



## The Pentaho Performance Tuning Guide



This document is copyright © 2011 Pentaho Corporation. No part may be reprinted without written permission from Pentaho Corporation. All trademarks are the property of their respective owners.

## Help and Support Resources

If you have questions that are not covered in this guide, or if you would like to report errors in the documentation, please contact your Pentaho technical support representative.

Support-related questions should be submitted through the Pentaho Customer Support Portal at <http://support.pentaho.com>.

For information about how to purchase support or enable an additional named support contact, please contact your sales representative, or send an email to [sales@pentaho.com](mailto:sales@pentaho.com).

For information about instructor-led training on the topics covered in this guide, visit <http://www.pentaho.com/training>.

## Limits of Liability and Disclaimer of Warranty

The author(s) of this document have used their best efforts in preparing the content and the programs contained in it. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, express or implied, with regard to these programs or the documentation contained in this book.

The author(s) and Pentaho shall not be liable in the event of incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of the programs, associated instructions, and/or claims.

## Trademarks

Pentaho (TM) and the Pentaho logo are registered trademarks of Pentaho Corporation. All other trademarks are the property of their respective owners. Trademarked names may appear throughout this document. Rather than list the names and entities that own the trademarks or insert a trademark symbol with each mention of the trademarked name, Pentaho states that it is using the names for editorial purposes only and to the benefit of the trademark owner, with no intention of infringing upon that trademark.

## Company Information

Pentaho Corporation  
Citadel International, Suite 340  
5950 Hazeltine National Drive  
Orlando, FL 32822  
Phone: +1 407 812-OPEN (6736)  
Fax: +1 407 517-4575  
<http://www.pentaho.com>

E-mail: [communityconnection@pentaho.com](mailto:communityconnection@pentaho.com)

Sales Inquiries: [sales@pentaho.com](mailto:sales@pentaho.com)

Documentation Suggestions: [documentation@pentaho.com](mailto:documentation@pentaho.com)

Sign-up for our newsletter: <http://community.pentaho.com/newsletter/>

# Contents

Introduction.....	4
System Requirements.....	5
Pentaho BI Server Performance Hacks.....	6
Switching to a File-Based Solution Repository.....	6
Turning Off Audit Logging.....	6
Pentaho Reporting Performance Tweaks.....	8
Caching Report Content.....	8
Result Set Caching.....	8
Streamlining Printed Output.....	9
Paginated Exports.....	9
Table Exports.....	10
HTML Exports.....	10
Pentaho Reporting Configuration Files.....	11
Pentaho Data Integration Performance-Tuning Tips.....	12
Upgrading to the Latest Release.....	12
Pentaho Data Integration Performance Tuning Tips.....	12
Limiting In-Memory Log Output.....	15
Pentaho Analysis (Mondrian) Performance Testing and Improvement.....	17
ROLAP Performance-Tuning and Design Theory.....	17
Partitioning High-Cardinality Dimensions.....	17
Tuple Segment Caching.....	17
Vertical Resource Scaling.....	18
Unfinished Content.....	18
Appropriate Upgrading.....	18
Switching to 64-bit.....	18
Horizontal Resource Scaling.....	19
Clustering the Application Server.....	19
Clustering Requirements.....	19
Sharing the Solution Repository.....	19
Installing and Configuring Apache as a Load Balancer.....	20
Tomcat Configuration.....	22
Copying WAR Files to the Nodes.....	23
Starting and Testing the Cluster.....	24
Changing the Java VM Memory Limits.....	25
Increasing the Memory Limit in JBoss.....	25
Increasing the Memory Limit in Tomcat.....	25
Testing BI Server Scalability.....	27

# Introduction

---

This guide is designed to help you discover where your BI Server performance bottlenecks are, along with instructions and suggestions on how to address them.

There are many ways to improve the speed and efficiency of Pentaho software documented in this guide. Each applies to a specific situation and **should never be blindly applied**. Some of the performance tweaks herein will remove functionality and in some cases security from your BI Server instance. Others will assign more system resources to the BI Server, which could in turn impact other services running on the same machine.

To put it more plainly: **Performance always comes at the cost of one or more of: functionality, security, or resources.**

The tips and tricks listed in this guide are meant as an initial set of self-service tasks for improving BI Suite performance. There are much more advanced techniques that may improve performance, but require code changes or major surgical changes to Pentaho software, none of which should ever be attempted without qualified assistance. These techniques are not included in this guide for safety reasons. A Pentaho partner or consultant can assist you with more advanced performance improvements, if required.

# System Requirements

---

Before continuing with this guide, you should already have a working and tested Pentaho BI Suite 3.8.0 installation. There are no operating system or hardware requirements beyond those implicit in specific performance-tuning tips. Requirements are listed on an individual basis for each tip.



**Note:** It may be possible to use other versions of the BI Suite, and other versions of the software mentioned in this guide, such as Tomcat and Apache, but those configurations are untested. If you stray from the tested configuration, be prepared to dynamically modify the instructions in this guide to accommodate your situation, and be warned that your Pentaho support representative may not be able to assist with your unsupported configuration.

# Pentaho BI Server Performance Hacks

---

The Pentaho BI Server ships in a condition designed to work well for the majority of customers. However, deployments that drift toward opposite extremes -- very large and very small -- will need to adjust certain settings, and possibly even remove certain unused functionality, in order to achieve the desired performance goals without adding hardware.

Read through the subsections below and decide which ones apply to your scenario.

## Switching to a File-Based Solution Repository

---

The Pentaho BI Server and Pentaho Enterprise Console must be stopped before executing these instructions.

This procedure changes your default database solution repository into a strictly file-based repository. There will no longer be a database mirroring the content in your pentaho-solutions directory, and there will be no way of isolating or securing BI content within the BI Server. However, removing the database overhead means that there will be a substantial performance increase in many instances. This procedure does not, in itself, remove security from the BI Platform, but it does remove access control lists from all content.

1. Edit the `/pentaho-solutions/system/pentahoObjects.spring.xml` file.
2. Comment out the current **ISolutionRepository** line, and uncomment the similar line above it.

Alternatively, you can switch the value of **ISolutionRepository** from **org.pentaho.platform.repository.solution.dbbased.DbBasedSolutionRepository** to **org.pentaho.platform.repository.solution.filebased.FileBasedSolutionRepository**.

```
<!-- Uncomment the following line to use a filesystem-based repository. Note: does not support ACLs. -->
<bean id="ISolutionRepository"
  class="org.pentaho.platform.repository.solution.filebased.FileBasedSolutionRepository"
  scope="session" />
<!-- Uncomment the following line to use a filesystem/db-based repository with meta information stored in a db -->
<!-- <bean id="ISolutionRepository"
  class="org.pentaho.platform.repository.solution.dbbased.DbBasedSolutionRepository"
  scope="session" /> -->
```

3. Comment out the **IMetadataDomainRepository** line, and uncomment the similar line below it.

Alternatively, you can switch the value of **IMetadataDomainRepository** from **org.pentaho.platform.plugin.services.metadata.SecurityAwareMetadataDomainRepository** to **org.pentaho.platform.plugin.services.metadata.MetadataDomainRepository**.

```
<!-- <bean id="IMetadataDomainRepository"
  class="org.pentaho.platform.plugin.services.metadata.SecurityAwareMetadataDomainRepository"
  scope="singleton"/> -->
<!-- Use this schema factory to disable PMD security -->
<bean id="IMetadataDomainRepository"
  class="org.pentaho.platform.plugin.services.metadata.MetadataDomainRepository"
  scope="singleton"/>
```

4. Save and close the file.

You've now switched over to a file-based solution repository. You can safely restart your BI Server and Pentaho Enterprise Console server.

## Turning Off Audit Logging

---

Your BI Server must be stopped before performing this procedure.

While audit logging can be useful for monitoring BI Server activity and performance, the act of collecting the necessary audit data can introduce significant memory overhead with the solution database. Follow the instructions below to disable audit logging in the BI Server.



**Note:** Performing this task will disable all audit functions in the BI Server's administration interface.

1. Open the `/pentaho-solutions/system/pentahoObjects-spring.xml` file with a text editor.
2. Locate the following line:

```
<bean id="IAuditEntry"  
  class="org.pentaho.platform.engine.services.audit.AuditSQLEntry" scope="singleton" />
```

3. Replace that line with this one:

```
<bean id="IAuditEntry" class="org.pentaho.platform.engine.core.audit.NullAuditEntry"  
  scope="singleton" />
```

4. Save and close the file
5. Using a database management tool or command line interface, connect to the Pentaho **hibernate** database.
6. Truncate (but do not drop) the following tables:
  - PRO\_AUDIT
  - PRO\_AUDIT\_TRANSFORM\_TRACKER
7. Exit your database utility and restart the BI Server.


If you need to reverse this process later, you can replace the old configuration line shown above, and use the **Initialize From File** audit function in the **Services** section of the **Administration** tab in the Pentaho Enterprise Console to restore the audit log table structure.

# Pentaho Reporting Performance Tweaks

Pentaho Reporting's default configuration makes certain assumptions about system resources and the size, features, and details of reports that may not meet your specific requirements. If you have large inline subreports, or many parameters, you can run into performance bottlenecks. Fortunately, many performance problems can be mitigated through specific engine and report options. Refer to the sections below that apply to your scenario.

## Caching Report Content

You can cache the result sets of parameterized reports so that every time you change a parameter during your user session (all caching is on a per-session basis) you don't have to retrieve a new result set. By default, Pentaho Reporting has result set caching turned on, but you may find some advantage in turning it off or changing the cache thresholds and settings.

 **Note:** When you publish a report to the BI Server, you switch cache and engine configurations from the local Report Designer versions of **ehcache.xml** and **classic-engine.properties** to the server's version inside the Pentaho WAR. These configurations may not be the same, so if you have made changes to the result set cache settings locally, you may want to port those changes over to the BI Server as well.

## Result Set Caching

When rendered, a parameterized report must account for every dataset required for every parameter. Every time a parameter field changes, every dataset is recalculated, which can negatively impact performance.

You can avoid gratuitous dataset recalculations by caching parameter datasets. This is accomplished through the EHCache framework built into the BI Server. You can configure specific settings for published reports by editing the **ehcache.xml** file in the `/WEB-INF/classes/` directory inside of the `pentaho.war`. The relevant element is:

Anything containing complex objects is not cached (CLOB and BLOB data types); neither are results coming from a scripting dataset, a Java method call, a table data source, an external data source (computed in an action sequence), or a CDA data source. In all of these cases there is either no point in caching because it would be more expensive than recalculating, or because there are not enough hints available in the involved parameters.

```
<!--
  Defines a cache used by the reporting engine to hold small datasets.
  This cache can be configured to have a separate instance for each
logged in user via the
  global report configuration. This per-user cache is required if role
  or other security and
  filter information is used in ways invisible for the reporting
engine.
-->
<cache name="report-dataset-cache"
  maxElementsInMemory="50"
  eternal="false"
  overflowToDisk="false"
  timeToIdleSeconds="300"
  timeToLiveSeconds="600"
  diskPersistent="false"
  diskExpiryThreadIntervalSeconds="120"
/>
```

The other side of the coin is that if a cache exists for too long, when the data source is updated it may not reflect in the report output because it's still using old data. So there is a balance between performance and accuracy that you must tune to your needs.

## Result Set Cache Options

These **classic-engine.properties** options control result set caching in parameterized reports.

Option	Purpose	Possible Values
<code>org.pentaho.reporting</code> <code>.platform.plugin.cache.</code> <code>PentahoDataCache.CachableRowLimit</code>	Number of rows in the dataset that will be cached; the higher the number, the	Integer; default value is <b>10000</b> .



Option	Purpose	Possible Values
	larger the cache and the more disk space is used while the cache is active.	

## Streamlining Printed Output

Pentaho Reporting's overall performance is chiefly affected by the amount of printed content that it has to generate. The more content you generate, the more time the Reporting engine will take to perform all layout computations.

Large inline subreports are notorious for poor performance. This is because the layouted output of an inline subreport is always stored in memory. The master report's layouting pauses until the subreport is fully generated, then it's inserted into the master report's layout model and subsequently printed. Memory consumption for this layouting model is high because the full layout model is kept in memory until the report is finished. If there is a large amount of content in the subreport, you will run into "out of memory" exceptions.

An inline subreport that consumes the full width of the root-level band should be converted into a banded subreport. Banded subreports are layouted and all output is generated while the subreport is processed. The memory footprint for that is small because only the active band or the active page has to be held in memory.

When images are embedded from remote servers (HTTP/FTP sources), you must ensure that the server produces a **LastModifiedDate** header. The Reporting engine uses that header as part of its caching system, and if it is missing, the remote images will not be cached, forcing the engine to retrieve them every time they're needed.

Caching must be configured properly via a valid ehcache configuration file, which is stored in the Pentaho Web app in the `/WEB-INF/classes/` directory. If caching is disabled or misconfigured, then there will be performance problems when loading reports and resources.

Within Pentaho Reporting there are three output types, each with its own memory and CPU consumption characteristics. Each is listed below with an explanation of how it is optimized.

## Paginated Exports

A pageable report generates a stream of pages. Each page has the same height, even if the page is not fully filled with content. When a page is filled, the layouted page will be passed over to the output target to render it in either a Graphics2D or a streaming output (PDF, Plaintext, HTML, etc.) context.

### Page break methods

When the content contains a manual pagebreak, the page will be considered full. If the pagebreak is a **before-print** break, then the break will be converted to an **after-break**, the internal report states will be rolled back, and the report processing restarts to regenerate the layout with the new constraints. A similar rollback happens if the current band does not fit on the page. Because of this, you would generally prefer break-before over break-after.

So for large reports, you might consider removing manual page breaks and limiting the width of bands.

### Page states

When processing a pageable report, the reporting engine assumes that the report will be run in **interactive mode**, which allows for parameterization control. To make browsing through the pages faster, a number of page states will be stored to allow report end-users to restart output processing at the point in the report where they adjust the parameters.

Reports that are run to fully export all pages usually do not need to store those page states. A series of Report engine settings controls the number and frequency of the page states stored:

- `org.pentaho.reporting.engine.classic.core.performance.pagestates.PrimaryPoolSize=20`
- `org.pentaho.reporting.engine.classic.core.performance.pagestates.SecondaryPoolFrequency=4`
- `org.pentaho.reporting.engine.classic.core.performance.pagestates.SecondaryPoolSize=100`
- `org.pentaho.reporting.engine.classic.core.performance.pagestates.TertiaryPoolFrequency=10`

The Reporting engine uses three lists to store page states. The default configuration looks as follows:

1. The first 20 states (Pages 1 to 20) are stored in the **primary pool**. All states are stored with strong references and will not be garbage collected.
2. The next 400 states (pages 21 to 421) are stored into the **secondary pool**. Of those, every fourth state is stored with a strong reference and cannot be garbage collected as long as the report processor is open.


3. All subsequent states (pages > 421) are stored in the **tertiary pool** and every tenth state is stored as strong reference.

So for a 2000-page report, a total of about 270 states will be stored with strong references.

In server mode, the settings could be cut down to:

```
org.pentaho.reporting.engine.classic.core.performance.pagestates.PrimaryPoolSize=1
org.pentaho.reporting.engine.classic.core.performance.pagestates.SecondaryPoolFrequency=1
org.pentaho.reporting.engine.classic.core.performance.pagestates.SecondaryPoolSize=1
org.pentaho.reporting.engine.classic.core.performance.pagestates.TertiaryPoolFrequency=100
```

This reduces the number of states stored for a 2000 page report to 22, thus cutting the memory consumption for the page states to a 1/10th.


 **Note:** In PRD 3.7 full exports no longer generate page states and thus these settings will have no effect on such exports. They still affect the interactive mode.

## Table Exports

A table export produces tabular output from a fully-layouted display model. A table export cannot handle overlapping elements and therefore has to remove them.

To support layout debugging, the Reporting engine stores a lot of extra information in the layout model. This increases memory consumption but makes it easier to develop Reporting solutions. These Reporting engine debug settings should never be enabled in production environments:

- org.pentaho.reporting.engine.classic.core.modules.output.table.base.ReportCellConflicts
- org.pentaho.reporting.engine.classic.core.modules.output.table.base.VerboseCellMarkers

 **Note:** As of Report Designer version 3.7, these settings are **false** by default. Report Designer comes with its own method to detect overlapping elements and does not rely on these settings.

## HTML Exports

In HTML exports, there are a few Reporting engine settings that can affect export performance. The first is **CopyExternalImages**:

```
org.pentaho.reporting.engine.classic.core.modules.output.table.html.CopyExternalImages=true
```

This controls whether images from HTTP/HTTPS or FTP sources are linked from their original source or copied (and possibly re-encoded) into the output directory. The default is **true**; this ensures that reports always have the same image. Set to **false** if the image is dynamically generated, in which case you'd want to display the most recent view.

The **Style** and **ForceBufferedWriting** settings control how stylesheets are produced and whether the generated HTML output will be held in a buffer until the report processing is finished:

```
org.pentaho.reporting.engine.classic.core.modules.output.table.html.ForceBufferedWriting=true
```

Style information can be stored inline, or in the <head> element of the generated HTML file:

```
org.pentaho.reporting.engine.classic.core.modules.output.table.html.InlineStyles=true
```

Or in an external CSS file:

```
org.pentaho.reporting.engine.classic.core.modules.output.table.html.ExternalStyle=true
```

**ForceBufferedWriting** should be set to true if a report uses an external CSS file. Browsers request all resources they find in the HTML stream, so if a browser requests a stylesheet that has not yet been fully generated, the report cannot display correctly. It is safe to disable buffering if the styles are inline because the browser will not need to fetch an external stylesheet in that case. Buffered content will appear slower to the user than non-buffered content because browsers render partial HTML pages while data is still being received from the server. Buffering will delay that rendering until the report is fully processed on the server.

## Pentaho Reporting Configuration Files

---

The following files contain various configuration options for Pentaho Reporting. The options are not particularly self-explanatory and their value limits are not obvious; therefore, you shouldn't change any options in these files unless you are following guidelines from Pentaho documentation or are assisted by a Pentaho support or consulting representative.

File	Purpose
/pentaho/design-tools/report-designer/resources/ <b>report-designer.properties</b>	Contains options for the Report Designer client tool. It does not change any report options.
/pentaho/design-tools/report-designer/resources/ <b>classic-engine.properties</b>	Contains global report rendering options for reports generated locally from Report Designer. Some of these options can be overridden in individual reports.
/tomcat/webapps/pentaho/WEB-INF/classes/ <b>classic-engine.properties</b>	Contains global report rendering options for published reports that are generated on the BI Server. Some of these options can be overridden in individual reports.

# Pentaho Data Integration Performance-Tuning Tips

Below are some tips and tricks for improving PDI performance. Most of them involve streamlining jobs and transformations.

## Upgrading to the Latest Release

If you are using PDI version 4.0.x or 4.1, you can see some substantial performance increases in enterprise repository transactions by upgrading to version 4.1.1 or later. Pentaho put a great deal of effort into improving repository efficiency for this release.

Refer to the upgrade documents that apply to your scenario. You can obtain all upgrade materials and instructions from the Pentaho Knowledge Base.


## Pentaho Data Integration Performance Tuning Tips

The tips described here may help you to identify and correct performance-related issues associated with PDI transformations.



**Note:** All tips described in this document are associated with PDI versions 3.0.0 and later unless otherwise specified.

Step	Tip	Description
JavaScript	Turn off compatibility mode	<p>Rewriting JavaScript to use a format that is not compatible with previous versions is, in most instances, easy to do and makes scripts easier to work with and to read. By default, old JavaScript programs run in compatibility mode. That means that the step will process like it did in a previous version. You may see a small performance drop because of the overload associated with forcing compatibility. If you want make use of the new architecture, disable compatibility mode and change the code as shown below:</p> <ul style="list-style-type: none"><li>• <code>intField.getInteger() --&gt; intField</code></li><li>• <code>numberField.getNumber() --&gt; numberField</code></li><li>• <code>dateField.getDate() --&gt; dateField</code></li><li>• <code>bigNumberField.getBigNumber() --&gt; bigNumberField</code></li><li>• and so on...</li></ul> <p>Instead of Java methods, use the built-in library. Notice that the resulting program code is more intuitive. For example :</p> <ul style="list-style-type: none"><li>• checking for null is now: <code>field.isNull() --&gt; field==null</code></li></ul>

Step	Tip	Description
		<ul style="list-style-type: none"> <li>• Converting string to date:  <code>field.Clone().str2dat() --&gt; str2date(field)</code></li> <li>• and so on...</li> </ul> <p>If you convert your code as shown above, you may get significant performance benefits.</p> <p> <b>Note:</b> It is no longer possible to modify data in-place using the value methods. This was a design decision to ensure that no data with the wrong type would end up in the output rows of the step. Instead of modifying fields in-place, create new fields using the table at the bottom of the Modified JavaScript transformation.</p>
JavaScript	Combine steps	One large JavaScript step runs faster than three consecutive smaller steps. Combining processes in one larger step helps to reduce overhead.
JavaScript	Avoid the JavaScript step or write a custom plug in	Remember that while JavaScript is the fastest scripting language for Java, it is still a scripting language. If you do the same amount of work in a native step or plugin, you avoid the overhead of the JS scripting engine. This has been known to result in significant performance gains. It is also the primary reason why the Calculator step was created — to avoid the use of JavaScript for simple calculations.
JavaScript	Create a copy of a field	No JavaScript is required for this; a "Select Values" step does the trick. You can specify the same field twice. Once without a rename, once (or more) with a rename. Another trick is to use <code>B=NVL(A,A)</code> in a Calculator step where B is forced to be a copy of A. In version 3.1, an explicit "create copy of field A" function was added to the Calculator.
JavaScript	Data conversion	Consider performing conversions between data types (dates, numeric data, and so on) in a "Select Values" step (version 3.0.2 or higher). You can do this in the Metadata tab of the step.
JavaScript	Variable creation	If you have variables that can be declared once at the beginning of the transformation, make sure you put them in a separate script and mark that script as a startup script (right click on the script name in the tab). JavaScript object creation is time consuming so if you can avoid creating a new object for every row you are transforming, this

Step	Tip	Description
		will translate to a performance boost for the step.
Not Applicable	Launch several copies of a step	<p>There are two important reasons why launching multiple copies of a step may result in better performance:</p> <ol style="list-style-type: none"> <li>1. The step uses a lot of CPU resources and you have multiple processor cores in your computer. Example: a JavaScript step</li> <li>2. Network latencies and launching multiple copies of a step can reduce average latency. If you have a low network latency of say 5ms and you need to do a round trip to the database, the maximum performance you get is 200 (x5) rows per second, even if the database is running smoothly. You can try to reduce the round trips with caching, but if not, you can try to run multiple copies. Example: a database lookup or table output</li> </ol>
Not Applicable	Manage thread priorities	In versions 3.0.2 and higher, this feature that is found in the "Transformation Settings" dialog box under the (Misc tab) improves performance by reducing the locking overhead in certain situations. This feature is enabled by default for new transformations that are created in recent versions, but for older transformations this can be different.
Select Value	If possible, don't remove fields in Select Value	Don't remove fields in Select Value unless you must. It's a CPU-intensive task as the engine needs to reconstruct the complete row. It is almost always faster to add fields to a row rather than delete fields from a row.
Get Variables	Watch your use of Get Variables	May cause bottlenecks if you use it in a high-volume stream (accepting input). To solve the problem, take the "Get Variables" step out of the transformation (right click, detach) then insert it in with a "Join Rows (cart prod)" step. Make sure to specify the main step from which to read in the "Join Rows" step. Set it to the step that originally provided the "Get Variables" step with data.
Not Applicable	Use new text file input	The new "CSV Input" or "Fixed Input" steps provide optimal performance. If you have a fixed width (field/row) input file, you can even read data in parallel. (multiple copies) These new steps have been rewritten using Non-blocking I/O (NIO) features. Typically, the larger the NIO buffer you

Step	Tip	Description
		specify in the step, the better your read performance will be.
Not applicable	When appropriate, use lazy conversion	In instances in which you are reading data from a text file and you write the data back to a text file, use Lazy conversion to speed up the process. The principle behind lazy conversion that it delays data conversion in hopes that it isn't necessary (reading from a file and writing it back comes to mind). Beyond helping with data conversion, lazy conversion also helps to keep the data in "binary" storage form. This, in turn, helps the internal Kettle engine to perform faster data serialization (sort, clustering, and so on). The Lazy Conversion option is available in the "CSV Input" and "Fixed input" text file reading steps.
Join Rows	Use Join Rows	You need to specify the main step from which to read. This prevents the step from performing any unnecessary spooling to disk. If you are joining with a set of data that can fit into memory, make sure that the cache size (in rows of data) is large enough. This prevents (slow) spooling to disk.
Not Applicable	Review the big picture: database, commit size, row set size and other factors	Consider how the whole environment influences performance. There can be limiting factors in the transformation itself and limiting factors that result from other applications and PDI. Performance depends on your database, your tables, indexes, the JDBC driver, your hardware, speed of the LAN connection to the database, the row size of data and your transformation itself. Test performance using different commit sizes and changing the number of rows in row sets in your transformation settings. Change buffer sizes in your JDBC drivers or database.
Not Applicable	Step Performance Monitoring	New in PDI version 3.1.0, is a way to track the performance of individual steps in a transformation. Step Performance Monitoring is an important tool that allows you identify the slowest step in your transformation.

## Limiting In-Memory Log Output

PDI logs data about transformations and jobs according to default parameters that control how many lines are allowed in the log and how long the oldest line should stay in memory before it is released. Obviously the more lines that are recorded and the longer they are kept, the more heap space is consumed by them. If you are experiencing memory shortages or slow performance in your PDI content, you can address the problem by modifying in-memory logging.

In Spoon, the following parameters control logging:

- **KETTLE\_MAX\_LOG\_SIZE\_IN\_LINES**, which sets the maximum number of log lines that are kept internally by Kettle. Setting this to **0** (the default) forces PDI to keep all rows.
- **KETTLE\_MAX\_LOG\_TIMEOUT\_IN\_MINUTES**, which represents the maximum age (in minutes) that a log line should be kept internally by PDI. Setting this to **0** (the default) keeps all rows indefinitely.

The equivalent parameters, which can be set on each KTR or KJB individually using Kitchen or Pan, are:

- maxloglines
- maxlogtimeout

Set these values to the lowest non-zero values that your operations can tolerate. If you are using logging for any purpose, you must balance between tolerable performance and necessary functionality.



# Pentaho Analysis (Mondrian) Performance Testing and Improvement

---

This section contains advice and procedures for testing and improving Mondrian performance.

## ROLAP Performance-Tuning and Design Theory

---

Concept definition.

## Partitioning High-Cardinality Dimensions

---

<http://devdonkey.blogspot.com/2011/03/mondrian-high-cardinality-dimensions.html>

## Tuple Segment Caching

---

<http://devdonkey.blogspot.com/2011/02/mondrian-spi-segmentcache.html>

# Vertical Resource Scaling

---

If your hardware has reached the limits of its performance capabilities, there are a few different approaches to expanding it without adding new machines. The cheapest solution is to switch from a 32-bit environment to a 64-bit one, which can often be accomplished without buying any new hardware. If you've gone as far as you can go with software and need to scale up, there are some key upgrade points to consider for business intelligence deployments.



**Note:** You may need to adjust your support entitlement contract to complete some of the upgrades described here. The Pentaho BI Server is licensed per CPU, so increasing the number of CPUs in your deployment will require a higher level of service. Contact your Pentaho sales representative for more details on support plan pricing.

## Unfinished Content

---

The topics listed below were originally planned for this guide, but weren't completed before publication. Customer feedback enables Pentaho to give priority to the most important documentation, so if there are sections you would like to see added or completed, let your Pentaho support representative know.

### Appropriate Upgrading

#### Understanding hardware upgrades

Perhaps your server hardware is not upgraded to its maximum potential.

#### More RAM

#### Faster CPUs and multiple cores

#### RAID and solid-state storage

#### Network infrastructure

### Switching to 64-bit


While all modern computing hardware is 64-bit capable, most operating systems currently shipping with it is 32-bit. This is primarily to maintain backwards compatibility with legacy software and proprietary hardware drivers. As time goes on, 64-bit software is becoming more available, and within the next few years it's likely that many operating systems, desktop software packages, and business software suites will be 64-bit only.

The first step in increasing performance on Pentaho servers and workstations is to ensure that all of the operating systems they run on are 64-bit. You can check this in Windows by going to the System Properties dialogue, and in Linux with the **uname -a** command.

# Horizontal Resource Scaling

---

If you've gone as far as you can (or care to) go with vertical scalability, then it's time to consider scaling horizontally. The Pentaho BI Server can scale out to a cluster, or further to a cloud environment. Clusters are excellent for permanently expanding resources commensurate with increasing load; cloud computing is particularly useful if you only need to scale out for specific periods of increased activity.

 **Note:** You may need to adjust your support entitlement contract to complete some of the upgrades described here. Contact your Pentaho sales representative for more details on support plan pricing.


## Clustering the Application Server

---

The Pentaho BI Server is scalable out to the limits of its Java application server. Most application servers are easily clustered and load balanced by using the Apache httpd Web server with the Tomcat Connector (**mod\_jk**) plugin.

The Tomcat Connector module forwards requests from httpd to Tomcat via the AJP protocol. It can operate locally through localhost or remotely through TCP. Because JBoss embeds Tomcat, the core methodology for clustering between the two application servers is the same:

1. Establish a load balancer that runs Apache httpd with the Tomcat Connector (mod\_jk) plugin
2. Copy the pentaho-solutions directory to a network drive; this will be shared among all nodes
3. Configure the application server nodes with identical configurations and Pentaho BI Server deployments

 **Note:** You may have to modify these instructions to accommodate your operating system's unique httpd, Apache module, and application server configuration. Many operating systems -- especially Linux distributions -- have non-standard ways of managing these services.

## Clustering Requirements

In order to successfully implement a Pentaho deployment on a Tomcat or JBoss cluster, you must meet the following requirements:

- **Each node and the load balancer must be time-synchronized via NTP.** All machines that comprise the cluster have to have the same system time. If they do not, it will be impossible to share sessions among nodes.
- **You must run only one node per machine (or NIC).** It is possible to run multiple application servers on each node with a modified configuration, but this scenario does not offer any benefit for load balancing (performance) or hardware failover (redundancy), and therefore not covered in this guide. Refer to your application server's clustering documentation for more information.
- **You must use either Tomcat version 6.0 or JBoss version 5.1.** You may be able to use this guide as a basic blueprint for configuring other application servers or versions of Tomcat and JBoss for a clustered environment, but Pentaho support will not be able to assist you if you run into any problems.
- **You must have permission to install software and modify service configurations.** Or you must have access to someone at your company who does have the correct permission levels (root access, typically).
- **Only the Pentaho BI Server will be deployed to the cluster.** It is possible to modify the configuration to deploy other WARs or EARs. However, for ease of testing and support, Pentaho only supports deployment of the **pentaho** and **pentaho-style** WARs to the cluster.
- **You must use a single repository location.** Most people use a database-based solution repository; remember that you are not clustering the database server in this procedure -- only the application server. If you are using a file-based repository, you will have to create one canonical location for the solution repository, preferably on a network share so that the location can be the same for all nodes.

## Sharing the Solution Repository

In order to keep all cluster nodes current with stored Pentaho content and settings, they must share the pentaho-solutions directory. This directory is mirrored in the Pentaho solution database, which is also shared among all nodes. Since the database configuration is partially stored in this same directory, this is a required step for sharing the BI Server configuration among all nodes as well. Follow the directions below to set up a common pentaho-solutions directory.

1. Select a secure location in your computing infrastructure for the pentaho-solutions directory, and configure it so that it can securely share directories over your corporate network.

2. If you have existing content and BI Server settings that you want to move to the cluster, copy it over to the shared drive now. You can also copy an extant pentaho-solutions directory from a standalone BI Server that you are moving to a clustered environment.
3. Configure each cluster node to automatically connect to the machine that will serve the pentaho-solutions directory. This is best accomplished by setting up a network drive that connects to the remote machine.
4. Modify each BI Server node configuration to point to the new location of the pentaho-solutions directory by editing the `/WEB-INF/web.xml` inside of the deployed **pentaho.war** or unpacked `/webapps/pentaho/` directory and modifying the **solution-path** element:

```
<context-param>
  <param-name>solution-path</param-name>
  <param-value>/mnt/sharedhost/pentaho-solutions/</param-value>
</context-param>
```

5. If you have been using a local solution database (on the same machine with a standalone BI Server that you are migrating to a cluster), you must either move the database to a machine that can be reliably shared among all nodes, or ensure that it is available to the other nodes. If you move the database, you must change the **hibernate** and **quartz** locations in the following files within the shared pentaho-solutions directory:
  - `/system/applicationContext-spring-security-hibernate.properties`
  - `/system/hibernate/hibernate-settings.xml`
  - One of these, depending on which database you are using:
    - `/system/hibernate/mysql.hibernate.cfg.xml`
    - `/system/hibernate/postgresql.hibernate.cfg.xml`
    - `/system/hibernate/oracle.hibernate.cfg.xml`

And **/META-INF/context.xml** inside the pentaho.war or unpacked pentaho directory on each Tomcat or JBoss node.

You now have a pentaho-solutions directory and solution database shared among all of your cluster nodes.

## Installing and Configuring Apache as a Load Balancer

Clustering requires a load balancer to manage each application server node. This is most easily accomplished via the Apache httpd Web server, which will connect to each application server node via the Tomcat Connector module.

1. Install Apache 2.2.x -- with SSL support -- through your operating system's preferred installation method. For most people, this will be through a package manager. It's also perfectly valid to download and install the reference implementation from <http://www.apache.org>.

It is possible to use Apache 1.3, but you will have to modify the instructions on your own from this point onward.

2. If it has started as a consequence of installing, stop the Apache server or service.
3. Retrieve or create your SSL keys.

If you do not know how to generate self-signed certificates, refer to the OpenSSL documentation. Most production environments have SSL certificates issued by a certificate authority such as Thawte or Verisign.

4. Check to see if you already have the Tomcat Connector installed on your system. You can generally accomplish this by searching your filesystem for **mod\_jk**, though you can also search your **httpd.conf** file for **mod\_jk**. If it is present, then you only need to be concerned with the Apache httpd configuration details and can skip this step. If it is not there, then the Tomcat Connector module needs to be installed. If you are using Linux or BSD, use your package manager or the Ports system to install **mod\_jk**. For all other platforms, visit the <http://www.apache.org/dist/tomcat/tomcat-connectors/jk/binaries/>, then click on the directory for your operating system. The module will be either an **.so** (for Linux, BSD, OS X, and Solaris) or **.dll** (for Windows) file. Save it to your Apache modules directory, which is generally `C:\Program Files\Apache Group\Apache2\modules\` on Windows, and `/usr/lib/apache2/modules/` on Unix-like operating systems, though this can vary depending on your Apache configuration.
5. Edit your **httpd.conf** file with a text editor and add the following text to the end of the file, modifying the paths and filenames as instructed in the comments:



**Note:** Some operating systems use modular httpd configuration files and have unique methods of including each separate piece into one canonical file. Ensure that you are not accidentally interfering with an auto-generated `mod_jk` configuration before you continue. In many cases, some of the configuration example below will have to be cut out (such as the **LoadModule** statement). In some cases (such as with Ubuntu

Linux), httpd.conf may be completely empty, in which case you should still be able to add the below lines to it. Replace **example.com** with your hostname or domain name.

```
# Load mod_jk module
# Update this path to match your mod_jk location; Windows users should change the .so
to .dll
LoadModule      jk_module      /usr/lib/apache/modules/mod_jk.so
# Where to find workers.properties
# Update this path to match your conf directory location
JkWorkersFile  /etc/httpd/conf/workers.properties
# Should mod_jk send SSL information to Tomcat (default is On)
JkExtractSSL  On
# What is the indicator for SSL (default is HTTPS)
JkHTTPSIndicator  HTTPS
# What is the indicator for SSL session (default is SSL_SESSION_ID)
JkSESSIONIndicator  SSL_SESSION_ID
# What is the indicator for client SSL cipher suit (default is SSL_CIPHER)
JkCIPHERIndicator  SSL_CIPHER
# What is the indicator for the client SSL certificated (default is SSL_CLIENT_CERT)
JkCERTSIndicator  SSL_CLIENT_CERT
# Where to put jk shared memory
# Update this path to match your local state directory or logs directory
JkShmFile      /var/log/httpd/mod_jk.shm
# Where to put jk logs
# Update this path to match your logs directory location (put mod_jk.log next to
access_log)
JkLogFile      /var/log/httpd/mod_jk.log
# Set the jk log level [debug/error/info]
JkLogLevel     info
# Select the timestamp log format
JkLogStampFormat "[%a %b %d %H:%M:%S %Y] "
# Send everything for context /examples to worker named worker1 (ajp13)
# JkOptions indicates to send SSK KEY SIZE
JkOptions      +ForwardKeySize +ForwardURICompat -ForwardDirectories
# JkRequestLogFormat
JkRequestLogFormat "%w %V %T"
# Mount your applications on the load balancer node
JkMount /pentaho/* loadbalancer
# There should be no need to cluster the style WAR, but just in case...
#JkMount /pentaho-style/* loadbalancer
# Add shared memory.
# This directive is present with 1.2.10 and
# later versions of mod_jk, and is needed for
# for load balancing to work properly
JkShmFile logs/jk.shm
# Add jkstatus for managing runtime data
<Location /jkstatus/>
JkMount status
Order deny,allow
Deny from all
Allow from 127.0.0.1
</Location>
<VirtualHost example.com
    ServerName example.com
    JkMount /pentaho default
    JkMount /pentaho/* default
    JkMount /sw-style default
    JkMount /sw-style/* default
    JkMount /pentaho-style default
    JkMount /pentaho-style/* default
</VirtualHost>
```

6. In your Apache configuration, ensure that SSL is enabled by uncommenting or adding and modifying the following lines:

```
LoadModule ssl_module modules/mod_ssl.so
Include conf/extra/httpd-ssl.conf
```

7. Save and close the file, then edit `/conf/extra/httpd-ssl.conf` and properly define the locations for your SSL certificate and key:

```
SSLCertificateFile "conf/ssl/mycert.cert"  
SSLCertificateKeyFile "conf/ssl/mycert.key"
```

8. Ensure that your SSL engine options contain these entries:

```
SSLOptions +StdEnvVars +ExportCertData
```

9. Add these lines to the end of the **VirtualHost** section:

```
JkMount /pentaho default  
JkMount /pentaho/* default  
JkMount /sw-style default  
JkMount /sw-style/* default  
JkMount /pentaho-style default  
JkMount /pentaho-style/* default
```

10. Save and close the file, then create a **workers.properties** file in your Apache conf directory. If it already exists, merge it with the example configuration in the next step.
11. Copy the following text into the new **workers.properties** file, changing the location of Tomcat and Java, and the port numbers and IP addresses to match your configuration:

 **Note:** Remove the **workers.tomcat\_home** setting if you are using JBoss.

```
# Load-balancer settings  
workers.tomcat_home=/home/pentaho/pentaho/server/biserver-ee/tomcat/  
workers.java_home=/home/pentaho/pentaho/java/  
# Define list of workers that will be used for mapping requests  
worker.list=loadbalancer,status  
# Define Node1  
# modify the host as your host IP or DNS name.  
worker.node1.port=8009  
worker.node1.host=192.168.3.6  
worker.node1.type=ajp13  
worker.node1.lbfactor=1  
worker.node1.cachesize=50  
# Define Node2  
# modify the host as your host IP or DNS name.  
worker.node2.port=8009  
worker.node2.host=192.168.3.7  
worker.node2.type=ajp13  
worker.node2.lbfactor=1  
worker.node2.cachesize=50  
# Load-balancing behaviour  
worker.loadbalancer.type=lb  
worker.loadbalancer.balance_workers=node1,node2  
worker.loadbalancer.sticky_session=1  
# Status worker for managing load balancer  
worker.status.type=status
```

Apache httpd is now configured to act as a load balancer for two nodes. You can come back and adjust this for more nodes later, but it is easier to test and adjust your configuration at this minimal level before you expand out further.

Proceed to the JBoss or Tomcat section below, depending on which application server you are using.

## Tomcat Configuration

Before continuing, you should have a working BI Server instance running in Tomcat 6. This should not be currently in production. The safest way to proceed -- and the quickest way to recover from a failed deployment -- is to make a backup archive of the **tomcat** directory before making any changes.

Follow the directions below to modify your Tomcat server to act as a member of a cluster.

1. Stop Tomcat on each cluster node.
2. Edit the `/tomcat/conf/server.xml` file on each node and add the **jvmRoute** parameter (change the value to match the node names you defined in **workers.properties**) to the **Engine** element:

```
<Engine name="Catalina" defaultHost="localhost" jvmRoute="node01">
```

3. Further down in the file, ensure that the AJP connector line is uncommented, and that the port number matches the node definition you defined in **workers.properties** (if it doesn't match, change the node entry in workers.properties, not the AJP connector entry):

```
<Connector URIEncoding="UTF-8" port="8009" enableLookups="false" redirectPort="8443"
protocol="AJP/1.3" />
```

4. Also ensure that the SSL version of the AJP connector line is uncommented, and add two properties at the end of the line that define your SSL keystore password:

```
<Connector URIEncoding="UTF-8" port="8443" maxHttpHeaderSize="8192"
maxThreads="150" minSpareThreads="25" maxSpareThreads="75" enableLookups="false"
disableUploadTimeout="true" acceptCount="100" scheme="https" secure="true"
clientAuth="false" sslProtocol="TLS" keystorePass="password" />
```

5. Further down in server.xml, uncomment the **Cluster** node, but do not make any changes to it:

```
<Cluster className="org.apache.catalina.cluster.tcp.SimpleTcpCluster"
managerClassName="org.apache.catalina.cluster.session.DeltaManager"
expireSessionsOnShutdown="false"
useDirtyFlag="true"
notifyListenersOnReplication="true">

  <Membership
    className="org.apache.catalina.cluster.mcast.McastService"
    mcastAddr="228.0.0.4"
    mcastPort="45564"
    mcastFrequency="500"
    mcastDropTime="3000" />

  <Receiver
    className="org.apache.catalina.cluster.tcp.ReplicationListener"
    tcpListenAddress="auto"
    tcpListenPort="4001"
    tcpSelectorTimeout="100"
    tcpThreadCount="6" />

  <Sender
    className="org.apache.catalina.cluster.tcp.ReplicationTransmitter"
    replicationMode="pooled"
    ackTimeout="15000"
    waitForAck="true" />

  <Valve className="org.apache.catalina.cluster.tcp.ReplicationValve"
    filter=".*\.(gif|.*\.(js|.*\.(jpg|.*\.(png|.*\.(htm|.*\.(html|.*\.(css|.*
\.txt); />

  <Deployer className="org.apache.catalina.cluster.deploy.FarmWarDeployer"
    tempDir="/tmp/war-temp/"
    deployDir="/tmp/war-deploy/"
    watchDir="/tmp/war-listen/"
    watchEnabled="false" />

  <ClusterListener
    className="org.apache.catalina.cluster.session.ClusterSessionListener" />
</Cluster>
```

6. Edit the /tomcat/conf/context.xml and add a **distributable** parameter to the main **Context** element:

```
<Context distributable="true">
```

## Copying WAR Files to the Nodes

Now that you have configured your application for clustering, you have to copy the Pentaho WARs to each node. Follow the instructions below.

1. If you have not already done so, install an identical version of Tomcat or JBoss (including the configuration files) to each node in the cluster. Alternatively you can copy the Tomcat or JBoss directory you've been working with to each node and configure the operating system to start and stop these services when the computer boots and shuts down.
2. If they were not copied over in the previous step, ensure that the Pentaho WARs have been copied to the /server/default/deploy/ or /webapps/ directory: **pentaho.war**, **pentaho-style.war**.

You now have a populated cluster.

## Starting and Testing the Cluster

Follow the below instructions to start the cluster and verify that it is working properly.

1. Start the solution database.
2. Ensure that the machine hosting the shared pentaho-solutions directory is available to each node.
3. Start the application server on each node.
4. Start Apache and your application server on the load balancer.
5. Test the cluster by accessing the BI Server through the load balancer's IP address, hostname, or domain name and commence whatever test procedure you have designed for this scenario.



# Changing the Java VM Memory Limits

---

If you are running out of memory even though your server has a lot of RAM, you probably need to increase the resources allocated to the JRE instance that runs the Pentaho software you're trying to improve performance on. Adjusting the memory limit is an easy configuration change to make, but the instructions differ depending on the client tool or Web application server you're using. Refer only to the sections below that apply to your situation.

## Increasing the Memory Limit in JBoss

---

By default, JBoss has a relatively low memory allotment. This can cause out of memory errors in the BI Server from time to time. To increase the memory limit, follow the below process.

1. Stop the JBoss server or service.
2. Open a terminal or command prompt window.
3. If you are running Tomcat as a service in Windows, paste in this command, changing the path to the BI Suite to accommodate your situation: `C:\pentaho\bi-suite-2.0.1\server\bi-server\tomcat\bin\tomcat5 //US//pentahobiserver --Jvms=512 --Jvmmx=1024`, and in Linux, the command would be: `/bi-suite-2.0.1/server/bi-server/tomcat/bin/tomcat5 //US//pentahobiserver --Jvms=512 --Jvmmx=1024`
4. If you are starting Tomcat manually as a server with the BI Platform's included scripts, edit the **start-pentaho** script (.bat for Windows, and .sh for Linux), and modify the **CATALINA\_OPTS** environment variable, modifying the values of **Xms** and **Xmx** in the same manner as the previous step.

```
export CATALINA_OPTS="-Xms512m -Xmx1024m -XX:MaxPermSize=256m -Dsun.rmi.dgc.client.gcInterval=3600000 -Dsun.rmi.dgc.server.gcInterval=3600000"
```

5. Start the Tomcat server or service.

Your Tomcat server now has increased minimum and maximum memory limits. You can adjust the **Jvmmx** number (this parameter specifies the maximum limit) to a higher number if you prefer. However, if the Java virtual machine refuses to start with increased limits, then you will have to add more RAM to your system, stop some memory-intensive services, or reduce the maximum memory limit to a lower number. This problem occurs when there is not enough contiguous memory available to assign to the JVM, and appears to happen on Windows at lower thresholds than on other operating systems.

## Increasing the Memory Limit in Tomcat

---

By default, Tomcat has a relatively low memory allotment. This can cause out of memory errors in the BI Server from time to time. To increase the memory limit, follow the below process.

1. Stop the Tomcat server or service.
2. **If you are modifying your own Tomcat instance** and performed a manual deployment of the BI Server WAR, edit the `~/ .bashrc` for the user account that starts the Tomcat service, or whatever configuration file or dialogue contains global system variables on your BI Server machine. Set or modify the **CATALINA\_OPTS** system variable to include reasonable minimum and maximum memory settings via the **-Xms** and **-Xmx** options.

```
export CATALINA_OPTS="-Xms512m -Xmx1024m"
```

3. **If you are using a Pentaho-supplied Tomcat instance** provided via BI Server archive packages or the BI Suite graphical installer, edit the **start-pentaho** script (.bat for Windows, and .sh for Linux), and modify the **CATALINA\_OPTS** environment variable, adjusting the values of **Xms** and **Xmx** in the same manner as the previous step.

```
export CATALINA_OPTS="-Xms512m -Xmx1024m -XX:MaxPermSize=256m -Dsun.rmi.dgc.client.gcInterval=3600000 -Dsun.rmi.dgc.server.gcInterval=3600000"
```

4. Start the Tomcat server or service.

Your Tomcat server now has increased minimum and maximum memory limits. You can adjust the **Jvmmx** number (this parameter specifies the maximum limit) to a higher number if you prefer. However, if the Java virtual machine refuses to start with increased limits, then you will have to add more RAM to your system, stop some memory-intensive services, or reduce the maximum memory limit to a lower number. This problem occurs when there is not enough

contiguous memory available to assign to the JVM, and appears to happen on Windows at lower thresholds than on other operating systems.

# Testing BI Server Scalability

---

Improper scalability testing can give you the wrong idea about changes you've made to your BI Server instance.

Before testing, ensure that you're reusing sessions, instead of creating successive new sessions. Creating multiple unnecessary sessions causes the BI Server to run out of memory unless the session timeout in `web.xml` is set extremely low (1 per minute, for instance); the default is 30 minutes.

Logging into the BI Server is resource-intensive: you must authenticate the user; create a bunch of session data; and run all startup action sequences, which usually store data in the user's session. So if, during testing, you simply string together a bunch of URLs and ignore the established session, you'll create a series of 30-minute sessions and almost certainly run out of memory. The correct way to test the server is to mimic a user's actions from a browser. The sections below explain how to do this.

## Sessions and URLs

Most stress test tools (Loadrunner, JMeter, etc.) have session/cookie management options to ensure that they behave like a human user. However, if you're creating your own test scripts, you should follow this process:

1. Log into the server
2. Execute a URL that contains the **userid** and **password** parameters (**&userid=joe&password=password** for example)
3. Using the same session, submit other URLs that don't have the `userid/password` in them.

Use this process for as many users as you need to test with.

To log out of a session, you can use the **`http://localhost:8080/pentaho/Logout`** URL; this will invalidate the session if you append the `userid` and `password` values of the logged-in user. Without passing those parameters (or, alternatively, specifying the session ID or cookie) on the Logout URL, you will create another new session instead of closing an old one.

This means that two back-to-back **wget** commands in Linux will create two different HTTP sessions on the server unless one of the following conditions is met:

1. **-cookies=on** is specified for both `wget` commands
2. **-save-cookies** is used on the first `wget` command to save the cookies to a file, and **-load-cookies** is used on the second `wget` command to load the session state

## Memory and sessions

Out of memory errors in the BI Platform can happen because of what your test script is doing, not necessarily because of any weakness in the platform. You can see just how robust the BI Platform is by taking a look at a production server's natural (human user) load. The following URL will show you what each day's maximum and present number of HTTP sessions are: **`http://testserver.example.com/pentaho/public/UserService`**.

You can see the Java virtual machine memory settings by examining the the options passed to the Tomcat or JBoss start scripts, or by looking at the **CATALINA\_OPTS** system variable, if there is one. The **Xms** and **Xmx** options define the minimum and maximum amount of memory assigned to the application server. The default settings are not particularly high, and even if you've adjusted them, take note of the number of sessions it takes to use up all of the memory. Also take note of the fact that closing sessions after an out of memory error will return the memory to the available pool, proving that there are no memory leaks or zombie sessions inherent in the BI Platform.