune the SIBWS, complete one of the following two main steps:

- **Optional:** To use the administrative console to tune the SIBWS, complete the following steps:
  1. Use the topic [Tuning Java virtual machines](#) to set the JVM heap size to a larger value than the default value (256 megabytes). The value should generally be as large as possible without incurring paging.

2. Use the topic Tuning service integration messaging to tune the maximum number of instances of a message-driven bean, the maximum batch size for batches of messages for a bean, and the number of threads available to service requests for a bean:

- In the administrative console navigation pane, click **Resources** → **Resource adapters** → **SIB JMS Resource Adapter** → **J2C activation specification** → **SIBWS\_OUTBOUND\_MDB**, then set `maxConcurrency` to  $n$  and `maxBatchSize` to 5 (where  $n$  is the maximum permitted number of instances of a message-driven bean).
- In the administrative console navigation pane, click **Servers** → **Application servers** → `server_name` → **Thread Pools**, then set the number of threads in the default and WebContainer thread pools to a maximum of  $n-1$  threads and a minimum of some figure approaching the maximum (to reduce lock contention).

- To use a Jacl script to tune the SIBWS, create and run a script based on the following example script:

```
puts "****Configuring server thread pools ****"

if {[length $argv] !=1} {
 puts "ERROR: Incorrect usage."
 puts ""
 puts {Usage: wsadmin -f configure-soapservice1.jacl serverName}
 puts ""
 puts "Where:"
 puts ""
 puts " serverName = name of server"
 exit
}

#-----
#--- Setup the arguments
#-----

#--- Get args
set serverName [lindex $argv 0]

#-----
#--- Set the thread pool sizes and the JVM heap size
#-----
set serverId [$AdminConfig getid /Server:$serverName/]

#--- First the threadPools...
puts "Modifying Server ThreadPool Sizes for $serverName"

set inactivityTimeout [list inactivityTimeout 120000]
set isGrowable [list isGrowable false]
set maximumSize [list maximumSize 150]
set minimumSize [list minimumSize 145]

set threadPoolList [list $inactivityTimeout \
 $isGrowable \
 $maximumSize \
 $minimumSize \
]
set threadpools [$AdminConfig list ThreadPool $serverId]

foreach pool $threadpools {
 set modifyToggle 0
 #puts [$AdminConfig showAttribute $pool name]
 if {[string compare [$AdminConfig showAttribute $pool name] WebContainer] == 0} {
 set modifyToggle 1
 #puts "WebContainer ThreadPool found for server $serverName"
 }
}
```

```

 if {[string compare [$AdminConfig showAttribute $pool name] Default] == 0} {
set modifyToggle 1
#puts "Default ThreadPool found for server $serverName"
 }
 if {$modifyToggle == 1} {
puts "modifying threadpool [$AdminConfig showAttribute $pool name]"
$AdminConfig modify $pool $threadPoolList
 }
}
puts "ThreadPool modifications complete"

#--- Now the JVM heap size
set heapSize 1536
puts "Modifying JVM Heap size to $heapSize for server $serverName"

set initialHeapSize [list initialHeapSize $heapSize]
set maximumHeapSize [list maximumHeapSize $heapSize]

set jvmModifications [list $initialHeapSize $maximumHeapSize]

set processDef [$AdminConfig list ProcessDef $serverId]
set jvm [$AdminConfig list JavaVirtualMachine $processDef]

$AdminConfig modify $jvm $jvmModifications
puts "Heap size modification complete"

#-----
#--- Modify the J2CActivationSpec for the service integration technologies Resource adapter
#-----
puts "Modifying the J2CActivationSpec"

set actSpec [$AdminConfig getid /J2CActivationSpec:SIBWS_OUTBOUND_MDB/]
set propSet [$AdminConfig showAttribute $actSpec resourceProperties]

set propSet [lindex $propSet 0]

set maxConcurrency [list value 200]
set maxConcurrency [list $maxConcurrency]

set maxBatchSize [list value 5]
set maxBatchSize [list $maxBatchSize]

foreach propId $propSet {
 if { [string compare [$AdminConfig showAttribute $propId name] maxConcurrency] == 0} {
$AdminConfig modify $propId $maxConcurrency
puts "Custom property changed : [$AdminConfig showall $propId] "
 }
 if { [string compare [$AdminConfig showAttribute $propId name] maxBatchSize] == 0} {
$AdminConfig modify $propId $maxBatchSize
puts "Custom property changed : [$AdminConfig showall $propId] "
 }
}

puts "J2CActivationSpec modifications complete"

puts "Saving actions..."
$AdminConfig save

```

## Security

### Tuning security configurations

Performance issues typically involve trade-offs between function and speed. Usually, the more function and the more processing involved, the slower the

performance. Consider what type of security is necessary and what you can disable in your environment. For example, if your application servers are running in a Virtual Private Network (VPN), consider whether you can disable Secure Sockets Layer (SSL). If you have a lot of users, can they be mapped to groups and then associated to your J2EE roles? These questions are things to consider when designing your security infrastructure.

Complete the following steps for general security tuning:

1. Consider disabling Java 2 Security Manager if you know exactly what code is put onto your server and you do not need to protect process resources. Remember that in doing so, you put your local resources at some risk.
2. Consider disabling security (realizing that this might cause a security exposure) for the specific application server that does not require resource protection because some application servers do not have protected resources. If the application server needs to go downstream with credentials, however, this action might not be feasible. There are two ways to disable security. Use either of the following two procedures and then restart the application server:

#### Using the administrative console

- a. Click **Servers > Application servers > *server\_name***.
- b. Under Security, click **Server security**. Under Additional properties, click **Server-level security**.
- c. Disable the **Enable global security** option.

#### Using the command line

- a. Type `wsadmin.sh -conntype NONE`.
  - b. When the system command prompt redisplay, type `securityoff`.
3. Consider propagating new security settings to all nodes before restarting the deployment manager and node agents to change the new security policy.  
If your security configurations are not consistent across all servers, you get access denied errors. Therefore, you must propagate new security settings when enabling or disabling global security in a Network Deployment environment.  
Configuration changes are generally propagated using configuration synchronization. If auto-synchronization is enabled, you can wait for the automatic synchronization interval to pass, or you can force synchronization before the synchronization interval expires. If you are using manual synchronization, you must synchronize all nodes.  
If the cell is in a configuration state (the security policy is mixed with nodes that have security enabled and disabled) you can use the `syncNode` utility to synchronize the nodes where the new settings are not propagated.  
Refer to the article, *Enabling global security in the WebSphere Application Server Network Deployment package* for more detailed information about enabling security in a distributed environment.
  4. Consider increasing the cache and token time-out if you feel your environment is secure enough. By doing so, you have to re-authenticate less often. This action supports subsequent requests to reuse the credentials that already are created. The downside of increasing the token time-out is the exposure of having a token hacked and providing the hacker more time to hack into the system before the token expires. You can use security cache properties to determine the initial size of the primary and secondary hashtable caches, which affect the frequency of rehashing and the distribution of the hash algorithms. See the article “Security cache properties” on page 106 for a list of these properties.

5. Consider changing your administrative connector from Simple Object Access Protocol (SOAP) to Remote Method Invocation (RMI) because RMI uses stateful connections while SOAP is completely stateless. Run a benchmark to determine if the performance is improved in your environment.
6. Use the `wsadmin` script to complete the access IDs for all the users and groups to speed up the application startup. Complete this action if applications contain many users, or groups, or if applications are stopped and started frequently.
7. Consider tuning the Object Request Broker (ORB) because it is a factor in enterprise bean performance with or without security enabled. Refer to ORB tuning guidelines.
8. If using SSL, then enable the SSL session tracking mechanism option as described in the article, Session management settings.
9. In some cases, using the unrestricted Java Cryptography Extension (JCE) policy file can improve performance. Refer to the article, Tuning Web services security.

#### Tuning CSiv2:

1. Consider using Secure Sockets Layer (SSL) client certificates instead of a user ID and password to authenticate Java clients. Because you are already making the SSL connection, using mutual authentication adds little overhead while removing the service context containing the user ID and password completely.
2. If you send a large amount of data that is not very security sensitive, reduce the strength of your ciphers. The more data you have to bulk encrypt and the stronger the cipher, the longer this action takes. If the data is not sensitive, do not waste your processing with 128-bit ciphers.
3. Consider putting just an asterisk (\*) in the trusted server ID list (meaning trust all servers) when you use Identity Assertion for downstream delegation. Use SSL mutual authentication between servers to provide this trust. Adding this extra step in the SSL handshake performs better than having to fully authenticate the upstream server and check the trusted list. When an asterisk is used, the identity token is trusted. The SSL connection trusts the server by way of client certificate authentication.
4. Ensure that stateful sessions are enabled for Common Secure Interoperability Version 2 (CSiv2). This is the default, but only requires authentication on the first request and any subsequent token expirations.
5. If you are only communicating with WebSphere Application Server Version 5 or higher servers, make the Active Authentication Protocol CSI, instead of CSI and SAS. This action removes an interceptor invocation for every request on both the client and server sides.
6. For a pure Java client, you can disable the creation of server sockets used for Object Request Broker (ORB) callbacks. Do this step only if you are communicating with servers running WebSphere Application Server, Version 5 or later.
  - a. In the `sas.client.props` file, add  
`com.ibm.CSI.claimTransportAssocSSLTLSRequired=false` and  
`com.ibm.CSI.claimTransportAssocSSLTLSSupported=false`.
  - b. Set the active protocol to `csiv2` instead of both in the `sas.client.props` file. The protocol property changes to `com.ibm.CSI.protocol=csiv2`.

#### Tuning LDAP authentication:

1. Select the **Ignore case for authorization** option in the WebSphere Application Server Lightweight Directory Access Protocol (LDAP) User Registry configuration, when case-sensitivity is not important.

2. Select the **Reuse connections** option in the WebSphere Application Server LDAP User Registry configuration.
3. Check to see which caches your LDAP server has and take advantage of them. This action is best with LDAP servers that do not change frequently.
4. Choose the directory type of either IBM Tivoli Directory Server or SecureWay, if you are using an IBM Tivoli Directory Server. The IBM Tivoli Directory Server yields improved performance because it is programmed to use the new group membership attributes to improve group membership searches. However, it is required that authorization is case insensitive to use IBM Tivoli Directory Server.
5. Choose either iPlanet Directory Server (also known as Sun ONE) or Netscape as the directory if you are an iPlanet Directory user. Using the iPlanet Directory Server directory increases performance in group membership lookup. However, only use **Role** for group mechanisms.

#### Tuning Web authentication:

1. Consider increasing the cache and token time-out if you feel your environment is secure enough. The Web authentication information is stored in these caches and as long as the authentication information is in the cache, the login module is not invoked to authenticate the user. This supports subsequent requests to reuse the credentials already created. The downside of increasing the token time-out is the exposure of having a token stolen and providing the thief more time to hack into the system before the token expires.  
See the article "Security cache properties" on page 106 for a list of these properties.
2. Consider enabling single signon (SSO). SSO is only available when you select **LTPA** as the authentication mechanism in the Global security panel. When you select SSO, a single authentication to one application server is enough to make requests to multiple application servers in the same SSO domain. There are some situations where SSO is not desirable and should not be used in those situations.
3. Consider disabling or enabling the **Web Inbound Security Attribute Propagation** option on the SSO panel if the function is not required. In some cases, having the function enabled improves performance. This improvement is most likely for higher volume cases where a considerable number of user registry calls reduces performance. In other cases, having the feature disabled improves performance. This improvement is most likely when the user registry calls do not take considerable resources.

#### Tuning authorization:

1. Consider mapping your users to groups in the user registry. Then associate the groups with your Java 2 Platform, Enterprise Edition (J2EE) roles. This association greatly improves performance as the number of users increases.
2. Judiciously assign method-permissions for enterprise beans. For example, you can use an asterisk (\*) to indicate all methods in the method-name element. When all the methods in enterprise beans require the same permission, use an asterisk (\*) for the method-name to indicate all methods. This indication reduces the size of deployment descriptors and reduces the memory required to load the deployment descriptor. It also reduces the search time during method-permission match for the enterprise beans method.
3. Judiciously assign security-constraints for servlets. For example, you can use the URL pattern \*.jsp to apply the same authentication data constraints to indicate all JSP files. For a given URL, the exact match in the deployment descriptor takes precedence over the longest path match. Use the extension

match (\*.jsp, \*.do, \*.html) if there is no exact match and longest path match for a given URL in the security constraints.

There is always a trade off between performance, feature and security. Security typically adds more processing time to your requests, but for a good reason. Not all security features are required in your environment. When you decide to tune security, you should create a benchmark before making any change to ensure the change is improving performance.

In a large scale deployment, performance is very important. Running benchmark measurements with different combinations of features can help you to determine the best performance versus the benefit configuration for your environment. Continue to run benchmarks if anything changes in your environment, to help determine the impact of these changes.

**Security cache properties:** The following security cache custom properties determine whether the authentication cache is enabled or disabled. If the authentication cache is enabled, as recommended, these custom properties specify the initial size of the primary and secondary hash table caches, which affect the frequency of rehashing and the distribution of the hash algorithms. To specify these security cache properties, complete the following steps:

1. Click **Servers > Application Servers > *server\_name*** .
2. Under Server infrastructure, click **Java and Process Management > Process definition > Control or Servant**.
3. Under Additional properties, click **Java Virtual Machine > Custom Properties**
4. Click **New** to specify a new custom property.

**Important:** The `com.ibm.websphere.security.util.tokenCacheSize` and `com.ibm.websphere.security.util.LTPAValidationCacheSize` properties were replaced with the `com.ibm.websphere.security.util.authCacheSize` property.

WebSphere Application Server includes the following security cache custom properties:

**`com.ibm.websphere.security.util.authCacheEnabled`**

Specifies whether to disable the authentication cache. It is recommended that you leave the authentication cache enabled for performance reasons. However, you might disable the authentication cache for debug or measurement purposes.

|          |      |
|----------|------|
| Default: | True |
|----------|------|

**`com.ibm.websphere.security.util.authCacheSize`**

Specifies the initial size of the primary and secondary hash table caches. A higher number of available hash values might decrease the occurrence of hash collisions. A hash collision results in a linear search for the hash bucket, which might decrease the retrieval time. If several entries compose a hash table cache, you can might consider creating a table with a larger capacity that supports more efficient hash entries instead of allowing automatic rehashing determine the growth of the table. Rehashing causes every entry to move each time.

|          |         |
|----------|---------|
| Default: | 200     |
| Type:    | Integer |



### Related reference

Example: User revocation from a cache

Chapter 5, "Tuning the application serving environment," on page 25

This topic describes the benefits of tuning for optimal performance, highlights the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

**Secure Sockets Layer performance tips:** The following are two types of Secure Sockets Layer (SSL) performance:

- Handshake
- Bulk encryption and decryption

When an SSL connection is established, an SSL handshake occurs. After a connection is made, SSL performs bulk encryption and decryption for each read-write. The performance cost of an SSL handshake is much larger than that of bulk encryption and decryption.

To enhance SSL performance, decrease the number of individual SSL connections and handshakes.

Decreasing the number of connections increases performance for secure communication through SSL connections, as well as non-secure communication through simple TCP/IP connections. One way to decrease individual SSL connections is to use a browser that supports HTTP 1.1. Decreasing individual SSL connections can be impossible if you cannot upgrade to HTTP 1.1.

Another common approach is to decrease the number of connections (both TCP/IP and SSL) between two WebSphere Application Server components. The following guidelines help to verify the HTTP transport of the application server is configured so that the Web server plug-in does not repeatedly reopen new connections to the application server:

- Verify that the maximum number of keep alives are, at minimum, as large as the maximum number of requests per thread of the Web server (or maximum number of processes for IBM HTTP Server on UNIX). Make sure that the Web server plug-in is capable of obtaining a keep alive connection for every possible concurrent connection to the application server. Otherwise, the application server closes the connection after a single request is processed. Also, the maximum number of threads in the Web container thread pool should be larger than the maximum number of keep alives, to prevent the keep alive connections from consuming the Web container threads.

**Note:** HTTP Transports have been deprecated. For instructions on how to set a maximum keep alive value for channel based configurations, see HTTP transport channel settings.

- Increase the maximum number of requests per keep alive connection. The default value is 100, which means the application server closes the connection from the plug-in after 100 requests. The plug-in then has to open a new connection. The purpose of this parameter is to prevent denial of service attacks when connecting to the application server and preventing continuous send requests to tie up threads in the application server.
- Use a hardware accelerator if the system performs several SSL handshakes. Hardware accelerators currently supported by WebSphere Application Server only increase the SSL handshake performance, not the bulk encryption and

decryption. An accelerator typically only benefits the Web server because Web server connections are short-lived. All other SSL connections in WebSphere Application Server are long-lived.

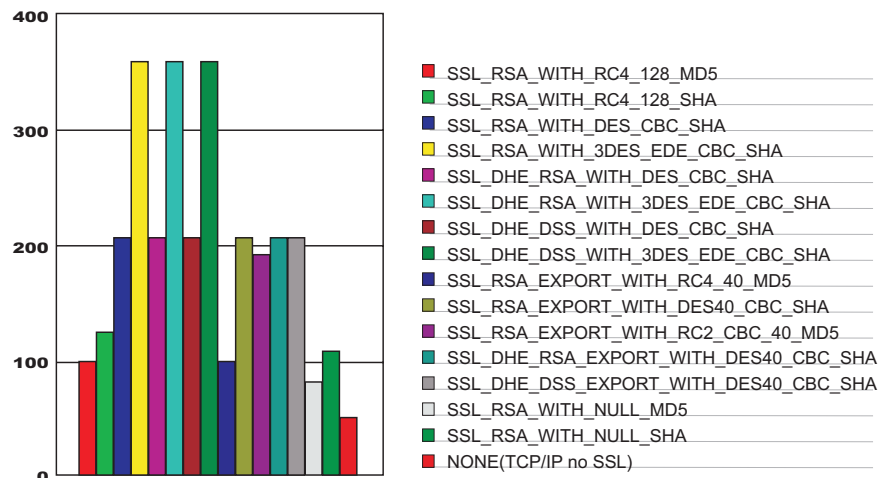
The IBM Cryptographic Coprocessor is not supported for use with WebSphere Application Server. However, you can use the IBM Cryptographic Coprocessor to improve SSL performance for other products, such as IBM HTTP Server for iSeries, which is powered by Apache.

- Use an alternative cipher suite with better performance.

The performance of a cipher suite is different with software and hardware. Just because a cipher suite performs better in software does not mean a cipher suite will perform better with hardware. Some algorithms are typically inefficient in hardware (for example, DES and 3DES), however, specialized hardware can provide efficient implementations of these same algorithms.

The performance of bulk encryption and decryption is affected by the cipher suite used for an individual SSL connection. The following chart displays the performance of each cipher suite. The test software calculating the data was Java Secure Socket Extension (JSSE) for both the client and server software, which used no crypto hardware support. The test did not include the time to establish a connection, but only the time to transmit data through an established connection. Therefore, the data reveals the relative SSL performance of various cipher suites for long running connections.

Before establishing a connection, the client enables a single cipher suite for each test case. After the connection is established, the client times how long it takes to write an integer to the server and for the server to write the specified number of bytes back to the client. Varying the amount of data had negligible effects on the relative performance of the cipher suites.



An analysis of the above data reveals the following:

- Bulk encryption performance is only affected by what follows the WITH in the cipher suite name. This is expected since the portion before the WITH identifies the algorithm used only during the SSL handshake.
- MD5 and SHA are the two hash algorithms used to provide data integrity. MD5 is generally faster than SHA, however, SHA is more secure than MD5.
- DES and RC2 are slower than RC4. Triple DES is the most secure, but the performance cost is high when using only software.

- The cipher suite providing the best performance while still providing privacy is `SSL_RSA_WITH_RC4_128_MD5`. Even though `SSL_RSA_EXPORT_WITH_RC4_40_MD5` is cryptographically weaker than `RSA_WITH_RC4_128_MD5`, the performance for bulk encryption is the same. Therefore, as long as the SSL connection is a long-running connection, the difference in the performance of high and medium security levels is negligible. It is recommended that a security level of high be used, instead of medium, for all components participating in communication only among WebSphere Application Server products. Make sure that the connections are long running connections.

### Tuning security:

Enabling security decreases performance. The following tuning parameters provide ways to minimize this performance impact.

While it is not practical to run WebSphere Application Server for z/OS without security enabled, it is possible to perform certain tuning techniques to make the Application Server run better on z/OS. These techniques are documented in detail in Security tuning tips for z/OS.

- Disable security on any application servers that do not need security. You can disable security in the administrative console by clicking **Security > Global Security** and deselecting the **Enable global security** option.
- Fine-tune the **Cache timeout** value on the Global security panel in the administrative console. For more information, see Global security settings.
- Configure the security cache properties. For more information, see “Security cache properties” on page 106.
- Enable the **Enable SSL ID tracking** option on the Session management panel in the administrative console. For more information, see Session management settings.
- Improve the performance of Web services security by downloading a Java Cryptography Extension (JCE) unlimited jurisdiction policy file that does not have restrictions on cryptography strength. For more information, see “Tuning Web services security for version 6.0.x applications” on page 65.
- Modify the RACF security settings as documented in the RACF tuning tips for z/OS article.
- Read the Secure Sockets Layer performance tips and “Tuning security configurations” on page 102 articles for more information.
- Read the “Tuning security configurations” on page 102 article for more information.

## Object Request Broker

### Object Request Broker tuning guidelines

Use the guidelines in this document any time the Object Request Broker (ORB) is used in a workload.

The ORB is used whenever enterprise beans are accessed through a remote interface. If you experience particularly high or low CPU consumption, you might have a problem with the value of one of the following parameters. Examine these core tuning parameters for every application deployment.

#### Related tasks

Chapter 5, “Tuning the application serving environment,” on page 25

This topic describes the benefits of tuning for optimal performance, highlights

the tunable parameters of the major WebSphere Application Server components, and provides insight about how these parameters affect performance.

#### **Related reference**

Object Request Broker service settings

Use this page to configure the Java Object Request Broker (ORB) service.

#### **Thread pool adjustments:**

##### **Size**

Tune the size of the ORB thread pool according to your workload. Avoid suspending threads because they have no work ready to process. If threads do not have work ready to process, CPU time is consumed by calling the `Object.wait` method, performing a context switch. Tune the thread pool size so that threads are not waiting long enough to get destroyed from remaining idle too long.

The thread pool size is dependent on your workload and system. In typical configurations, applications need 10 or fewer threads per processor.

However, if your application is performing a very slow backend request, like a request to a database system, a server thread blocks waiting for the backend request to complete. With backend requests, CPU use is fairly low. In this case, increasing the load does not increase CPU use or throughput. Your thread dumps indicate that nearly all the threads are in a call out to the backend resource. In this case, consider increasing the number of threads per processor until throughput improves and thread dumps show that the threads are in other areas of the run time besides the backend call. You should adjust the number of threads only if your backend resource is tuned correctly.

The **Allow thread allocation beyond maximum thread size** parameter also affects thread pool size, but do not use this parameter unless your back end stops for long periods of time, causing the blocking of all the run-time threads waiting for the backend system instead of processing other work that does not involve the backend system.

You can adjust the thread pool size settings in the administrative console. Click **Servers > Application servers > *server\_name* > Container services > ORB service > Thread pool**. You can adjust the minimum and maximum number of threads. See **Thread pool settings** for more information.

#### **Pass by reference:**

Specifies how the ORB passes parameters. If enabled, the ORB passes parameters by reference instead of by value, to avoid making an object copy. If you do not enable the pass by reference option, a copy of the parameter passes rather than the parameter object itself. This can be expensive because the ORB must first make a copy of each parameter object.

If the Enterprise JavaBeans (EJB) client and server are installed in the same WebSphere Application Server instance, and the client and server use remote interfaces, enabling the pass by reference option can improve performance up to 50%. The pass by reference option helps performance only where non-primitive object types are passed as parameters. Therefore, int and floats are always copied, regardless of the call model.

**CAUTION:**

Enable this property with caution because unexpected behavior can occur. If an object reference is modified by the callee, the caller's object is modified as well, since they are the same object.

If you use command-line scripting, the full name of this system property is `com.ibm.CORBA.iiop.noLocalCopies`.

|                  |                     |
|------------------|---------------------|
| <b>Data type</b> | Boolean             |
| <b>Default</b>   | Not enabled (false) |

The use of this option for enterprise beans with remote interfaces violates EJB Specification, Version 2.0 (see section 5.4). Object references passed to EJB methods or to EJB home methods are not copied and can be subject to corruption.

Consider the following example:

```
Iterator iterator = collection.iterator();
MyPrimaryKey pk = new MyPrimaryKey();
while (iterator.hasNext()) {
 pk.id = (String) iterator.next();
 MyEJB myEJB = myEJBHome.findByPrimaryKey(pk);
}
```

In this example, a reference to the same `MyPrimaryKey` object passes into WebSphere Application Server with a different ID value each time. Running this code with pass by reference enabled causes a problem within the application server because multiple enterprise beans are referencing the same `MyPrimaryKey` object. To avoid this problem, set the `com.ibm.websphere.ejbcontainer.allowPrimaryKeyMutation` system property to true when the pass by reference option is enabled. Setting the pass by reference option to true causes the EJB container to make a local copy of the `PrimaryKey` object. As a result, however, a small portion of the performance advantage of setting the pass by reference option is lost.

As a general rule, any application code that passes an object reference as a parameter to an enterprise bean method or to an EJB home method must be scrutinized to determine if passing that object reference results in loss of data integrity or in other problems.

After examining your code, you can enable the pass by reference option by setting the `com.ibm.CORBA.iiop.noLocalCopies` system property to true. You can also enable the pass by reference option in the administrative console. Click **Servers > Application servers > *server\_name* > Container services > ORB Service** and select **Pass by reference**.

**Fragment size:** The ORB separates messages into fragments to send over the ORB connection. You can configure this fragment size through the `com.ibm.CORBA.FragmentSize` parameter.

To determine and change the size of the messages that transfer over the ORB and the number of required fragments, perform the following steps:

1. In the administrative console, enable ORB tracing in the ORB Properties page. See Object Request Broker service settings for more information.
2. Enable ORBRas tracing from the logging and tracing page.
3. Increase the trace file sizes because tracing can generate a lot of data.

4. Restart the server and run at least one iteration (preferably several) of the case that you are measuring.
5. Look at the traceable file and do a search for `Fragment to follow: Yes`.

This message indicates that the ORB transmitted a fragment, but it still has at least one remaining fragment to send before the entire message arrives. A `Fragment to follow: No` value indicates that the particular fragment is the last in the entire message. This fragment can also be the first, if the message fit entirely into one fragment.

If you go to the spot where `Fragment to follow: Yes` is located, you find a block that looks similar to the following example:

```

Fragment to follow: Yes
Message size: 4988 (0x137C)
--
Request ID: 1411

```

This example indicates that the amount of data in the fragment is 4988 bytes and the Request ID is 1411. If you search for all occurrences of Request ID: 1411, you can see the number of fragments that are used to send that particular message. If you add all the associated message sizes, you have the total size of the message that is being sent through the ORB.

6. You can configure the fragment size by setting the `com.ibm.CORBA.FragmentSize` property. See Object Request Broker custom properties for more information.

**Interceptors:** Interceptors are ORB extensions that can set up the context before the ORB runs a request. For example, the context might include transactions or activity sessions to import. If the client creates a transaction, and then flows the transaction context to the server, then the server imports the transaction context onto the server request through the interceptors.

Most clients do not start transactions or activity sessions, so most systems can benefit from removing the interceptors that are not required.

To remove the interceptors, manually edit the `server.xml` file and remove the interceptor lines that are not needed from the ORB section.

**Connection Cache Adjustments:** Depending on an application server's workload, and throughput or response-time requirements, you might need to adjust the size of the ORB's connection cache. Each entry in the connection cache is an object that represents a distinct TCP/IP socket endpoint, identified by the hostname or TCP/IP address, and the port number used by the ORB to send a GIOP request or a GIOP reply to the remote target endpoint. The purpose of the connection cache is to minimize the time required to establish a connection by reusing ORB connection objects for subsequent requests or replies. (The same TCP/IP socket is used for the request and corresponding reply.)

For each application server, the number of entries in the connection cache relates directly to the number of concurrent ORB connections. These connections consist of both the inbound requests made from remote clients and outbound requests made by the application server. When the server-side ORB receives a connection request, it uses an existing connection from an entry in the cache, or establishes a new connection and adds an entry for that connection to the cache.

The ORB Connection cache maximum and Connection cache minimum properties are used to control the maximum and minimum number of entries in the connection cache at a given time. When the number of entries reaches the value specified for the Connection cache maximum property, and a new connection is needed, the ORB creates the requested connection, adds an entry to the cache and searches for and attempts to remove up to five inactive connection entries from the cache. Because the new connection is added before inactive entries are removed, it is possible for the number of cache entries to temporarily exceed the value specified for the Connection cache maximum property.

An ORB connection is considered inactive if the TCP/IP socket stream is not in use and there are no GIOP replies pending for any requests made on that connection. As the application workload diminishes, the ORB closes the connections and removes the entries for these connections from the cache. The ORB continues to remove entries from the cache until the number of remaining entries is at or below the value specified for the Connection cache maximum property. The number of cache entries is never less than the value specified for the Connection cache minimum property, which must be at least five connections less than the value specified for the Connection cache maximum property.

Adjustments to the connection cache in the client-side ORB are usually not necessary because only a small number of connections are made on that side.

**JNI Reader Threads:** By default, the ORB uses a Java thread for processing each inbound connection request it receives. As the number of concurrent requests increases, the storage consumed by a large number of reader threads increases and can become a bottleneck in resource-constrained environments. Eventually, the number of Java threads created can cause out-of-memory exceptions if the number of concurrent requests exceeds the system's available resources.

To help address this potential problem, you can configure the ORB to use JNI reader threads where a finite number of reader threads, implemented using native OS threads instead of Java threads, are created during ORB initialization. JNI reader threads rely on the native OS TCP/IP asynchronous mechanism that enables a single native OS thread to handle I/O events from multiple sockets at the same time. The ORB manages the use of the JNI reader threads and assigns one of the available threads to handle the connection request, using a round-robin algorithm. Ordinarily, JNI reader threads should only be configured when using Java threads is too memory-intensive for your application environment.




The number of JNI reader threads you should allocate for an ORB depends on many factors and varies significantly from one environment to another, depending on available system resources and workload requirements. The following potential benefits might be achieved if you use JNI threads:

- Because a fixed number of threads is allocated, memory usage is reduced. This reduction provides significant benefit in environments with unusually large and sustained client-request workloads.
- The time needed to dynamically create and destroy Java threads is eliminated because a fixed number of JNI threads is created and allocated during ORB initialization.
- Each JNI thread can handle up to 1024 socket connections and interacts directly with the asynchronous I/O native OS mechanism, which might provide enhanced performance of network I/O processing.

## Learn about WebSphere programming extensions

Use this section as a starting point to investigate the WebSphere programming model extensions for enhancing your application development and deployment.

See Learn about WebSphere applications: Overview and new features for an introduction to each WebSphere extension.

|                       |  |  |  |
|-----------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| ActivitySessions      | How do I?...                                                                      | Overview                                                                            | Samples                                                                             |
| Application profiling | How do I?...                                                                      | Overview                                                                            | Samples                                                                             |
| Asynchronous beans    | How do I?...                                                                      | Overview                                                                            | Samples                                                                             |
| Dynamic caching       | How do I?...                                                                      | Overview                                                                            |                                                                                     |
| Dynamic query         | How do I?...                                                                      | Overview                                                                            | Samples                                                                             |
| Internationalization  | How do I?...                                                                      | Overview                                                                            | Samples                                                                             |
| Object pools          | How do I?...                                                                      | Overview                                                                            |                                                                                     |
| Scheduler             | How do I?...                                                                      | Overview                                                                            | Samples                                                                             |
| Startup beans         | How do I?...                                                                      | Overview                                                                            |                                                                                     |
| Work areas            | How do I?...                                                                      | Overview                                                                            |                                                                                     |

### Related concepts

Learn about WebSphere applications: Overview and new features

Use the **Learn about WebSphere applications** section as a starting point to study technologies used in and by applications deployed on the application server.

## Dynamic cache

### Managing cache entries stored on a disk:

Use this page to set Java virtual machine (JVM) custom properties to maintain cache entries that are saved to disk.

### Steps for this task

You can set the custom properties globally to affect all cache instances, or you can set the custom property on a single cache instance. In most cases, set the properties on the individual cache instances. To set the custom properties on the default cache instance, use the global option. If you set the same property both globally and on a cache instance, the value that is set on the cache instance overrides the global value.

To configure the custom properties on a single object cache instance or servlet cache instance, perform the following steps:

1. In the administrative console, click one of the following paths:
  - To configure a servlet cache instance, click **Resources > Cache instances > Servlet cache instances > *servlet\_cache\_instance\_name* > Custom properties > New**.



- To configure an object cache instance, click **Resources > Cache instances > Object cache instances > *object\_cache\_instance\_name* > Custom properties > New**.
2. Type the name of the custom property. When configuring these custom properties on a single cache instance, you do not use the full property path. For example, type `htodCleanupFrequency` to configure the `com.ibm.ws.cache.CacheConfig.htodCleanupFrequency` custom property.
  3. Type a valid value for the property in the **Value** field.
  4. Save the property and restart WebSphere Application Server.

To configure the custom property globally across all configured cache instances, perform the following steps:

1. In the administrative console, click **Servers > Application servers > *server\_name* > Java and process management > Process definition > Java virtual machine > Custom properties > New**.
2. Type the name of the custom property (for example, `com.ibm.ws.cache.CacheConfig.htodCleanupFrequency`) in the **Name** field.
3. Type a valid value for the property in the **Value** field.
4. Save the property and restart WebSphere Application Server.

### **com.ibm.ws.cache.CacheConfig.htodCleanupFrequency**

Use this property to change the amount of time between disk cache cleanup.

By default, the disk cache cleanup is scheduled to run at midnight to remove expired cache entries and cache entries that have not been accessed in the past 24 hours. However, if you have thousands of cache entries that might expire within one or two hours, the files that are in the disk cache can grow large and become unmanageable. Use the `com.ibm.ws.cache.CacheConfig.htodCleanupFrequency` custom property to change the time interval between disk cache cleanup.

|         |                                                                                             |
|---------|---------------------------------------------------------------------------------------------|
| Units   | minutes<br><br>For example, a value of 60 means 60 minutes between each disk cache cleanup. |
| Default | 0<br><br>The disk cache cleanup occurs at midnight every 24 hours.                          |

### **Tune the delay offload function**

Use these properties to tune the delay offload function for the disk cache. The delay offload function uses extra memory buffers for dependency IDs and templates to delay the disk offload and minimize the input and output operations. However, if most of your cache IDs are longer than 100 bytes, the delay offload function might use too much memory. Use any combination of the following properties to tune your configuration:

- To increase or decrease the in-memory limit of cache IDs for dependency ID and template buffers, use the `com.ibm.ws.cache.CacheConfig.htodDelayOffloadEntriesLimit` custom property.

- To disable the disk cache delay offload function, use the `com.ibm.ws.cache.CacheConfig.htodDelayOffload` custom property. Disabling this property saves all cache entries to disk immediately after removing them from the memory cache.

### **com.ibm.ws.cache.CacheConfig.htodDelayOffloadEntriesLimit**

Use this property to specify the number of different cache IDs that can be saved in memory for the dependency ID and template buffers. Consider increasing this value if you have a lot of memory in your server and you want to increase the performance of your disk cache.

|         |                                                                                                                                                         |
|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Units   | number of cache IDs<br><br>For example, a value of 1000 means that each dependency ID or template ID can have up to 1000 different cache IDs in memory. |
| Default | 1000                                                                                                                                                    |
| Minimum | 100                                                                                                                                                     |

### **com.ibm.ws.cache.CacheConfig.htodDelayOffload**

Use this property to specify if extra memory buffers should be used in memory for dependency IDs and templates to delay disk offload and to minimize input and output operations to the disk. This property is enabled by default. However, consider disabling this property if your cache IDs are larger than 100 bytes because this option might use too much memory when it buffers your data. If you set this property to `false`, all the cache entries are copied to disk immediately after they are removed from the memory cache.

|         |      |
|---------|------|
| Default | true |
|---------|------|

#### **Related reference**

Troubleshooting tips for the dynamic cache service

The dynamic cache service works within an application server Java virtual machine (JVM), intercepting calls to cacheable objects. This article describes some common runtime and configuration problems and remedies.

#### **Tuning dynamic cache with the cache monitor:**

Use this task to interpret cache monitor statistics to improve the performance of the dynamic cache service.

Verify that dynamic cache is enabled and that the cache monitor application is installed on your application server. See [Displaying cache information](#) to configure the cache monitor application.

Use the cache monitor to watch cache hits versus misses. By comparing these two values, you can determine how much dynamic cache is helping your application, and if you can take any additional steps to further improve performance and decrease the cost of processing for your application server.

1. Start cache monitor and click on **Cache Statistics**. You can view the following cache statistics:

| <u>Cache statistic</u> | <u>Description</u>                                                                                                                                  |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Cache Size             | The maximum number of entries that the cache can hold.                                                                                              |
| Used Entries           | The number of cache entries used.                                                                                                                   |
| Cache Hits             | The number of request responses that are served from the cache.                                                                                     |
| Cache Misses           | The number of request responses that are cacheable but cannot be served from the cache.                                                             |
| LRU Evictions          | The number of cache entries removed to make room for new cache entries.                                                                             |
| Explicit Removals      | The number of cache entries removed or invalidated from the cache based on cache policies or were deleted from the cache through the cache monitor. |

2. You can also view the following cache configuration values:

| <u>Cache configuration value</u> | <u>Description</u>                                                                                                                                                                                                                          |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Default priority                 | Specifies the default priority for all cache entries. Lower priority entries are moved from the cache before higher priority entries when the cache is full. You can specify the priority for individual cache entries in the cache policy. |
| Servlet Caching Enabled          | If servlet caching is enabled, results from servlets and JavaServer Pages (JSP) files are cached. See Configuring servlet caching for more information.                                                                                     |
| Disk Offload Enabled             | Specifies if entries that are being removed from the cache are saved to disk. See Configuring dynamic cache disk offload for more information.                                                                                              |

- Wait for the application server to add data to the cache. You want the number of used cache entries in the cache monitor to be as high as it can go. When the number of used entries is at its highest, the cache can serve responses to as many requests as possible.
- When the cache has a high number of used entries, reset the statistics. Watch the number of cache hits versus cache misses. If the number of hits is far greater than the number of misses, your cache configuration is optimal. You do not need to take any further actions. If you find a higher number of misses with a lower number of hits, the application server is working hard to generate responses instead of serving the request using a cached value. The application server might be making database queries, or running logic to respond to the requests.
- If you have a large number of cache misses, increase the number of cache hits by improving the probability that a request can be served from the cache. To improve the number of cache hits, you can increase the cache size or configure additional cache policies. See Dynamic cache service settings to increase the cache size and Configuring cacheable objects with the cachespec.xml file for information about configuring cache policies.

By using the cache monitor application, you optimized the performance of the dynamic cache service.

See Task overview: Using the dynamic cache service to improve performance for more information about the dynamic cache.

#### **Related concepts**

Cache monitor

Cache monitor is an installable Web application that provides a real-time view of the current state of dynamic cache. You use it to help verify that dynamic cache is operating as expected. The only way to manipulate the data in the cache is by using the cache monitor. It provides a GUI interface to manually change data.

#### **Related tasks**

Troubleshooting the dynamic cache service

## **Work area**

**Work area service performance considerations:** The work area service is designed to address complex data passing patterns that can quickly grow beyond convenient maintenance. A *work area* is a note pad that is accessible to any client that is capable of looking up Java Naming Directory Interface (JNDI). After a work area is established, data can be placed there for future use in any subsequent method calls to both remote and local resources.

You can utilize a work area when a large number of methods require common information or if information is only needed by a method that is significantly further down the call graph. The former avoids the need for complex parameter passing models where the number of arguments passed becomes excessive and hard to maintain. You can improve application function by placing the information in a work area and subsequently accessing it independently in each method, eliminating the need to pass these parameters from method to method. The latter case also avoids unnecessary parameter passing and helps to improve performance by reducing the cost of marshalling and de-marshalling these parameters over the Object Request Broker (ORB) when they are only needed occasionally throughout the call graph.

When attempting to maximize performance by using a work area, cache the UserWorkArea partition that is retrieved from JNDI wherever it is accessed. You can reduce the time spent looking up information in JNDI by retrieving it once and keeping a reference for the future. JNDI lookup takes time and can be costly.

Additional caching mechanisms available to a user-defined partition are defined by the configuration property, "Deferred Attribute Serialization". This mechanism attempts to minimize the number of serialization and deserialization calls. See Work area partition service for further explanation of this configuration attribute.

The `maxSendSize` and `maxReceiveSize` configuration parameters can affect the performance of the work area. Setting these two values to 0 (zero) effectively turns off the policing of the size of context that can be sent in a work area. This action can enhance performance, depending on the number of nested work areas an application uses. In applications that use only one work area, the performance enhancement might be negligible. In applications that have a large number of nested work areas, there might be a performance enhancement. However, a user must note that by turning off this policing it is possible that an extremely large amount of data might be sent to a server.

Performance is degraded if you use a work area as a direct replacement to passing a single parameter over a single method call. The reason is that you incur more

overhead than just passing that parameter between method calls. Although the degradation is usually within acceptable tolerances and scales similarly to passing parameters with regard to object size, consider degradation a potential problem before utilizing the service. As with most functional services, intelligent use of the work areas yields the best results.

The work area service is a tool to simplify the job of passing information from resource to resource, and in some cases can improve performance by reducing the overhead that is associated with a parameter passing when the information is only sparsely accessed within the call graph. Caching the instance retrieved from JNDI is important to effectively maximize performance during runtime.

**Related concepts**

Work area partition service

The work area partition service is an extension of the work area service that allows the creation of multiple custom work areas. The work area partition service is an optional service to users. Any user that currently uses the work area service and the UserWorkArea partition can continue using it in the same manner. The UserWorkArea partition is created automatically (if it has not been disabled) by the work area partition service. By allowing a user the option to create their own work area partition through the work area partition service, they can have more control over configuration and access to their partition.

**Related reference**

UserWorkArea interface

The Work area partition manager interface

**Related information**

Work area service - Overview



---

## Chapter 6. Troubleshooting performance

This topic illustrates that solving a performance problem is an iterative process and shows how to troubleshoot performance problems.

It is recommended that you review the tuning parameter hot list before reading this topic.

Solving a performance problem is frequently an iterative process of:

- Measuring system performance and collecting performance data
- Locating a bottleneck
- Eliminating a bottleneck

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

### Measuring system performance and collecting performance data

Begin by choosing a *benchmark*, 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, optimize code paths, and so on. System performance during the warm-up period is usually much slower than after the warm-up period. The benchmark must be able to generate work that warms up the system prior to 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 only occurs when a large number of 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, or the measurements are made during the warm-up period, or that the system is running additional workloads.

Several tools facilitate benchmark development. The tools range from tools that simply invoke a URL to script-based products that can interact with dynamic data generated by the application. 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 needs to 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 is provided by WebSphere Performance Monitoring Infrastructure (PMI) helps to monitor and tune the application server performance. Request metrics is another sources of performance data that is provided by WebSphere Application Server. Request

metrics allows a request to be timed at WebSphere Application Server component boundaries, enabling a determination of the time that is spent in each major component.

### Locating a bottleneck

Consult the following scenarios and suggested solutions:

- **Scenario:** Poor performance occurs with only a single user.

**Suggested solution:** Utilize request metrics to determine how much each component is contributing to the overall response time. Focus on the component accounting for the most time. Use Tivoli Performance Viewer to check for resource consumption, including frequency of garbage collections. You might need code profiling tools to isolate the problem to a specific method.

- **Scenario:** Poor performance only occurs with multiple users.

**Suggested solution:** Check to determine if any systems have high CPU, network or disk utilization and address those. For clustered configurations, check for uneven loading across cluster members.

- **Scenario:** None of the systems seems to have a CPU, memory, network, or disk constraint but performance problems occur with multiple users.

**Suggested solutions:**

- Check that work is reaching the system under test. Ensure that some external device does not limit the amount of work reaching the system. Tivoli Performance Viewer helps determine the number of requests in the system.
- A thread dump might reveal a bottleneck at a synchronized method or a large number of threads waiting for a resource.
- Make sure that enough threads are available to process the work both in IBM HTTP Server, database, and the application servers. Conversely, too many threads can increase resource contention and reduce throughput.
- Monitor garbage collections with Tivoli Performance Viewer or the `verbosegc` option of your Java virtual machine. Excessive garbage collection can limit throughput.

### Eliminating a bottleneck

Consider the following methods to eliminate a bottleneck:

- Reduce the demand
- Increase resources
- Improve workload distribution
- Reduce synchronization

Reducing the demand for resources can be accomplished in several ways. Caching can greatly reduce the use of system resources by returning a previously cached response, thereby avoiding the work needed to construct the original response.

Caching is supported at several points in the following systems:

- IBM HTTP Server
- Command
- Enterprise bean
- Operating system

Application code profiling can lead to a reduction in the CPU demand by pointing out hot spots you can optimize. IBM Rational and other companies have tools to perform code profiling. An analysis of the application might reveal areas where some work might be reduced for some types of transactions.



Change tuning parameters to increase some resources, for example, the number of file handles, while other resources might need a hardware change, for example, more or faster CPUs, or additional application servers. Key tuning parameters are described for each major WebSphere Application Server component to facilitate solving performance problems. Also, the performance advisors can provide advice on tuning a production system under a real or simulated load.

Workload distribution can affect performance when some resources are underutilized and others are overloaded. WebSphere Application Server workload management functions provide several ways to determine how the work is distributed. Workload distribution applies to both a single server and configurations with multiple servers and nodes.

Workload distribution can affect performance when some resources are underutilized and others are overloaded. WebSphere Application Server workload management functions provide several ways to determine how the work is distributed. Workload distribution applies to both a single server and configurations with multiple servers and nodes.

Some critical sections of the application and server code require synchronization to prevent multiple threads from running this code simultaneously and leading to incorrect results. Synchronization preserves correctness, but it can also reduce throughput when several threads must wait for one thread to exit the critical section. When several threads are waiting to enter a critical section, a thread dump shows these threads waiting in the same procedure. Synchronization can often be reduced by: changing the code to only use synchronization when necessary; reducing the path length of the synchronized code; or reducing the frequency of invoking the synchronized code.

#### **Additional references**

IBM WebSphere V6: Performance, Scalability, and High Availability Concepts

WebSphere Application Server Performance Web site

All SPEC jAppServer2004 Results Published by SPEC.



---

## Notices

References in this publication to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program, or service. Evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, is the user's responsibility.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Intellectual Property & Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
USA



---

## Trademarks and service marks

For trademark attribution, visit the IBM Terms of Use Web site (<http://www.ibm.com/legal/us/>).