

Getting Started Guide

Sun™ ONE Application Server

Version 7, Enterprise Edition

817-2147-10
September 2003

Sun Microsystems, Inc.
4150 Network Circle
Santa Clara, CA 95054 U.S.A.

Copyright 2003 Sun Microsystems, Inc. All rights reserved.

Sun, Sun Microsystems, the Sun logo, Java, Solaris, Sun ONE, iPlanet, and all Sun, Java, and Sun ONE based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries.

UNIX is a registered trademark in the United States and other countries, exclusively licensed through X/Open Company, Ltd.

Netscape is a trademark or registered trademark of Netscape Communications Corporation in the United States and other countries.

Federal Acquisitions: Commercial Software—Government Users Subject to Standard License Terms and Conditions

The product described in this document is distributed under licenses restricting its use, copying, distribution, and decompilation. No part of the product or this document may be reproduced in any form by any means without prior written authorization of Sun Microsystems, Inc. and its licensors, if any.

THIS DOCUMENTATION IS PROVIDED “AS IS” AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

Copyright 2003 Sun Microsystems, Inc. Tous droits réservés.

Sun, Sun Microsystems, le logo Sun, Java, Solaris, Sun ONE, et iPlanet sont des marques de fabrique ou des marques déposées de Sun Microsystems, Inc. aux Etats-Unis et d'autre pays.

UNIX est une marque enregistrée aux Etats-Unis et dans d'autres pays et licenciée exclusivement par X/Open Company Ltd.

Netscape est une marque de Netscape Communications Corporation aux Etats-Unis et dans d'autres pays.

Le produit décrit dans ce document est distribué selon des conditions de licence qui en restreignent l'utilisation, la copie, la distribution et la décompilation. Aucune partie de ce produit ni de ce document ne peut être reproduite sous quelque forme ou par quelque moyen que ce soit sans l'autorisation écrite préalable de Sun Microsystems, Inc. et, le cas échéant, de ses bailleurs de licence.

CETTE DOCUMENTATION EST FOURNIE “EN L'ÉTAT”, ET TOUTES CONDITIONS EXPRESSES OU IMPLICITES, TOUTES REPRÉSENTATIONS ET TOUTES GARANTIES, Y COMPRIS TOUTE GARANTIE IMPLICITE D'APTITUDE À LA VENTE, OU À UN BUT PARTICULIER OU DE NON CONTREFAÇON SONT EXCLUES, EXCEPTÉ DANS LA MESURE OÙ DE TELLES EXCLUSIONS SERAIENT CONTRAIRES À LA LOI.

Contents

About this Guide	7
What's In This Guide	7
How This Guide is Organized	8
Using the Documentation	8
Documentation Conventions	11
General Conventions	11
Conventions Referring to Directories	11
Product Support	12
Chapter 1 Introduction to Sun ONE Application Server 7, Enterprise Edition	13
About Sun ONE Application Server	13
About Enterprise Edition Features	16
Clustering	16
Load Balancing	17
Session Persistence	18
About Session Persistence Types	18
About Session Persistence Configuration	19
About Single Sign-on Session Information	19
High-Availability Database	19
Example Clustering Scenario	20
Tools for Configuring and Administering the Application Server	22
Chapter 2 Enterprise Features Configuration Tutorial	25
Preparing to Use the Tutorials	25
Installation Requirements	26
Required Steps Before Using the Tutorials	26
Overview of Tutorial Steps	27

Starting the Server	29
Setting the PATH Variable	29
Running asadmin start-domain	30
Verifying Server Start-up	31
Verifying Admin Server Start-up	31
Accessing the Administration Interface	31
Viewing the Admin Server's Event Log	33
Verifying Start-up of Server Instances	34
Verifying Instances using asadmin	34
Verifying Instances by Accessing the HTTP Server	35
Creating the loadbalancer.xml File	37
Adding a Cluster to the loadbalancer.xml File	38
Configuring Load Balancing	40
Configuring the Health Checker	41
Enabling Load Balancer Monitoring	42
Other Load Balancer Properties	43
Dynamic Reconfiguration using the Reload Poll Interval	44
Response Timeout	44
HTTPS Routing	44
Sample loadbalancer.xml File	45
 Chapter 3 Cluster JSP Sample Application Tutorial	47
Preparing to use the Cluster JSP Sample Application Tutorial	47
Deploying the Cluster JSP Sample Application to a Cluster	48
Input Files for the cladadmin Command	49
The cladadmin Command Syntax	50
Running cladadmin deploy	50
The asadmin Commands Supported by the cladadmin Command	51
Requirements and Limitations	52
Starting the Application Server Instances Using cladadmin	52
Verifying the Application Deployment	53
Monitoring the Sample in the Application Server	55
Using the Administration Interface to View Logs	55
Viewing Logs Using the tail Command	57
Application-Generated Messages in the Event Log	58
Application-Generated Messages in the Access Log	58
Troubleshooting Deployment to an Application Server Instance	59
Adding the Sample Application to the Cluster	59
Applying Configuration Changes and Restarting the Web Server	61
Running the Application	62
Verifying Load Balancing	67
Steps for Verifying Load Balancing	67
Troubleshooting the Load-Balancer Plug-in	68

Finding Errors in Your loadbalancer.xml File	69
Using the Health Checker	70
Verifying HTTP Session Persistence	71
Quiescing a Server Instance	72
 Chapter 4 Summary and Next Steps	 75
 Index	 77

About this Guide

This preface describes the contents of Sun™ Open Net Environment (Sun ONE) Application Server 7 *Getting Started Guide*.

This preface contains the following sections:

- [What's In This Guide](#)
- [How This Guide is Organized](#)
- [Using the Documentation](#)
- [Documentation Conventions](#)
- [Product Support](#)

What's In This Guide

This *Getting Started Guide* is intended for first-time users of the Sun ONE Application Server. It offers a brief, hands-on means of gaining familiarity with Sun ONE Application Server, Enterprise Edition features, with an emphasis on demonstrating the load balancing and HTTP session persistence features. Prior application server and development experience are not prerequisites for the exercises in this guide, though you must install the server using the *Sun ONE Application Server Installation Guide* before starting.

The guide first introduces you to the Sun ONE Application Server, including a brief overview of the clustering, load balancing, and HTTP session persistence features. Next, it steps you through the configuration of an environment for running the sample application in the exercises. It shows you how to deploy a sample application to demonstrate and verify load balancing and session persistence. It contains a list of further areas of investigation, and references to where to find more information.

How This Guide is Organized

The following chapters are included in this guide:

- [Chapter 1, “Introduction to Sun ONE Application Server 7, Enterprise Edition”](#) provides an overview of the product, including a brief introduction to clustering, load balancing, HTTP session persistence, and the high-availability database.
- [Chapter 2, “Enterprise Features Configuration Tutorial”](#) contains a list of task to do before starting the tutorials, a summary of steps to configure your server to use Enterprise features, and a tutorial on configuring your server to use clusters and HTTP session failover.
- [Chapter 3, “Cluster JSP Sample Application Tutorial”](#) describes the Cluster JSP application, how to deploy it to a cluster, and how to run it. It describes how you can use this application to verify load balancing and HTTP session persistence.
- [Chapter 4, “Summary and Next Steps”](#) describes the tasks you’ve completed in this guide and the next steps to take in working with the product.

Using the Documentation

The Sun ONE Application Server manuals are available as online files in Portable Document Format (PDF) and Hypertext Markup Language (HTML) formats at:

<http://docs.sun.com/>

The following table lists tasks and concepts described in the Sun ONE Application Server manuals. The left column lists the tasks and concepts, and the right column lists the corresponding manuals.

Table 1 Sun ONE Application Server Documentation Roadmap

For information about	See the following
Late-breaking information about the software and the documentation.	<i>Release Notes</i>
Comprehensive, table-based summary of supported hardware, operating system, JDK, and JDBC/RDBMS.	<i>Platform Summary</i>
Sun ONE Application Server 7 overview, including the features available with each product edition.	<i>Product Overview</i>
Diagrams and descriptions of server architecture and the benefits of the Sun ONE Application Server architectural approach.	<i>Server Architecture</i>

Table 1 Sun ONE Application Server Documentation Roadmap

For information about	See the following
New enterprise, developer, and operational features of Sun ONE Application Server 7.	<i>What's New</i>
How to get started with the Sun ONE Application Server 7 product. Includes a sample application tutorial.	<i>Getting Started Guide</i>
Installing the Sun ONE Application Server software and its components, such as sample applications, the Administration interface, and the high-availability components. Instructions for implementing a basic high-availability configuration are included.	<i>Installation Guide</i>
Evaluating your system needs and enterprise to ensure that you deploy Sun ONE Application Server in a manner that best suits your site. General issues and concerns that you must be aware of when deploying an application server are also discussed.	<i>System Deployment Guide</i>
Best practices for HTTP session availability that application architects and developers can use.	<i>Application Design Guidelines for Storing Session State</i>
Creating and implementing Java™ 2 Platform, Enterprise Edition (J2EE™ platform) applications intended to run on the Sun ONE Application Server 7 that follow the open Java standards model for J2EE components such as servlets, Enterprise JavaBeans™ (EJBs™), and JavaServer Pages™ (JSPs™). Includes general information about application design, developer tools, security, assembly, deployment, debugging, and creating lifecycle modules. A comprehensive Sun ONE Application Server glossary is included.	<i>Developer's Guide</i>
Creating and implementing J2EE web applications that follow the Java™ Servlet and JavaServer Pages (JSP) specifications on the Sun ONE Application Server 7. Discusses web application programming concepts and tasks, and provides sample code, implementation tips, and reference material. Topics include results caching, JSP precompilation, session management, security, deployment, SHTML, and CGI.	<i>Developer's Guide to Web Applications</i>
Creating and implementing J2EE applications that follow the open Java standards model for enterprise beans on the Sun ONE Application Server 7. Discusses Enterprise JavaBeans (EJB) programming concepts and tasks, and provides sample code, implementation tips, and reference material. Topics include container-managed persistence, read-only beans, and the XML and DTD files associated with enterprise beans.	<i>Developer's Guide to Enterprise JavaBeans Technology</i>
Creating Application Client Container (ACC) clients that access J2EE applications on the Sun ONE Application Server 7.	<i>Developer's Guide to Clients</i>
Creating web services in the Sun ONE Application Server environment.	<i>Developer's Guide to Web Services</i>

Table 1 Sun ONE Application Server Documentation Roadmap

For information about	See the following
Java™ Database Connectivity (JDBC™), transaction, Java Naming and Directory Interface™ (JNDI), Java™ Message Service (JMS), and JavaMail™ APIs.	<i>Developer's Guide to J2EE Services and APIs</i>
Creating custom NSAPI plug-ins.	<i>Developer's Guide to NSAPI</i>
Information and instructions on the configuration, management, and deployment of the Sun ONE Application Server subsystems and components, from both the Administration interface and the command-line interface. Topics include cluster management, the high-availability database, load balancing, and session persistence. A comprehensive Sun ONE Application Server glossary is included.	<i>Administrator's Guide</i>
Editing Sun ONE Application Server configuration files, such as the <code>server.xml</code> file.	<i>Administrator's Configuration File Reference</i>
Configuring and administering security for the Sun ONE Application Server operational environment. Includes information on general security, certificates, and SSL/TLS encryption. HTTP server-based security is also addressed.	<i>Administrator's Guide to Security</i>
Configuring and administering service provider implementation for J2EE™ Connector Architecture (CA) connectors for the Sun ONE Application Server 7. Topics include the Administration Tool, Pooling Monitor, deploying a JCA connector, and sample connectors and sample applications.	<i>J2EE CA Service Provider Implementation Administrator's Guide</i>
Migrating your applications to the new Sun ONE Application Server 7 programming model, specifically from iPlanet Application Server 6.x and from Netscape Application Server 4.0. Includes a sample migration.	<i>Migrating and Redeploying Server Applications Guide</i>
How and why to tune your Sun ONE Application Server to improve performance.	<i>Performance Tuning Guide</i>
Information on solving Sun ONE Application Server problems.	<i>Troubleshooting Guide</i>
Messages that you may encounter while running Sun ONE Application Server 7. Includes a description of the likely cause and guidelines on how to address the condition that caused the message to be generated.	<i>Error Message Reference</i>
Utility commands available with the Sun ONE Application Server; written in manpage style.	<i>Utility Reference Manual</i>
Using the Sun™ Open Net Environment (Sun ONE) Message Queue software.	The Sun ONE Message Queue documentation at: http://docs.sun.com/db?p=prod/sl.slmsgqu

Documentation Conventions

This section describes the types of conventions used throughout this guide:

- [General Conventions](#)
- [Conventions Referring to Directories](#)

General Conventions

The following general conventions are used in this guide:

- **File and directory paths** are given in UNIX[®] format (with forward slashes separating directory names).
- **URLs** are given in the format:

`http://server.domain/path/file.html`

In these URLs, *server* is the server name where applications are run; *domain* is your Internet domain name; *path* is the server's directory structure; and *file* is an individual filename. Italic items in URLs are placeholders.

- **Font conventions** include:
 - The `monospace` font is used for sample code and code listings, API and language elements (such as function names and class names), file names, path names, directory names, and HTML tags.
 - *Italic* type is used for book titles, emphasis, variables and placeholders, and words used in the literal sense.
 - **Bold** type is used as either a paragraph lead-in or to indicate words used in the literal sense.

Conventions Referring to Directories

By default, the Sun ONE Application Server files are spread across several root directories. These directories are described in this section.

- *install_dir* refers to the directory which contains the static portion of the installation image. All utilities, executables, and libraries that make up the Sun ONE Application Server reside in this location. The default location is `/opt/SUNWappserver7`.

- *default_config_dir* refers to the directory which is the default location for any domains that are created. The default location is `/var/opt/SUNWappserver7/domains`.
- *install_config_dir* refers to the directory which contains installation-wide configuration information such as licenses and the master list of administrative domains configured for this installation. The default location is `/etc/opt/SUNWappserver7`.

Product Support

If you have problems with your system, contact customer support using one of the following mechanisms:

- The online support web site at:
<http://www.sun.com/supporttraining/>
- The telephone dispatch number associated with your maintenance contract

Please have the following information available prior to contacting support. This helps to ensure that our support staff can best assist you in resolving problems:

- Description of the problem, including the situation where the problem occurs and its impact on your operation
- Machine type, operating system version, and product version, including any patches and other software that might be affecting the problem
- Detailed steps on the methods you have used to reproduce the problem
- Any error logs or core dumps

Introduction to Sun ONE Application Server 7, Enterprise Edition

The following topics are included in this chapter:

- [About Sun ONE Application Server](#)
- [About Enterprise Edition Features](#)
- [Tools for Configuring and Administering the Application Server](#)

About Sun ONE Application Server

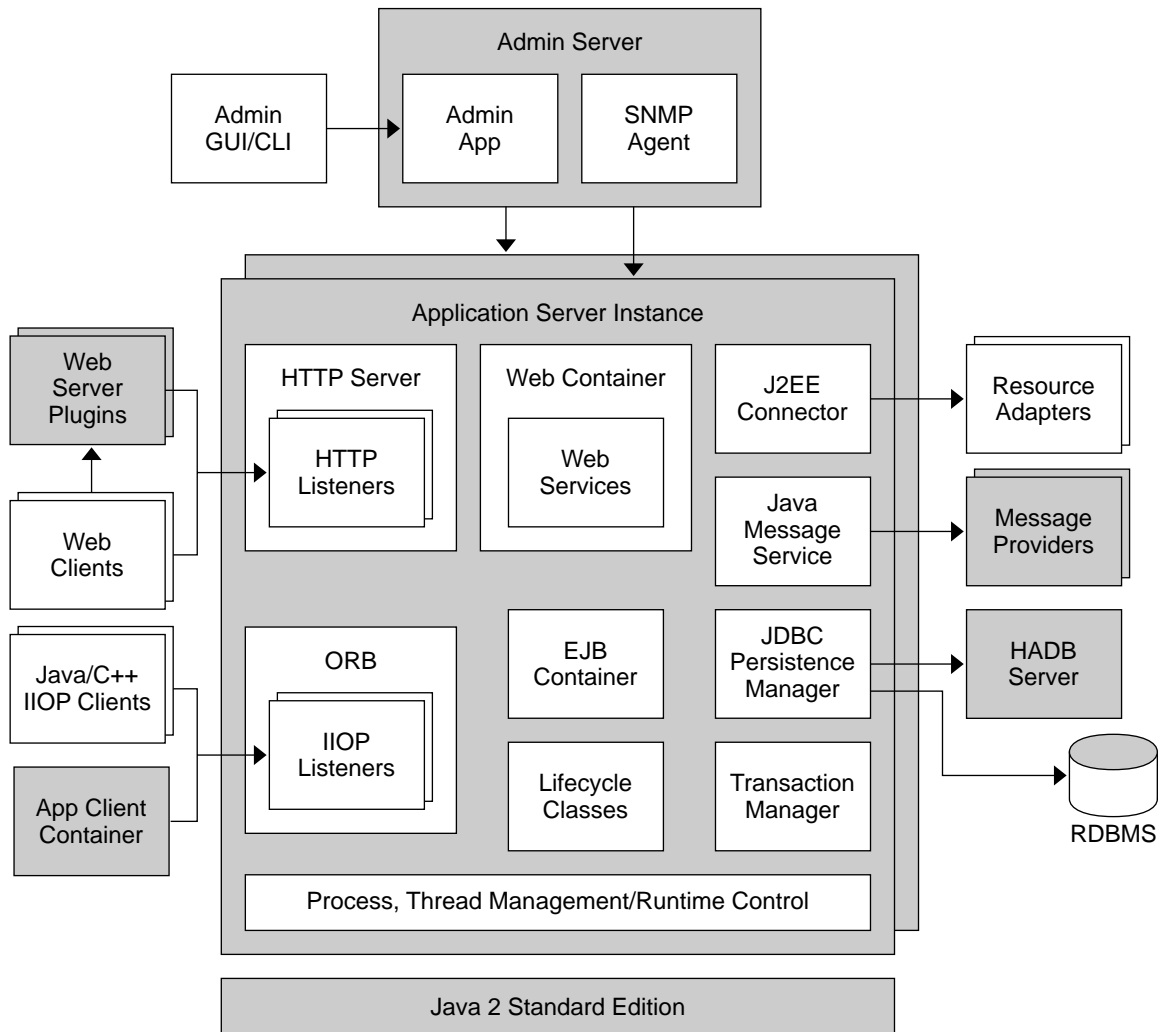
The Sun™ Open Net Environment (ONE) Application Server 7, Enterprise Edition provides a high-performance Java™ 2 Platform, Enterprise Edition (J2EE™) platform suitable for broad deployment of application services and web services. It offers a new modular architecture based on industry standard components, including proven implementations of the HTTP server infrastructure, Java Message Service (JMS), and rigorous support of the latest J2EE and web services specifications with J2EE version 1.3, Java 2 Platform, Standard Edition version 1.4, and the Java APIs for XML (JAX) from the Java Web Services Developer Pack software. In addition, it offers clustering, load balancing, and HTTP session failover. For a complete list of features, see the *Sun ONE Application Server Product Introduction*.

Application server instances form the basis of an application server deployment. Each instance has its own directory structure, configuration, and deployed applications. Each server instance also includes the J2EE platform web and EJB containers. The Sun ONE Application Server also includes a proven, high-performance HTTP server. An Object Request Broker (ORB) module enables EJB invocation using RMI-IIOP.

To access backend systems, applications can leverage J2EE connector architecture support and third-party resource adapters, either the built-in JMS provider or third-party providers, and any combination of popular third-party JDBC drivers. Access to backend systems can be managed within the scope of distributed transactions using the built-in, all-Java Transaction Manager.

The Admin Server (also called the Administration Server) is a special server instance that runs the core administration application and an SNMP agent. All remote management activity flows through the Admin Server. Both command-line and web-browser -based administrative clients access the Admin Server directly either through HTTP, or securely through HTTPS.

[Figure 1-1](#) shows an application server instance in detail. The application server instance is a building block in the clustering, load balancing, and session persistence features described in [“About Enterprise Edition Features” on page 16](#).

Figure 1-1 Sun ONE Application Server Instance

Web server plug-ins (such as the load balancer plug-in) enable you to deploy the Application Server behind one or more web servers housed in a demilitarized zone (DMZ) that is bracketed by one or more layers of firewalls. The plug-ins provide a means for the front-end web server tier to direct incoming HTTP and HTTPS traffic received from the Internet to one or more Application Servers housed in a back-end application server tier.

A variety of client applications can access business services deployed to the Application Server. Web service clients and browser-based clients can use either HTTP or HTTPS to access the web services, server-side end points, and J2EE web applications.

[Figure 1-1](#) shows a single administrative domain. Administrative domains enable you to define multiple, completely separate application server runtime configurations that reuse the same installation image. Each administrative domain has an Admin Server which in turn controls one or more application server instances. In the tutorials, you will be working with a single administrative domain configured during installation of the product, as well as multiple server instances in a cluster.

About Enterprise Edition Features

You can group application server instances into clusters, and with a web server and a load balancer, provide load balancing and HTTP session persistence. This guide shows you how to use the Enterprise Edition features described in the following sections:

- [Clustering](#)
- [Load Balancing](#)
- [Session Persistence](#)
- [High-Availability Database](#)
- [Example Clustering Scenario](#)

Clustering

A cluster is a group of application server instances that work together as one logical entity. Each application server instance in the cluster has the same configuration and the same applications deployed to it.

Using clusters in Sun ONE Application Server helps you to achieve:

- High availability by allowing for failover protection for the application server instances in a cluster. If one application server instance goes down, other application server instances take over the sessions that the unavailable server was serving. Sun ONE Application Server supports failover for HTTP sessions and EJB references within HTTP sessions.

- Scalability by allowing for the addition of application server instances to a cluster, thus increasing the capacity of the system. The load balancer of Sun ONE Application Server distributes requests to the healthy application server instances within the cluster. You can add more application server instances to a cluster without disrupting service.

The application server instances within a cluster can be hosted on different machines or on the same machine. That is, you can group application server instances across different machines into a cluster. For this guide, the default configuration is for two instances on the same machine.

For more information about clustering, see the *Sun ONE Application Server Administrator's Guide*.

Load Balancing

The goal of load balancing is to evenly distribute the workload among multiple Sun ONE Application Server instances. You can use either the load balancer plug-in supplied with the Sun ONE Application Server, or you can use various third party hardware and software load balancers. This guide describes using the load balancer plug-in.

Sun ONE Application Server's load balancer uses a sticky round robin algorithm to load balance incoming HTTP and HTTPS requests. When a new HTTP request is sent to the load balancer plug-in, the request is forwarded to an application server instance based on a simple round robin scheme. Subsequently, this request is "stuck" to this particular application server instance, either by using cookies or by explicit URL rewriting.

From the sticky information, the load balancer plug-in first determines the instance to which the request was previously forwarded. If that instance is found to be healthy, the load balancer plug-in forwards the request to that specific application server instance.

Because all requests for a given session are sent to the same application server instance, the session data can be cached on a single application server rather than having to be distributed to all instances in a cluster. This caching means that the sticky round robin scheme provides significant performance benefits that normally override the benefits of a more evenly distributed load that can be obtained with pure round robin.

For more information about load balancing, see the *Sun ONE Application Server Administrator's Guide*.

Session Persistence

Session persistence ensures that if a Sun ONE Application Server instance or machine fails, the HTTP/HTTPS request will be picked up by another server. Sun ONE Application Server supports persistence of HTTP sessions as well as failover of EJB references within HTTP sessions.

The high-availability database (HADB) is bundled with the Sun ONE Application Server as the persistence store to provide high availability for web applications. For more information on the high-availability database, see [“High-Availability Database” on page 19](#).

This section contains the following topics:

- [About Session Persistence Types](#)
- [About Session Persistence Configuration](#)
- [About Single Sign-on Session Information](#)

About Session Persistence Types

Sun ONE Application Server supports three types of persistence: `ha`, `file`, and `memory`.

- The `ha` persistence type stores the session information in the HADB. The `ha` persistence type is supported for production environments that require failover capabilities.
- The `file` persistence type stores session information periodically to files, but you may experience loss of data when an instance becomes unavailable. The `file` persistence type is *not* supported for production environments that require failover capabilities; however it is sometimes used in development environments for testing applications.
- The `memory` persistence type stores session information in a file in cases of graceful shutdown of the server instance. It does not store session information for unexpected shutdown. The `memory` persistence type is *not* supported for production environments that require failover capabilities.

The `ha` session persistence is the only type covered in the exercises in this guide. For more information on configuring and using the other types of session persistence, see the *Sun ONE Application Server Administrator's Guide*.

About Session Persistence Configuration

When you install the Sun ONE Application Server and run the `clsetup` command, your session persistence information is configured to default values. However, you can change the default values to balance your particular needs for performance, reliability, and high availability.

For example, every time your data is saved, you can store the entire session, or you can store the session only if the session has been modified. Or, you can also configure the persistence setting so that only the modified attributes of the session are stored.

Similarly, you can choose the persistence frequency. For example, you can choose to store the session after every web request, thus providing high availability and reliability of updated session states, or you can choose to store the session after a time interval you specify, thus providing better performance.

For more information about the various configuration options for session persistence, see the *Sun ONE Application Server Administrator's Guide*.

About Single Sign-on Session Information

In a single application server instance, once you are authenticated by a web application, you are not required to reauthenticate individually to other web applications running on the same instance. This is called single sign-on.

In order for this feature to continue to work even when a session fails over to another instance in a cluster, single sign-on information must be persisted to the HADB. This persistence is enabled when you enable high availability for the application server instance. High-availability is automatically configured when you run the `clsetup` command.

High-Availability Database

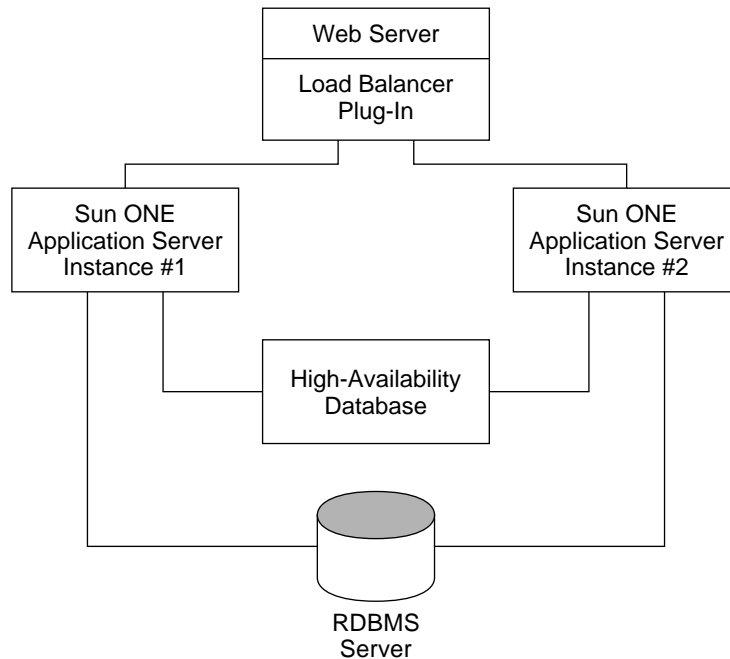
The high-availability database (HADB) is used to store session information so that HTTP sessions can persist. High availability means that the system remains available despite unplanned outages caused by hardware or software failures. The HADB, a JDBC-compliant database, is based on the Always-On technology. The HADB is capable of exceeding 99.999% data availability and offers an ideal platform for delivering all types of HTTP session state persistence within a highly-loaded enterprise application server environment.

The HADB achieves this data availability through fragmentation and replication of data. All tables in the database are partitioned to create subsets of approximately the same size called fragments. This process of fragmentation is based on a hash function that fragments and evenly distributes the data among the database's nodes. Each fragment is stored twice in the database, in mirror nodes. This ensures fault tolerance and fast recovery of data. In addition, if a node fails or is shut down, a spare node can take over until the node is active again.

Without high-availability, when an unplanned outage occurs, the HTTP session state is lost during failover from one web container to another. Use of the HADB as the session persistence mechanism overcomes this situation. The HADB stores and retrieves all state information in a separate but tightly-integrated persistent storage tier.

Example Clustering Scenario

The following diagram shows a simple clustering scenario consisting of a web server with a load balancer plug-in, two Sun ONE Application Server instances configured to use the HADB to store HTTP session data, and a remote database management system (RDBMS) to store application data. Note that your actual operational deployment may vary. For example, you could use a third-party load balancer in place of the load balancer plug-in.

Figure 1-2 Example Clustering Scenario

The following steps describe how a request is serviced:

1. An incoming client sends an HTTP request to a URL handled by the web server. The web server has been configured to allow the load balancer plug-in to handle the incoming HTTP request.
2. The load balancer plug-in then forwards the request to one of the Sun ONE Application Server instances in the cluster. The plug-in uses sticky round-robin load balancing to determine the target instance.
3. The target instance receives the request forwarded to it from the load balancer plug-in and begins an HTTP session, storing the HTTP session data to the HADB and the J2EE application data to the RDBMS. As the client progresses through the application, HTTP session data is updated and stored to the HADB and application data is updated in the RDBMS.
4. If the instance fails due to a system crash, the load balancer detects that the instance has ceased responding to requests. When subsequent requests come in, the load balancer forwards them to a healthy instance in the cluster.

5. The new target instance retrieves the failed over HTTP session information from the HADB and continues responding to the client's requests, allowing the client to complete the HTTP session with no loss of session data.

For more information about deployment scenarios for the Sun ONE Application Server, see the *Sun ONE Application Server System Deployment Guide*.

Tools for Configuring and Administering the Application Server

Sun ONE Application Server provides the following tools for configuring and administering the Sun ONE Application Server. You can use these tools to start and stop the server, as well as perform various other functions. Detailed steps for using some of these tools to perform specific configuration tasks are included in the tutorials.

- **The `asadmin` utility.** The `asadmin` utility is a command-line interface that you can use to perform administrative tasks on a single application server instance or administrative domain.
- **The `cladmin` command.** The `cladmin` command runs certain `asadmin` commands simultaneously on all application server instances in a cluster. Using the `cladmin` command assures you that all instances in the cluster are being configured identically, so you should use `cladmin`, rather than `asadmin`, whenever possible. Not all `asadmin` commands are supported by `cladmin`.
- **The Administration interface.** The Administration interface is a web-based user interface that you can use to configure the Admin Sever and individual application server instances. You can also use it to view the log files of individual application server instances.
- **The `clsetup` command.** The `clsetup` command allows you to easily set up a cluster, including initial configuration of the HADB database. For full instructions on the `clsetup` command, see the *Sun ONE Application Server Installation Guide*.

NOTE Some aspects of server configuration and corresponding interfaces are still undergoing change. An unstable interface may be removed and replaced with a cleaner and more stable version in a subsequent product release. Most server configuration and administrative interfaces will remain the same or change in a compatible manner; however some aspects may change incompatibly. Product documentation for future releases will clearly describe incompatible changes when they do occur.

Enterprise Features Configuration Tutorial

This chapter includes a tutorial to set up and use features in your Sun ONE Application Server environment. This chapter contains the following sections:

- [Preparing to Use the Tutorials](#)
- [Overview of Tutorial Steps](#)
- [Starting the Server](#)
- [Verifying Server Start-up](#)
- [Creating the loadbalancer.xml File](#)
- [Adding a Cluster to the loadbalancer.xml File](#)
- [Configuring Load Balancing](#)
- [Sample loadbalancer.xml File](#)

Preparing to Use the Tutorials

This guide walks you through setting up a clustered environment, deploying an application to it, and testing the load balancing and HTTP session persistence features with the application. It includes two tutorials: the [Enterprise Features Configuration Tutorial](#), which covers configuring the cluster and load balancer, and the [Cluster JSP Sample Application Tutorial](#), which covers deploying the application and verifying that load balancing and session persistence work.

This section contains the following topics:

- [Installation Requirements](#)
- [Required Steps Before Using the Tutorials](#)

Installation Requirements

The installation configuration suggested here is a basic, single-machine configuration for use with this guide. It is not necessarily the right configuration for your environment. You need to carefully consider your needs before deploying the Sun ONE Application Server in an your environment. For more information on recommended configurations, see the *Sun ONE Application Server System Deployment Guide*.

For the purposes of doing the tutorials included in this book, use the following basic configuration:

- One Sun ONE Web Server
- One load balancer plug-in
- ONE Sun ONE Application Server with two server instances configured to use high-availability databases
- Two high-availability databases

NOTE	All the instructions in this guide assume that you are using the Sun ONE Web Server. You can also use the Apache web server. For more information on using the Apache web server, see the <i>Sun ONE Application Server Administrator's Guide</i> .
-------------	---

Required Steps Before Using the Tutorials

Before you start the tutorials, you must have already completed the following steps.

Table 2-1 Steps for Preparing to Use the Tutorials

Step	Location of Instructions
1. Install Sun ONE Web Server	<i>Sun ONE Web Server Installation Guide</i>
2. Install Sun ONE Application Server, including: <ul style="list-style-type: none"> • The load balancer plug-in • The HADB 	<i>Sun ONE Application Server Installation Guide</i>
3. Set up the following for your machine: <ul style="list-style-type: none"> • Shared memory • Communication using rsh or ssh • Environment variables for using the HADB 	<i>Sun ONE Application Server Installation Guide</i>
4. Run <code>clsetup</code> to perform the following configuration: <ul style="list-style-type: none"> • Create and configure two application server instances, <code>server1</code> and <code>server2</code>, in the domain <code>domain1</code> • Create the HADB. • Create the database tables required to store session information in the HADB. • Create a connection pool in the instances • Create a JDBC resource in all the instances. • Configure the session persistence information in the application server instances. • Enable high availability in the application server instances. 	<i>Sun ONE Application Server Installation Guide</i>

Overview of Tutorial Steps

The following table summarizes the steps required in the tutorials to configure your system, deploy an application, and verify the enterprise features. The left column describes the step, and the right column gives the location of the instructions needed to perform the step.

Table 2-2 Steps for Using Tutorials

Step	Location of Instructions
1. Start the Admin Server and application server instances.	“Starting the Server” on page 29
2. Verify that the Admin Server and application server instances are running.	“Verifying Server Start-up” on page 31
3. Create a <code>loadbalancer.xml</code> file.	“Creating the loadblancer.xml File” on page 37
4. Set up a cluster in <code>loadbalancer.xml</code> .	“Adding a Cluster to the loadbalancer.xml File” on page 38
5. Configure the load balancer in the <code>loadbalancer.xml</code> file.	“Configuring Load Balancing” on page 40
6. Verify your <code>loadbalancer.xml</code> file against the sample <code>loadbalancer.xml</code> file.	“Sample loadbalancer.xml File” on page 45
7. Deploy the sample application to the instances in your cluster using <code>cladmin</code> .	“Deploying the Cluster JSP Sample Application to a Cluster” on page 48
8. Start the application server instances using <code>cladmin</code> .	“Starting the Application Server Instances Using cladmin” on page 52
9. Verify that you deployed your application successfully.	“Verifying the Application Deployment” on page 53
10. Add the sample application to the cluster information in <code>loadbalancer.xml</code> .	“Adding the Sample Application to the Cluster” on page 59
11. Apply changes and restart the web server.	“Applying Configuration Changes and Restarting the Web Server” on page 61
12. Run the sample.	“Running the Application” on page 62
13. Verify load balancing.	“Verifying Load Balancing” on page 67
14. Verify session persistence.	“Verifying HTTP Session Persistence” on page 71
15. Quiesce a server instance.	“Quiescing a Server Instance” on page 72

Starting the Server

Before you can use the tutorials in this guide, you need to start the Admin Server and all the application server instances in your cluster. Your Sun ONE Web Server should also be running.

The most straightforward way to start the Admin Server and the application server instances when all the server instances and the Admin Server are in a single domain is to use the `asadmin` utility's `start-domain` command.

This section contains the following topics:

- [Setting the PATH Variable](#)
- [Running `asadmin start-domain`](#)

Setting the PATH Variable

Before running `asadmin`, for added convenience you can set your PATH variable.

Setting the PATH variable allows you to run `asadmin`, `cladmin`, and the `asant` (the application server's Ant utility) from the command line at any location. Add the following directory to your PATH environment variable:

```
install_dir/bin
```

If you add the Application Server's `bin` directory to your login profile it is automatically added to your environment's PATH setting during login.

For example, add the following to your `.cshrc` file:

```
set path=( /opt/SUNWAppserver7/bin)
```

After you have saved the file, source it:

```
source .cshrc
```

You can test your change by executing the `asadmin` utility at the command prompt. Type:

```
asadmin
```

The following result should appear:

```
Use "exit" to exit and "help" for online help
```

```
asadmin>
```

Type `exit` to quit the `asadmin` interface.

If the command is not found:

- Check that you have updated your PATH setting correctly.
- If you updated your login file (for example, your `.cshrc` file), make sure you sourced the file. You might also need to start a new terminal window.

If you do not set the PATH environment variable, you can still run the commands and utilities from the directory where they are located (*install_dir/bin*), for example:

```
cd /opt/SUNWappserver7/bin
./asadmin
```

Running asadmin start-domain

To start the Admin Server and all the application server instances in it, run the `asadmin start-domain` command. At the command prompt, type:

```
asadmin start-domain --domain domain1
```

`domain1` is the default domain configured when you installed the server and ran the `clsetup` command.

NOTE In the configuration described in this guide, all the instances are in the same domain. If the server instances are not in the same domain, you need to run `asadmin start-domain` for every domain you plan to use. Once you have created a cluster, you can also use `cladmin start-instance` to start all the instances in a cluster. For information on `cladmin` syntax, see [“The cladmin Command Syntax” on page 50](#).

Use the following command to stop both the Admin Server and the application server instances of the initially configured domain:

```
asadmin stop-domain --domain domain1
```

where `domain1` is the name of the administrative domain defined during installation of the application server.

For a full list of `asadmin` commands, see the `asadmin help`. To access help, at the command prompt, type:

```
asadmin
```

Your prompt should change to the `asadmin` prompt, if you have set your `PATH` variable. Type `help` at the prompt to see a list of all `asadmin` commands. To get help on a specific `asadmin` command, at the `asadmin` prompt type the command name followed by the `help` option. For example:

```
asadmin> start-domain --help
```

Verifying Server Start-up

After starting the Admin Server and application server instances, verify that they have started successfully, as described in the following sections:

- [Verifying Admin Server Start-up](#)
- [Verifying Start-up of Server Instances](#)
- [Verifying Instances by Accessing the HTTP Server](#)

Verifying Admin Server Start-up

You can verify that the Admin Server started by accessing the Sun ONE Application Server's Administration interface, or by viewing the Admin Server's event log, as described in the following sections:

- [Accessing the Administration Interface](#)
- [Viewing the Admin Server's Event Log](#)

Accessing the Administration Interface

To verify that the Admin Server started correctly, you can access the Sun ONE Application Server's web-based Administration interface. You can use this interface to administer the Sun ONE Application server instances, though for clustered instances it is more convenient and less error-prone to use the `cladmin` command.

NOTE For a list of browsers compatible with the Administration interface, see the *Sun ONE Application Server Platform Summary*.

To access the Administration interface, follow these steps:

1. Open a browser window and specify the Admin Server port.

During installation, the default port number for the Admin Server is set to 4848. If this port was already in use or you selected another port number, then specify that port number.

For example:

`http://test.sun.com:4848`

2. Sign into the Administration interface using the administration user name and password supplied during product installation.

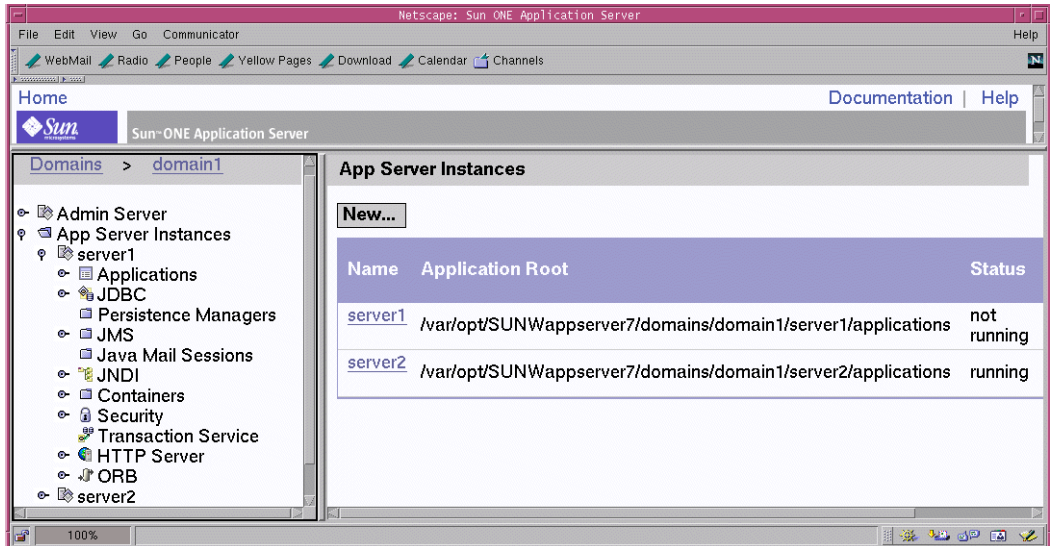
TIP	Forgot the user name or password? If you do not remember the Admin Server user name that was supplied during installation, try the user name <code>admin</code> , the default user name specified in the server configuration dialog during installation.
------------	--

If you don't remember the administrator's password, look in the `clpassword.conf` file in the *install_config_dir*, created when you installed the server. You must be root to access this file.

TIP	Port not accessible? If the connection was refused when attempting to contact the Admin Server's Administration interface, it is likely that the Admin Server is not running. Check your start-up procedures and the content of the Admin Server's log file to determine the reason why the server is not running. For instructions on viewing the log file, see “Viewing the Admin Server's Event Log” on page 33 .
------------	---

Once you've authenticated successfully, the initial screen of the Administration interface appears:

Figure 2-1 Administration Interface



When you click on an item in the left pane, the corresponding page appears in the right pane.

Viewing the Admin Server's Event Log

You can also look in your Admin Server event log for a server start-up message. To open and view the event log file, follow these steps:

1. Navigate to the server logs for the Admin Server:

```
domain_config_dir/domain1/admin-server/logs/
```

2. Open `server.log` in an editor.

If your server has been started, you will find `successful server startup` at the end of the log file:

```
[20/Feb/2003:00:06:00] INFO ( 4318): CORE3274: successful server startup
```

If you don't see the successful start-up message, you may have opened the event log file prior to the Admin Server completing its startup procedures. Close the log file and open it again to see the very latest event messages.

You can also use the `tail -f` command to view messages in the server log.

```
tail -f server.log
```

The `-f` option leaves the `tail` command running so new log entries are displayed as they are written to the file.

You can also access the Admin Server and the application server instance event log files through the Administration interface. Click the server name in the left pane (Admin Server or the application server instance) and click the Logging tab in the right pane.

For more information, see the *Sun ONE Application Server Administrator's Guide*.

Verifying Start-up of Server Instances

After running `clsetup`, you should have two server instances in domain1: `server1` and `server2`. This section discusses two ways to verify that those instances exist, and to verify that they are running:

- [Verifying Instances using asadmin](#)
- [Verifying Instances by Accessing the HTTP Server](#)

Verifying Instances using asadmin

To verify, use the `asadmin list-instances` command:

1. Launch `asadmin` by typing `asadmin` at the command prompt:

```
asadmin
```

Your prompt becomes `asadmin>`

2. Use the `list-instances` command at the `asadmin` prompt to show all current instances.

```
list-instances
```

This command lists the instances and whether or not they are currently running. You should see two running instances, `server1` and `server2`.

Verifying Instances by Accessing the HTTP Server

Another means of determining whether or not an application server instance has started is to access the instance's HTTP server welcome page through a web browser.

Using a browser, access the following location:

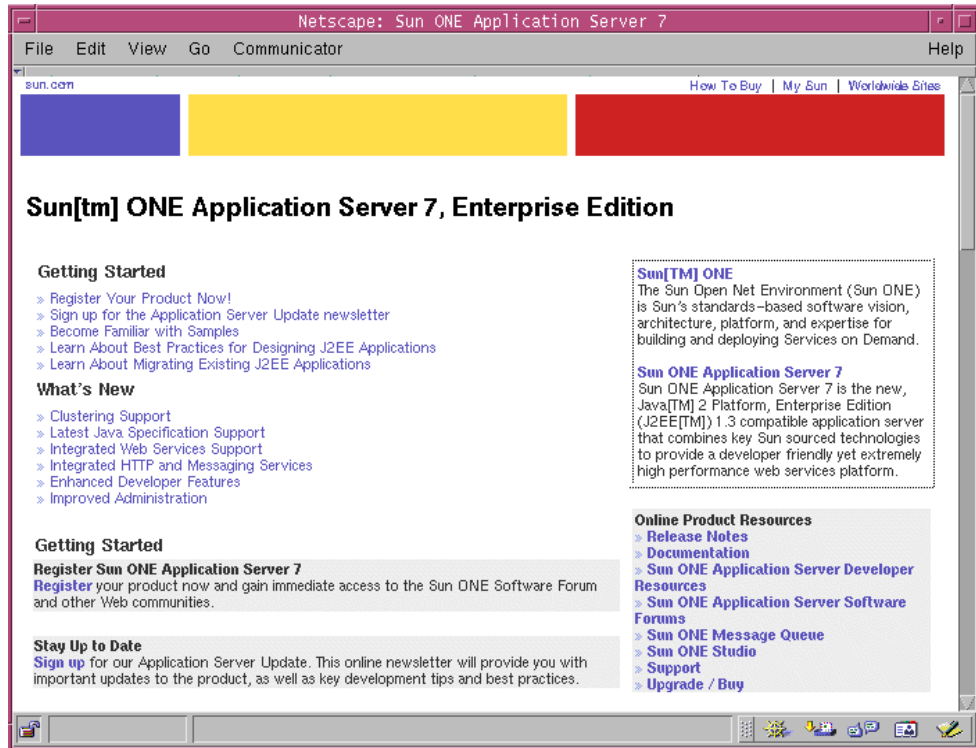
`http://server_instance:server_instance_port_number`

where *server_instance_port_number* is the HTTP server port number specified during installation or server instance creation.

TIP	<p>If you do not remember the HTTP server port number of your server, you can inspect the application server instance's configuration file to determine the HTTP server port number:</p> <ol style="list-style-type: none">1. Navigate to the <i>domain_config_dir</i>/domain1/<i>server_instance</i>/config/ directory and open the <code>server.xml</code> file in your favorite editor.2. Look for the <code>http-listener</code> element, for example: <pre>http-listener id="http-listener-1" address="0.0.0.0" port="81"...</pre> <p>In this case, port 81 is the HTTP port number in use.</p>
------------	---

If the application server instance is up and running normally, you should see the following default HTTP server welcome page in your browser:

Figure 2-2 HTTP Server Welcome Page



TIP **HTTP server welcome page:** The HTTP server welcome page is an HTML page named `index.html` in the application server instance's default document directory. The application server instance's `server.xml` configuration file contains the setting for the default document root of the instance. After installation, the document root for the instance named `server1` is set to `domain_config_dir/domain1/server1/docroot/`. You can find the welcome page at that location.

Creating the loadbalancer.xml File

After you have installed the components, run `clsetup`, and started the Admin Server, you are ready to configure your installation. The first step is to create a cluster.

Since the configuration of the cluster for this guide uses one load balancer plug-in, you must define the cluster in the web server's `loadbalancer.xml` configuration file. This file belongs in the configuration file directory of your web server. The `loadbalancer.xml` file does not exist by default. You must create it.

To create the `loadbalancer.xml` file, follow these steps:

1. Find the sample `loadbalancer.xml` file, called `loadbalancer.xml.example`, located in your web server's `config` directory.

By default, this directory is:

```
/usr/iplanet/servers/https-server_name/config
```

2. Save a copy of `loadbalancer.xml.example` as `loadbalancer.xml`.

You will edit this copy to include information for your environment, including the information that will create a cluster and allow load balancing to work.

Here is the example file:

Code Example 2-1 The loadbalancer.xml.example file

```
<!DOCTYPE loadbalancer PUBLIC "-//Sun Microsystems Inc.//DTD Sun ONE Application
Server 7.0//EN" "sun-loadbalancer_1_0.dtd">

<loadbalancer>

  <cluster name="cluster1">

    <instance name="instance1" enabled="true" disable-timeout-in-minutes="60"
listeners="<REPLACE_WITH_LISTENER1> <REPLACE_WITH_LISTENER2>" />

    <web-module context-root="/abc" enabled="true"
disable-timeout-in-minutes="60" enabled="true" />

    <health-checker url="/" interval-in-seconds="10" timeout-in-seconds="30"
/>

  </cluster>

  <property name="reload-poll-interval-in-seconds" value="60"/>

  <property name="response-timeout-in-seconds" value="30"/>
```

```
<property name="https-routing" value="true"/>
<property name="require-monitor-data" value="false"/>
</loadbalancer>
```

Notice the example file refers to the `sun-loadbalancer_1_0.dtd`, also found in the web server's `config` directory.

Adding a Cluster to the loadbalancer.xml File

Next you need to add a cluster, for this tutorial called `cluster1`, to your `loadbalancer.xml` file. The cluster includes two Sun ONE Application Servers or server instances.

Edit your newly-created `loadbalancer.xml` to include the instances and the URLs to the HTTP listeners:

1. Go to your web server's `config` directory and open `loadbalancer.xml` using a text editor.

Notice that the sample file already includes a cluster called `cluster1`.

```
<cluster name="cluster1">
```

2. Add the two application server instances that you created with `clsetup`, `server1` and `server2`, to `cluster1`.

Create an instance subelement of the cluster for each of the instances. The instances need to be enabled (`enabled=true`). You also need to add the URLs to the HTTP listeners. In the case of the default server instance (`server1`) and the second server instance you added (`server2`) there is only one HTTP listener for each. You could have more in a different environment.

The sample `loadbalancer.xml` file contains the following lines that you can edit:

```
<instance name="instance1" enabled="true"
disable-timeout-in-minutes="60"
listeners="<REPLACE_WITH_LISTENER1> <REPLACE_WITH_LISTENER2>" />
```

3. Replace `"instance1"` with `"server1"` for this tutorial.

4. Replace `<REPLACE_WITH_LISTENER1>` with the listener URL for your server instance, for example:

```
http://test.sun.com:81.
```

The listener URL consists of the server name and port. If the listener has security enabled, the URL starts with HTTPS. If the listener is not secure, the URL starts with HTTP. To find information on the HTTP listener for each Sun ONE Application Server instance:

- a. In the Administration interface, click the instance name to expand the tree view.
 - b. Click HTTP Server to expand it.
 - c. Under HTTP Server, click HTTP Listener.
 - d. The HTTP listener page gives you the port, the server name, and whether security is enabled.
5. Save your changes to `loadbalancer.xml`. Your file now has a line similar to the following:

```
<instance name="server1" enabled="true"
disable-timeout-in-minutes="60"
listeners="http://test.sun.com:81"/>
```

6. Change the `disable-timeout-in-minutes` value to 5.

The disable timeout in minutes is the instance-level quiescing period, in which you allow time for sessions to terminate for a server instance. The server is still running, but no requests that would result in new sessions are sent to it, though the server does accept requests that are for existing sessions. In the real world, this number would be quite high. The time specified in the example is 60 minutes. However, because the tutorial requires you to shut down the server, in this tutorial you are asked to set the timeout artificially low.

```
<instance name="server1" enabled="true"
disable-timeout-in-minutes="5"
listeners="http://test.sun.com:81"/>
```

7. Add a second instance to the cluster element. You can do that by copying the line above and pasting it below the existing line, then changing the server name and the listener URL. For example:

```
<instance name="server2" enabled="true"
disable-timeout-in-minutes="5"
listeners="http://test.sun.com:82"/>
```

8. Save your changes.

So far, your `loadbalancer.xml` file should look similar to the following:

```
<!DOCTYPE loadbalancer PUBLIC "-//Sun Microsystems Inc.//DTD Sun ONE Application
Server 7.0//EN" "sun-loadbalancer_1_0.dtd">

<loadbalancer>

  <cluster name="cluster1">

    <instance name="server1" enabled="true" disable-timeout-in-minutes="5"
listeners="http://test.sun.com:81"/>

    <instance name="server2" enabled="true" disable-timeout-in-minutes="5"
listeners="http://test.sun.com:82"/>

    <web-module context-root="/abc" enabled="true"
disable-timeout-in-minutes="60" enabled="true" />

    <health-checker url="/" interval-in-seconds="10" timeout-in-seconds="30"
/>

  </cluster>

  <property name="reload-poll-interval-in-seconds" value="60"/>
  <property name="response-timeout-in-seconds" value="30"/>
  <property name="https-routing" value="true"/>
  <property name="require-monitor-data" value="false"/>
</loadbalancer>
```

You'll add information associated with the deployed application (the sample line that begins with "web-module") during a later procedure.

NOTE	Enabling an application server instance in the <code>loadbalancer.xml</code> file is different from starting an application server instance. Even if the server is running and you have added it to the cluster, the load balancer does not route requests to the server instance until you enable it in <code>loadbalancer.xml</code> by setting the instance element's <code>enabled</code> attribute to <code>true</code> .
-------------	--

Configuring Load Balancing

After creating the `loadbalancer.xml` file and adding a cluster, edit the `loadbalancer.xml` file to include the load balancer settings.

This section contains the following topics:

- [Configuring the Health Checker](#)
- [Enabling Load Balancer Monitoring](#)
- [Other Load Balancer Properties](#)

Configuring the Health Checker

If you enable the health checker, the load balancer periodically checks all the Sun ONE Application Server instances that are flagged as unhealthy to see whether they have become healthy. If an application server instance that was marked as unhealthy has become healthy, the instance is added to the list of healthy instances, and requests are once more routed to it.

You configure the health checker in the `health-checker` element in `loadbalancer.xml`. You configure the listener URL that the health checker pings to see if the server is running, as well as how often the health checker pings the server and how long the health checker waits for a response before marking the server as unhealthy.

The sample `loadbalancer.xml` file includes the following line, which you do not need to edit in order to complete the tasks in this guide.

```
health-checker url="/" interval-in-seconds="10"
timeout-in-seconds="30" />
```

- The `health-checker url` specifies the URL to ping to determine whether the server instance is healthy (that is, whether it is responding). The example shows a URL of `"/"`. For example, if your listener URL is `http://www.example.com:80`, then a health checker URL of `"/"` pings `http://www.example.com:80/`. For the purposes of this guide, leave the value `"/"`.
- The `interval-in-seconds` attribute specifies the interval between load balancer health checks. You do not need to change the value from the default 10.
- The `timeout-in-seconds` attribute specifies the interval within which the load balancer must receive a response from the pinged server instance in order for the server instance to be considered healthy. You do not need to change the value from the default 30.

For example, using the default values, the health checker follows these steps:

1. The health checker pings the unhealthy listeners using the health checker URL.
2. For each listener, it waits 30 seconds (`timeout-in-seconds`) for a response.

If the server responds within 30 seconds, the listener is marked healthy. If the server does not respond, the health checker considers the server to still be unhealthy.

3. The health checker waits ten seconds (the `interval-in-seconds`) before starting another cycle of health checks.

If there are no unhealthy instances when the health checker starts, it skips the first two steps.

Enabling Load Balancer Monitoring

The load balancer plug-in uses the web server logging mechanism to write log messages. When monitoring is enabled, the following information is logged to the web server log file:

- Start/stop information for every request
- Failover request information when a request is failed over from an unhealthy to a healthy instance
- List of unhealthy instances logged at the end of every health checker cycle

CAUTION When logging is enabled on the load balancer plug-in, and if the web server logging level is set to `DEBUG` or to print verbose messages, the load balancer writes HTTP session IDs in the web server log files. Therefore, if the web server hosting the load balancer plug-in is in the DMZ, we recommend that you do not use the `DEBUG` or similar log level in production environments.

If you must use the `DEBUG` logging level, then you should turn off load balancer logging by setting `require-monitor-data` property to `false` in `loadbalancer.xml`.

To monitor load balancing, you need to:

- Enable monitoring in `loadbalancer.xml`
- Configure the web server to use verbose logging.

To enable monitoring in `loadbalancer.xml`, follow these steps:

1. Open `loadbalancer.xml` using a text editor.
2. Find the following line (it was part of the example `loadbalancer.xml` file):

```
<property name="require-monitor-data" value="false"/>
```

3. Change “false” to “true” to enable monitoring.
4. Save your changes and exit the file.

In addition, to change the default web server log level, follow these steps:

Turn on the Sun ONE Web Server’s verbose logging:

1. Go to the web server’s `config` directory. If you installed your Sun ONE Web Server in the default location, the path will be:

```
/web_server_install_dir/https-server_name/config
```

2. Open the `magnus.conf` file for editing.
3. Add the following line:

```
LogVerbose on
```

4. Save the changes to the file.
5. Restart the web server to apply the changes.

Other Load Balancer Properties

The example `loadbalancer.xml` contains the following additional load balancer properties:

```
<property name="reload-poll-interval-in-seconds" value="60"/>
```

```
<property name="response-timeout-in-seconds" value="30"/>
```

```
<property name="https-routing" value="true"/>
```

For this tutorial, you do not need to change the default values. For more information on these properties, see the following sections:

- [Dynamic Reconfiguration using the Reload Poll Interval](#)
- [Response Timeout](#)
- [HTTPS Routing](#)

Dynamic Reconfiguration using the Reload Poll Interval

After the initial configuration, the load balancer plug-in detects changes to its configuration and loads them automatically. It detects changes by looking at the time stamp on the `loadbalancer.xml` file. If the time stamp has changed, the load balancer automatically reconfigures itself. The reload poll interval specifies how often the load balancer checks the time stamp.

NOTE

- If the changes to the `loadbalancer.xml` file are not in the correct format as specified in the `sun-loadbalancer_1_0.dtd` file, the reconfigure operation fails. The failure notice is logged to the web server's error log files.

The load balancer will continue to use the old configuration that's already loaded in memory.

- If the load balancer encounters a hard disk read error while attempting to reconfigure itself, it uses the configuration that is currently in memory.

If a disk read error is encountered, a warning message is logged to the web server's error log file.

The Sun ONE Web Server's error log file can be found at:
web_server_install_dir/web_server_instance/logs/.

Response Timeout

The response timeout designates how long the server waits for the instance to respond to a request. After the specified number of seconds, if the instance has not responded an error message is sent to the browser.

HTTPS Routing

When HTTPS routing is disabled, HTTPS listeners listed in the `loadbalancer.xml` file are ignored, and only HTTP listeners are used for load balancing. When HTTPS routing is enabled, HTTPS requests are failed over only to HTTPS ports.

Sample loadbalancer.xml File

The following sample `loadbalancer.xml` file includes one cluster of two instances. Check the syntax of your `loadbalancer.xml` file against this example.

Code Example 2-2 The loadbalancer.xml File

```
<!DOCTYPE loadbalancer PUBLIC "-//Sun Microsystems Inc.//DTD Sun ONE Application
Server 7.0//EN" "sun-loadbalancer_1_0.dtd">
```

```
<loadbalancer>

  <cluster name="cluster1">

    <instance name="server1" enabled="true" disable-timeout-in-minutes="5"
listeners="http://test.sun.com:81"/>

    <instance name="server2" enabled="true" disable-timeout-in-minutes="5"
listeners="http://test.sun.com:82"/>

    <web-module context-root="/abc" enabled="true"
disable-timeout-in-minutes="60" enabled="true" />

    <health-checker url="/" interval-in-seconds="10" timeout-in-seconds="30"
/>

  </cluster>

  <property name="reload-poll-interval-in-seconds" value="60"/>
  <property name="response-timeout-in-seconds" value="30"/>
  <property name="https-routing" value="true"/>
  <property name="require-monitor-data" value="true"/>

</loadbalancer>
```

Now that you have configured the enterprise features, you can deploy and run the sample application. Continue on to [Chapter 3, “Cluster JSP Sample Application Tutorial.”](#)

Sample loadbalancer.xml File

Cluster JSP Sample Application Tutorial

This chapter contains the following sections:

- [Preparing to use the Cluster JSP Sample Application Tutorial](#)
- [Deploying the Cluster JSP Sample Application to a Cluster](#)
- [Starting the Application Server Instances Using cladmin](#)
- [Verifying the Application Deployment](#)
- [Adding the Sample Application to the Cluster](#)
- [Applying Configuration Changes and Restarting the Web Server](#)
- [Running the Application](#)
- [Verifying Load Balancing](#)
- [Verifying HTTP Session Persistence](#)
- [Quiescing a Server Instance](#)

Preparing to use the Cluster JSP Sample Application Tutorial

Before you use this tutorial, you must have already:

- Successfully completed all the tutorial steps in [Chapter 2, “Enterprise Features Configuration Tutorial.”](#)
- Made a copy of the sample applications, if necessary.

You may need to make a copy of the samples if you are either sharing your application server installation with other users or your system user ID does not have write permissions to the area in which the application server is installed.

To copy the samples, copy the *install_dir/samples* directory to a location in which your user ID has write permissions.

For example:

```
cp -r install_dir/samples user_sample_directory
```

The *samples* directory contains the subdirectory *ee-samples* which contains the sample used in this guide.

If you are using a copy of the samples, as you proceed with this guide, replace *install_dir/samples/* with the location of your own copy of the sample applications.

Deploying the Cluster JSP Sample Application to a Cluster

The Cluster JSP sample application demonstrates how a request to a JSP can be load balanced between application servers in a cluster. It has a shopping cart which demonstrates how HTTP session information persists, even if an application server has an outage.

The sample application comes with a WAR file that is ready to deploy as a web application. You do not need to compile or assemble the application.

NOTE All applications that use session failover must be distributable. The `clusterjsp` sample application is already distributable, as you can see if you look in its `web.xml` file.

To see this:

1. Go to the Cluster JSP sample's `src` directory:

```
cd /install_dir/samples/ee-samples/clusterjsp/src
```

2. Look at the `web.xml` file. You see the following code:

```
<web-app>
    <display-name>clusterjsp</display-name>
    <distributable/>
```

The `web-app` element has the `distributable` subelement specified. No changes are required for the application code.

To deploy the sample application to a cluster, you must first deploy it to every instance in the cluster.

Use the `cladmin` command to deploy an application to all instances in your cluster at once. The `cladmin` command runs an `asadmin` command on all instances in a cluster simultaneously. The `cladmin` command is located in `install_dir/bin`.

This section contains the following topics:

- [Input Files for the cladmin Command](#)
- [The cladmin Command Syntax](#)
- [Running cladmin deploy](#)
- [The asadmin Commands Supported by the cladmin Command](#)
- [Requirements and Limitations](#)

Input Files for the cladmin Command

The `cladmin` command uses two input files: `clinstance.conf` and `clpassword.conf`.

- `clinstance.conf`: This file contains information about the application server instances that are part of the cluster.

- `clpassword.conf`: This file contains the administration server password and is pre-populated with the correct password during a standard installation.

These files are located in the *install_config_dir*. Since you used these files to run `clsetup` earlier, these files should have the correct values for your environment.

Because many of the values you would have to specify to run `cladmin deploy` are included in these files, you can run the `deploy` command with a minimum of options.

The cladmin Command Syntax

The syntax of the `cladmin` command is as follows:

```
cladmin [--help] [--instancefile instance_file_location] [--passwordfile  
password_file_location] asadmin_command
```

where:

- *instance_file_location* is the location of the input file `clinstance.conf`
- *password_file_location* is the location of the input file `clpassword.conf`
- *asadmin_command* is the `asadmin` command that you want to run on the application server instances in the cluster.

If the input files are located in the default location `/etc/opt/SUNWappserver7`, you can omit the `instancefile` and `passwordfile` options and run the command as follows:

```
cladmin asadmin_command
```

Running cladmin deploy

To deploy the application to all the instances in cluster, type:

```
cladmin deploy filepath
```

You can deploy the sample as a web application using the WAR file provided with the sample, the *filepath* is the path to the WAR file.

For example, if your Cluster JSP application is in the default location, type:

```
cladmin deploy /opt/SUNWappserver7/samples/ee-samples/clusterjsp/clusterjsp.war
```

If you have set your PATH variable, you can also go to the directory where the WAR file is located and deploy the application from there. For example:

```
cd install_dir/samples/ee-samples/clusterjsp
```

```
cladmin deploy clusterjsp.war
```

When you run the `cladmin deploy` command, you see messages as the application is deployed to each instance in the cluster.

If you encounter errors, consult the log file at `/var/tmp/cladmin.log` for more information.

NOTE If you are running as root and have not set the PATH variable for root, you may not be able to run `cladmin` from any directory. Change to the `install_dir/bin` directory, then type `./cladmin asadmin_command` at the command prompt.

The asadmin Commands Supported by the cladmin Command

You can use the `cladmin` command to run the following `asadmin` commands simultaneously for each application server instance in a cluster:

Table 3-1 The asadmin Commands Supported by cladmin

Command	Use
<code>start-instance</code>	Starts the application server instances
<code>stop-instance</code>	Stops the application server instances
<code>deploy</code>	Deploys an EJB, WEB, connector, appclient, or application component that is in the directory to the application server instances
<code>undeploy</code>	Removes the deployed component from the application server instances
<code>create-jdbc-resource</code>	Creates a JDBC resource for the application server instances
<code>create-jdbc-connection-pool</code>	Creates a JDBC connection pool for the application server instances
<code>configure-session-persistence</code>	Configures session persistence for the application server instances

Table 3-1 The asadmin Commands Supported by cladmin

Command	Use
delete-jdbc-resource	Deletes the JDBC resource for the application server instances
delete-jdbc-connection-pool	Deletes the JDBC connection pool for the application server instances

For more information about the `cladmin` command, see the *Sun ONE Application Server Administrator's Guide*.

Requirements and Limitations

The `cladmin` command has the following requirements and limitations:

- All instances that are part of the cluster must have the same user name and password for the administrator.
- Before running the command you must start the Admin Servers associated with all application server instances that are part of the cluster.
- The values specified in the input files will be same for all the instances in a cluster. The `cladmin` command is not designed to set up each instance with different values. For example, this command cannot create a JDBC Connection Pool with different settings for each instance.

Starting the Application Server Instances Using cladmin

After you deploy the sample application, verify that the server instances are running.

1. At the command prompt, type:

```
asadmin list-instances
```

The command displays the instances and whether they are running or not.

2. If the instances are not running, start them using `cladmin` as follows:

```
cladmin start-instance
```

You see messages as each server instance is started.

Verifying the Application Deployment

Once the application is deployed, you can run it in the Sun ONE Application Server instance to test the deployment. To run the sample application, follow these steps:

1. From a browser, access the following URL:

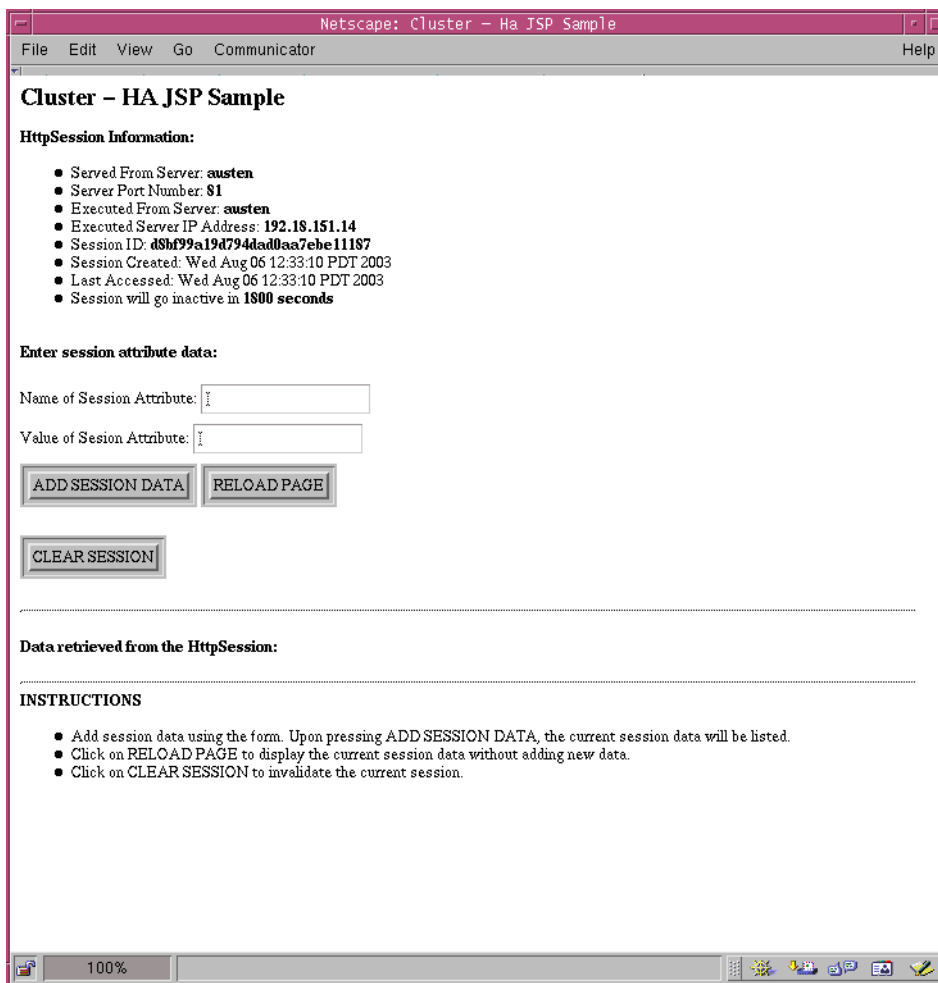
```
http://host:application_server_instance_port/clusterjsp
```

For example:

```
http://test.sun.com:81/clusterjsp
```

2. You see the page for this sample application.

Figure 3-1 Cluster JSP Sample Application Page



If you encounter an error, see [“Troubleshooting Deployment to an Application Server Instance”](#) on page 59.

3. Enter some information in the session attribute fields and click ADD SESSION DATA.

The application displays your data.

4. To verify deployment to the second instance, type the other instance's port and hostname in the URL:

`http://host:application_server_instance_port/clusterjsp`

For example:

`http://test.sun.com:82/clusterjsp`

These URLs verify that the application was deployed to the application server instances. The application has not yet been added to the cluster in `loadbalancer.xml`, so you cannot yet verify it through the web server.

Monitoring the Sample in the Application Server

You can monitor the sample application by looking in the event log file of the application server instance. Each application server instance has its own event log.

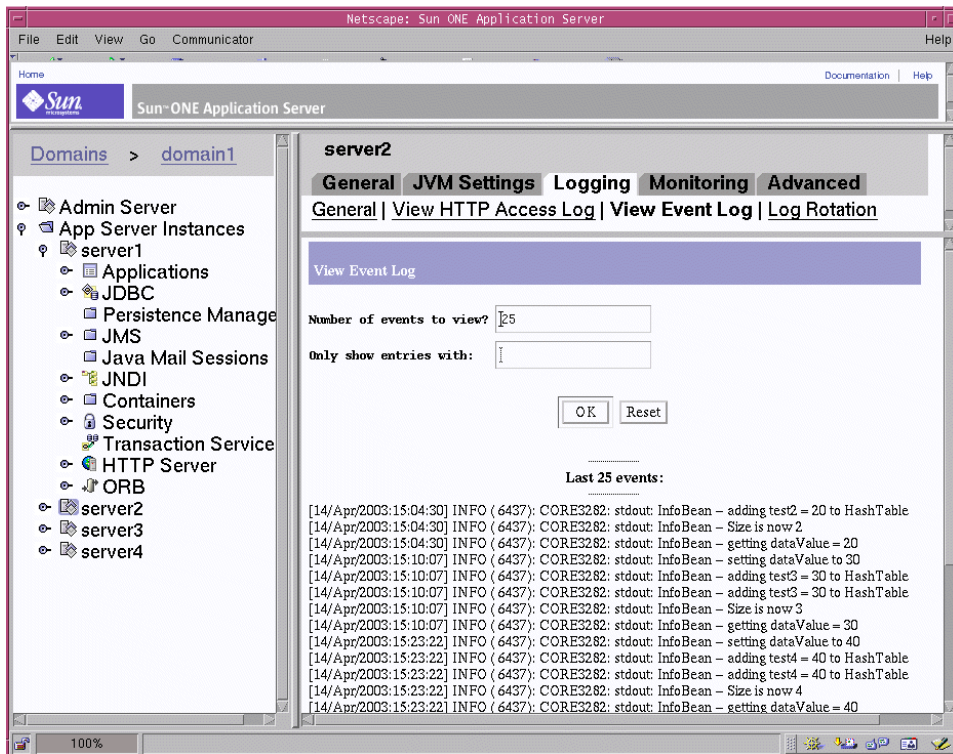
This section contains the following topics:

- [Using the Administration Interface to View Logs](#)
- [Viewing Logs Using the tail Command](#)
- [Application-Generated Messages in the Event Log](#)
- [Application-Generated Messages in the Access Log](#)

Using the Administration Interface to View Logs

To use the Administration interface to view server instance log files:

1. Access the Administration interface.
2. In the left pane, click the instance for which you want to check logs.
3. Click the Logging tab.
4. Click View Event Log.

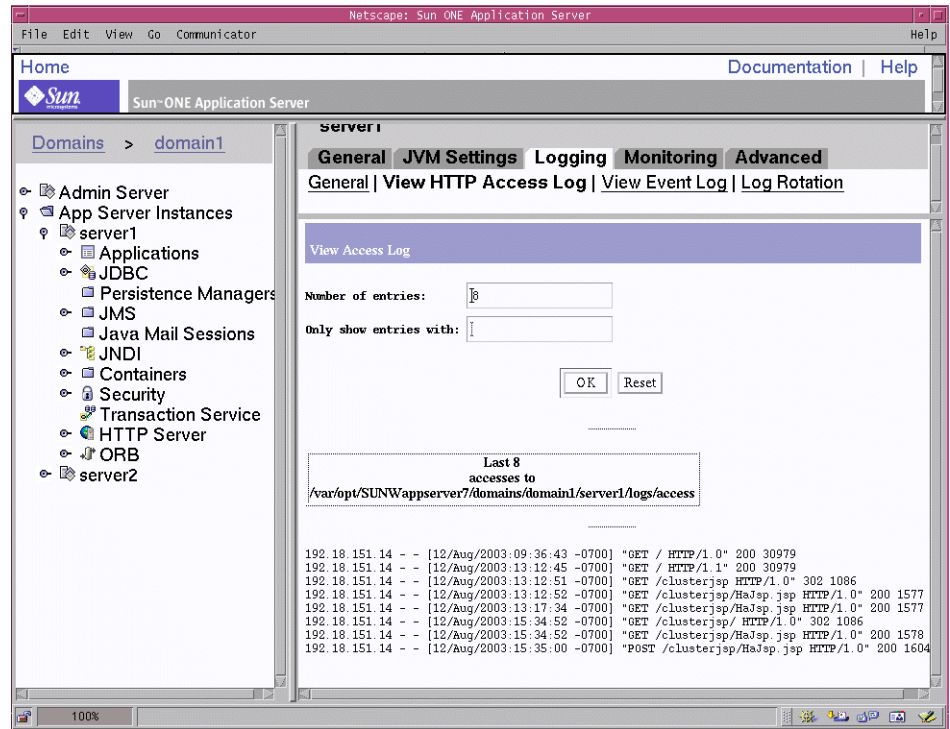
Figure 3-2 View Event Log

You can refresh the log view by clicking OK.

If you would like to see more than 25 log entries, simply enter a larger number in the “Number of errors to view?” area and click OK to refresh the log display.

To use the Administration interface to view the HTTP access log for an application server instance:

1. Access the Administration interface.
2. In the left pane, click the instance for which you want to check logs.
3. Click the Logging tab.
4. Click View HTTP Access Log.

Figure 3-3 View HTTP Access Log

The HTTP access log for a server instance shows the accesses to the sample application made for the server instance.

The name of the access log is `access`. By default, the access log file is located in the same directory as the server event log:

`domain_config_dir/domain1/server1/logs/`

Viewing Logs Using the tail Command

Using the `tail -f` command to monitor log files lets you see messages automatically as they are logged. In the Administration interface you must click a button to reflect new information on the page.

To use the `tail -f` command, follow these steps:

1. Navigate to the log directory of the server instance you want to monitor, for example:

```
cd domain_config/domain1/server1/logs/
```

2. Run `tail` on the server event log file:

```
tail -f server.log
```

The `-f` option leaves the `tail` command running so new log entries are displayed as they are written to the file.

Application-Generated Messages in the Event Log

When your application writes information to `stdout` and/or `stderr`, by default, this information is recorded in the server instance's event log as INFO-level messages with the message ID `CORE3282 (stdout)` and `CORE3283 (stderr)`. While running the Cluster JSP sample, a number of application generated messages are written the server event log. The following excerpt provides a snapshot of some of these messages:

For example:

```
[13/Aug/2003:17:32:28] INFO ( 9657): CORE3282: stdout: No parameter entered for this request
```

```
[13/Aug/2003:17:32:34] INFO ( 9657): CORE3282: stdout: Add to session: name1 = 1
```

```
[13/Aug/2003:17:32:45] INFO ( 9657): CORE3282: stdout: Add to session: name2 = 2
```

```
[13/Aug/2003:17:32:52] INFO ( 9657): CORE3282: stdout: Add to session: name3 = 3
```

These messages show the changes in values as data is entered into the application's user interface.

Application-Generated Messages in the Access Log

When you run the Cluster JSP sample, access log messages are logged to the application server instance that served the request:

```
192.18.151.14 - - [12/Aug/2003:15:34:52 -0700] "GET /clusterjsp/HTTP/1.0" 302 1086
```

```
192.18.151.14 - - [12/Aug/2003:15:34:52 -0700] "GET /clusterjsp/HaJsp.jsp HTTP/1.0" 200 1578
```

Troubleshooting Deployment to an Application Server Instance

The most common problems when attempting to run this sample application are described in the following table. The left column shows the symptom, the middle column the probable cause of the problem, and the right column shows suggested solutions to try.

Table 3-2 Troubleshooting Deployment to Application Server Instance

Symptom	Probable cause	Solution
Connection failure when attempting to access first page.	Application server not started. Wrong port specified in URL.	Ensure that the application server has been started. Determine the correct HTTP server port number. See “Starting the Server” on page 29 for more details.
404 error when accessing main page.	Application not deployed.	See “Deploying the Cluster JSP Sample Application to a Cluster” on page 48 .

When troubleshooting problems, it is very important that you monitor the application server log file. You may also find it useful to review the HTTP access log file to verify that HTTP requests are arriving at the application server as expected.

Adding the Sample Application to the Cluster

Once you have deployed the sample application to the two server instances in your cluster, you must add the application to the cluster you created in `loadbalancer.xml`:

1. Go to your web server’s `config` directory and open `loadbalancer.xml` using a text editor.

2. The sample `loadbalancer.xml` file contains the following line:

```
<web-module context-root="/abc" enabled="true"
disable-timeout-in-minutes="60" enabled="true" />
```

Change this line to reflect the `clusterjsp` sample application:

```
<web-module context-root="clusterjsp" enabled="true"
disable-timeout-in-minutes="60"/>
```

This line adds the `clusterjsp` application to the cluster and enables it.

The context root is the same value as is set for the application in the Sun ONE Application Server instance's `server.xml` file.

The `disable-timeout-in-minutes` attribute is set to 60. If the application is disabled, there is a 60-minute application quiescing period for the server to finish servicing any outstanding requests for the application. Earlier we set the instance-level `disable-timeout-in-minutes`, which controls the quiescing period for a server instance.

3. Save your changes.

Your `loadbalancer.xml` file should now be similar to the following file:

Code Example 3-1 The `loadbalancer.xml` File with Sample Application Configured

```
<!DOCTYPE loadbalancer PUBLIC "-//Sun Microsystems Inc.//DTD Sun ONE
Application Server 7.0//EN" "sun-loadbalancer_1_0.dtd">

<loadbalancer>

    <cluster name="cluster1">

        <instance name="server1" enabled="true"
disable-timeout-in-minutes="5" listeners="http://test.sun.com:81"/>

        <instance name="server2" enabled="true"
disable-timeout-in-minutes="5" listeners="http://test.sun.com:82"/>

        <web-module context-root="clusterjsp" enabled="true"
disable-timeout-in-minutes="60"/>

        <health-checker url="/" interval-in-seconds="10"
timeout-in-seconds="30" />

    </cluster>

    <property name="reload-poll-interval-in-seconds" value="60"/>

    <property name="response-timeout-in-seconds" value="30"/>

    <property name="https-routing" value="true"/>
```

```
<property name="require-monitor-data" value="true"/>
</loadbalancer>
```

Applying Configuration Changes and Restarting the Web Server

After you have completed your `loadbalancer.xml` edits, you need to start your web server, or if your web server is already running, you need to apply your configuration changes and restart the web server.

To start your web server:

1. Go to `web_server_install_dir/https-instance_name`.

For example:

```
/usr/iplanet/servers/https-test.sun.com
```

2. Type `./start` to start the server

You may also need to start the web server's administration server, located at `web_server_install_dir/https-admin-serv`.

To apply changes to a running web server and restart it:

1. Access the Server Manager for the web server by typing the appropriate URL into your web browser.

```
http://web_server:web_server_admin_port
```

For example:

```
http://test.sun.com:8888
```

2. Choose your server from the server manager and click Manage.

The Server Manager for the web server instance is displayed.

3. A message warns that you have made manual changes to your configuration files and you have to apply changes.
4. Click the Apply link in the upper right of the page.
5. Click Apply Changes to apply the changes and restart the server.

For subsequent `loadbalancer.xml` changes, you do not need to apply changes and restart the server. All subsequent changes are loaded automatically because the reload poll interval is set. For more information, see [“Dynamic Reconfiguration using the Reload Poll Interval” on page 44](#).

Running the Application

After you have deployed the application to all instances and updated the `loadbalancer.xml` file, test the application on the web server to make sure it works.

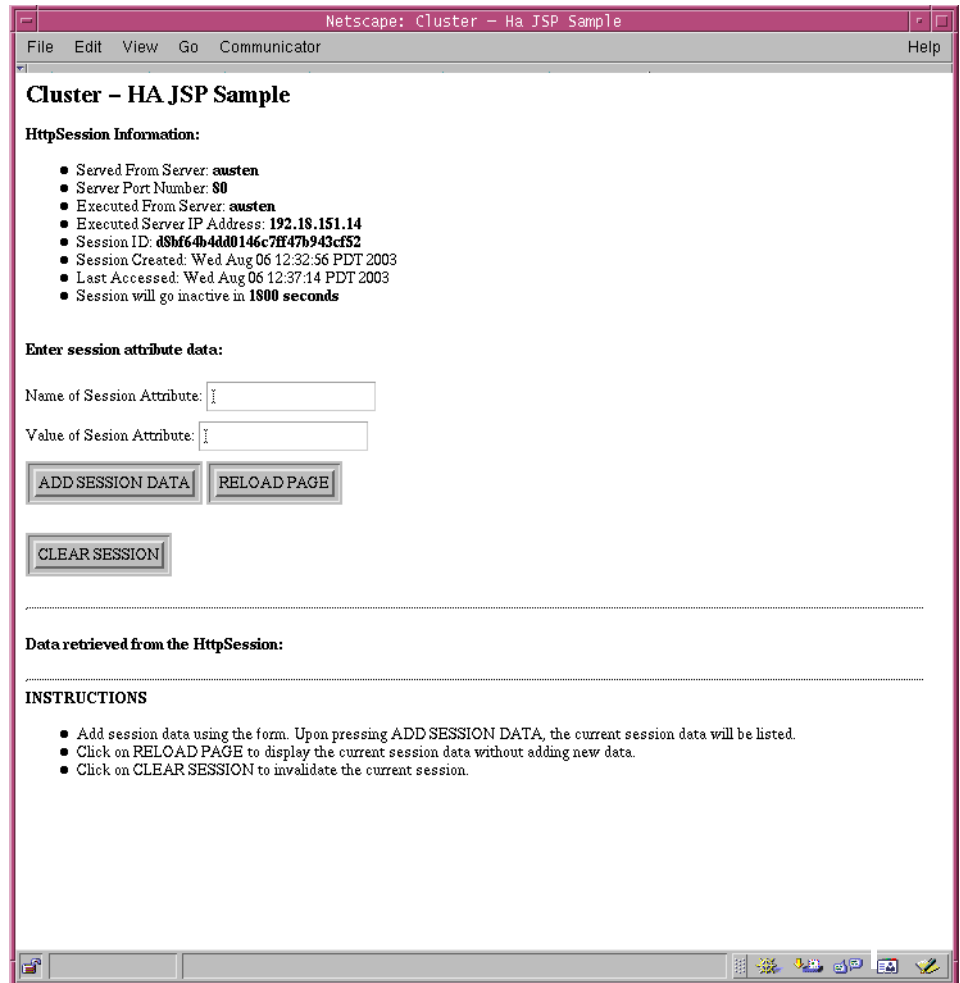
1. From a browser, access the following URL:

```
http://host:web_server_port/clusterjsp
```

For example:

```
http://test.sun.com:80/clusterjsp
```

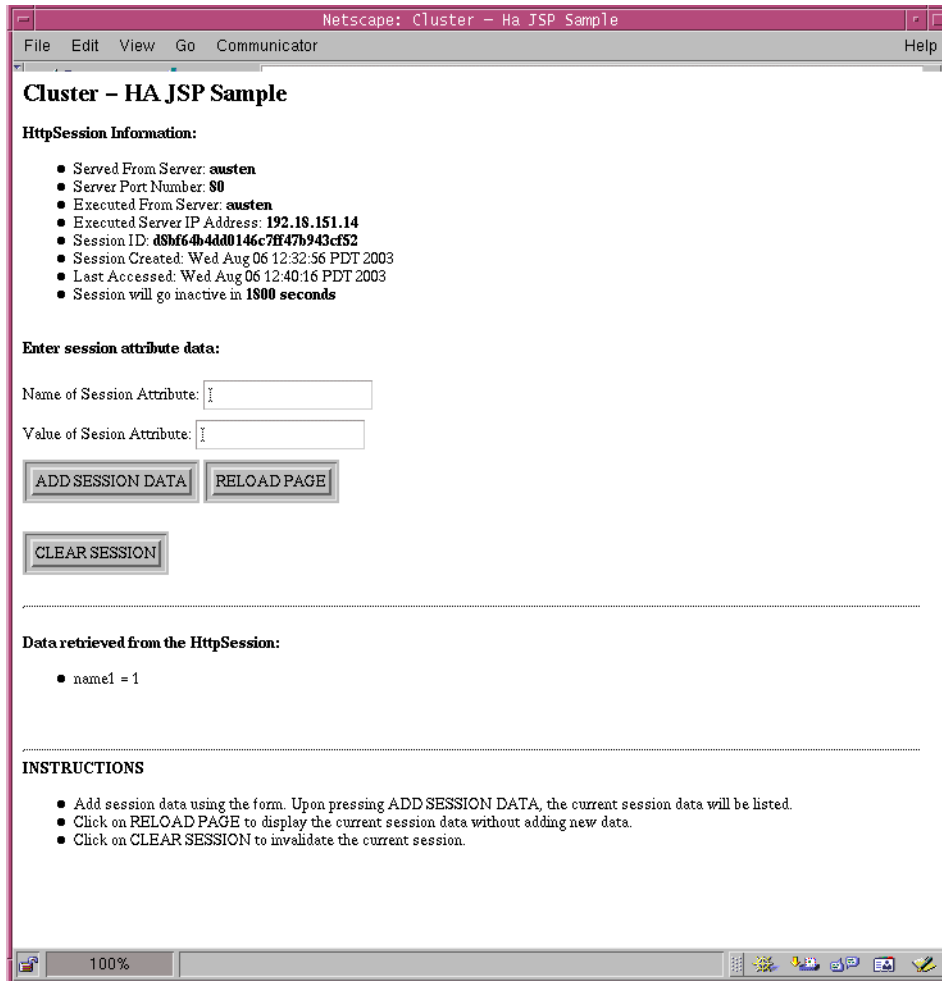
You see the page for this sample application.

Figure 3-4 Cluster JSP Sample Application Page

Your server information is listed at the top of the page, along with a unique Session ID. If you are accessing this application on the application server instance directly, the port information at the top of the page is the application server information. If you are accessing a clustered application through the web server, the port information is the web server information.

2. Enter a name and value for the session attribute and click Add Session Data. Your session data is displayed.

Figure 3-5 Session Data Displayed in Cluster JSP



Notice the Session ID at the top of the page. If you entered the session attributes and clicked ADD SESSION DATA before the session timed out, the Session ID at the top of the page will be the same as the first access.

In the example above, the Session ID d8bf64b4dd0146c7ff47b943cf52 is the same in both cases.

If the session timed out, the Session ID will be different.

3. Enter a new session attribute name and value. If the session has not timed out, all attributes are listed in the Data Retrieved from HttpSession area.

For example, if you enter session attribute data twice more before the session timed out, a page similar to the following is displayed:

Figure 3-6 Cluster JSP Sample Application with Multiple HTTP Session Attributes

Cluster - HA JSP Sample

HttpSession Information:

- Served From Server: **austen**
- Server Port Number: **80**
- Executed From Server: **austen**
- Executed Server IP Address: **192.18.151.14**
- Session ID: **d8hf64b4dd0146c7ff47b943cf52**
- Session Created: Wed Aug 06 12:32:56 PDT 2003
- Last Accessed: Wed Aug 06 12:42:49 PDT 2003
- Session will go inactive in **1800 seconds**

Enter session attribute data:

Name of Session Attribute:

Value of Session Attribute:

Data retrieved from the HttpSession:

- name2 = 2
- name3 = 3
- name1 = 1

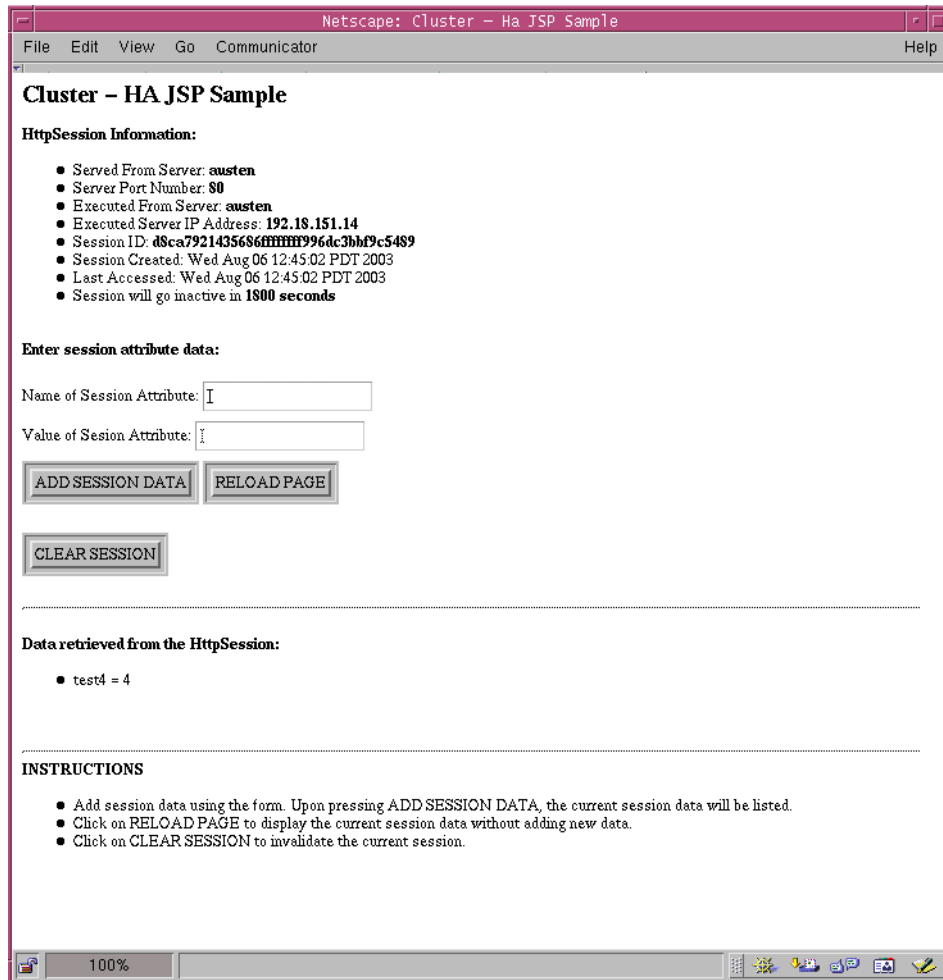
INSTRUCTIONS

- Add session data using the form. Upon pressing ADD SESSION DATA, the current session data will be listed.
- Click on RELOAD PAGE to display the current session data without adding new data.
- Click on CLEAR SESSION to invalidate the current session.

4. If you wait until the session times out (1800 seconds) and enter new session attribute data, you see only the latest data you entered, and a new session ID at the top of the page.

You can also simulate a timeout by clicking **CLEAR SESSION**, which will clear your session data and give you a new session ID.

Figure 3-7 Cluster JSP with New HTTP Session Information



Notice a new Session ID (in this example, d8ca7921435686ffffff996dc3bbf9c5489) appears at the top of the page and that only the latest information entered (in this example, test4=4) appears in the Data Retrieved from HttpSession section.

Verifying Load Balancing

Now that you have run the sample application and are familiar with how the session information is presented, you can demonstrate that your load balancer is working. If not, you can troubleshoot your load balancer.

This section contains the following topics:

- [Steps for Verifying Load Balancing](#)
- [Troubleshooting the Load-Balancer Plug-in](#)

Steps for Verifying Load Balancing

You can demonstrate that the load balancer is working by following these steps.

1. Open two separate browsers and run the application from your web server by typing `http://web_server_name:web_server_port/clusterjsp`. For example:

```
http://test.sun.com:80/clusterjsp
```

Because session data is sticky when you use the load balancer, if you open a browser window and access the application, all reloads or other operations from that browser are considered to be in the same session and will be serviced by a single Sun ONE Application Server instance.

In order to verify that load balancing is working for the application, open a second browser and access the application. Because the session data is stored in cookies, you must either open the second browser on a different machine, or use different browser software on the same machine. Check the Session ID to make sure both browser windows have separate session data.

2. Once you have accessed the application in two separate session windows, open another browser window and look at the web server error log.
 - a. Access your Sun ONE Web Server's Administration Server with the URL described in your web server documentation, for example:

```
http://test.sun.com:8888
```

- b. Choose your web server instance and click Manage.

- c. In the Server Manager for your web server instance, click the Logs tab.
 - d. In the left pane, click View Error Logs.
- 3. Look for entries with RequestExit in them.

You can either scan for them visually or enter RequestExit into the “Only show entries with” field.
- 4. The entries with RequestExit should show HTTP listener IDs for both Sun ONE Application Server instances, so you can see that both application server instances are serving requests.
- 5. To verify that the Sun ONE Application Server instance serviced the request, check the HTTP access logs for each application server instance.

You should see the requests for the `clusterjsp` application. For more information on checking the access logs, see [“Using the Administration Interface to View Logs” on page 55](#).

NOTE	In order to see the RequestExit messages, you must have enabled verbose logging for your web server. For instructions, see “Enabling Load Balancer Monitoring” on page 42 .
-------------	---

If you are having trouble getting your load balancer to work, see the following section on troubleshooting your load-balancer plug-in.

Troubleshooting the Load-Balancer Plug-in

Use the following table to help you troubleshoot load balancing. The symptom is described in the left column, the probable cause in the middle column, and the solution in the right column.

Table 3-3 Troubleshooting Load Balancing

Symptom	Probable cause	Solution
Connection failure when attempting to access application	Web server is not started Wrong port specified in URL.	Ensure that the web server has been started. Determine the correct web server port number. See your web server documentation for more details.
404 error when accessing application	Application is not deployed. Error in your <code>loadbalancer.xml</code> file	See “Deploying the Cluster JSP Sample Application to a Cluster” on page 48. Errors in your <code>loadbalancer.xml</code> file are noted in the web server’s error log. For more information, see “Finding Errors in Your loadbalancer.xml File” on page 69.
Error page when accessing application	One or more of the servers in the cluster is not responding.	Make sure that the server instances in your cluster are running. Check your error log to see if the health checker has found an unhealthy instance. For more information, see “Using the Health Checker” on page 70.

Finding Errors in Your `loadbalancer.xml` File

If you cannot reach the sample application through the web server URL, you might have a problem with your load balancer. If you can reach the applications through the individual application servers, you can determine that the problem is not with the application server instances.

Because the load balancer plug-in logs messages to the web server log files, you should check the web server error log files for more information.

You can view the error logs for the Sun ONE Web Server through the Server Manager:

1. Access your Sun ONE Web Server's Administration Server with the URL described in your web server documentation, for example:
`http://test.sun.com:8888`
2. Choose your web server instance and click Manage.
3. In the Server Manager for your web server instance, click the Logs tab.
4. In the left pane, click View Error Logs.
5. The default view shows 25 messages; you can increase the number if necessary.

If you are experiencing problems with the load balancer, you see messages that include "Failed to initialize load balancing subsystem."

If the problem is that the your `loadbalancer.xml` file has errors, you may see a parser error that points you to the line in your `loadbalancer.xml` file that contains the error. For example, you might see something like:

```
[17/Jun/2003:09:57:44] catastrophe ( 5643): LBConfigParser.cpp@434:
reports: lb.configurator: CNFG1000: Parsing of file
:/usr/iplanet/servers/https-test.sun.com/config/loadbalancer.xml

Failed at line : 7 and column : 3.The error message is : No character
data is allowed by content model
```

This message points you to the line with the error in the `loadbalancer.xml` file.

Correct the problem in the `loadbalancer.xml` file, apply changes and restart the web server, and try again.

Using the Health Checker

If your load balancer is running correctly, and you have set your web server's LogVerbose to on, you see the health checker working in the web server's log files:

```
[09/Apr/2003:13:36:02] verbose ( 7576): HealthChecker.cpp@153:
reports: lb.monitor: HLCK1006:UnhealthyInstances cluster1
1049920562631 NoUnhealthyInstances
```

For instructions about turning on verbose logging, see ["Enabling Load Balancer Monitoring" on page 42](#).

Regardless of whether you are using verbose logging, if one of your instances becomes unhealthy, the load balancer will flag it as unhealthy when an access attempt fails. If it remains unhealthy, the health-checker flags it as unhealthy again.

For example:

```
[17/Jun/2003:10:37:27] warning ( 5700): reports: lb.runtime:
RNTM2024: Daemon http://test.sun.com:81 is unhealthy.
```

```
[17/Jun/2003:10:37:27] warning ( 5700): reports: lb.healthchecker:
HLCK3003: Listener: http://test.sun.com:81 is detected to be still
unHealthy in cluster: cluster1
```

The health checker continues to monitor the unhealthy instance and flags it as healthy again once it is running.

Verifying HTTP Session Persistence

Once you have verified that load balancing works, you can verify HTTP session persistence. In session persistence, if one Sun ONE Application Server instance fails for some reason, the second instance will pick up the session and process the request.

NOTE	Session persistence does not work between clusters, only between instances in a cluster.
-------------	--

To demonstrate that HTTP session persistence is working:

1. Open a browser window.
2. Access the application from the web server. For example:

```
http://test.sun.com:80/clusterjsp
```
3. View the web server error log to see which server serviced the request (again, look for `RequestExit` entries).
4. Shut down the Sun ONE Application Server instance that is serving the page.

You can use the Administration interface to shut it down:

- a. Access the Administration interface using the administration port number, for example:

```
http://test.sun.com:4848
```
- b. In the left pane, click the server instance you want to stop.
- c. Click Stop in the right pane.

5. Reload the Cluster JSP sample application page.
The session ID and session attribute data is retained.
6. Check the web server error log. Notice that the other Application Server instance is now servicing the request.

Session information should be saved after a server goes offline. The only time you might lose session data during a failure is if the failure occurs while the application is transferring data. In that case, you may see slightly older session data.

Quiescing a Server Instance

In the real world, you would not shut down a server without attempting to have it complete existing sessions. Instead, you would shut down the server in a phased manner, in a process called quiescing.

To quiesce an application server instance, follow these steps:

1. Open a browser window and access the Cluster JSP application through the web server port. Keep this window open.
2. Determine which application server instance is serving the request.

Open a second browser window and access the server1 application server instance HTTP access log in the Administration interface. Look in the log for an access of the Cluster JSP application timed when you sent the request. If you see the access in the server1 log, then the server1 application server instance is serving the request. If you do not see such an entry in the access log, check the access log for the server2 application server instance. Leave the access log window open.

3. Once you have verified the application server instance that is serving the request, disable that server instance in the `loadbalancer.xml` file.

Change `enabled="true"` to `enabled="false"` for the correct server instance. For example:

```
<instance name="server1" enabled="false"
disable-timeout-in-minutes="5"
listeners="http://test.sun.com:81"/>
```

4. Save your `loadbalancer.xml` changes.

5. Wait for the reload poll interval to pass, so that your new configuration is loaded.

For this tutorial, the reload poll interval is set to 60 seconds, so you only need to wait a minute.

6. Once the reload poll interval has passed and your new configuration is loaded, the load balancer plug-in no longer sends requests that do not have existing sessions to that server instance. However, requests with existing sessions are still forwarded to the server instance for the quiescing period.

To see this, in your open Cluster JSP window, enter name and attribute information and click ADD SESSION DATA

7. Go to the browser window that displays the HTTP access log for the server that served the initial request. Click OK to refresh the information.

Notice that even though the server is disabled in `loadbalancer.xml`, it is serving this request.

8. Wait five minutes for the quiescing period to end, then use your open Cluster JSP application window to enter additional name attribute information and click ADD SESSION DATA.

The quiescing period is controlled by the disable timeout in minutes for the server instance. Earlier in the tutorial we set this to five minutes.

9. Because the quiescing period has passed, the request is sent to a different application server instance.

To see this result, refresh your server instance's HTTP access log. You should see no new requests.

10. Open a third browser window and view the HTTP access log for the other server instance. You should see the Cluster JSP request being serviced by the second server instance.

You can then shut down the application server instance safely.

You can also quiesce an application to take it offline safely. For more information on application server instance and application quiescing, see the *Sun ONE Application Server Administrator's Guide*.

Summary and Next Steps

By following the *Getting Started Guide*, you explored these key aspects of the Sun ONE Application Server:

- Key features and architecture
- Using tools for administering the application server
- Deploying and using a J2EE application
- Configuring and using load balancing
- Configuring and using HTTP session persistence
- Quiescing a server instance

Review the next steps in the following table to determine the resources that are best suited to your interests. The left column contains topics of interest and the right column contains information on the next steps to take to pursue your interest.

Table 4-1 Next Steps

Interest	Next Steps
Learn more about Enterprise Edition features implemented in Sun ONE Application Server sample applications.	See the sample applications for Enterprise Edition installed with the Application Server at http://application_server_host_name:port/samples . An index page provides information on all the available Enterprise Edition samples.
Learn about J2EE design and development best practices.	See the Java BluePrints and try out the Java Pet Store and Smart Ticket sample applications that are bundled with the application server.
Learn more about failover, session persistence, and high-availability session configuration in applications.	See the <i>Sun ONE Application Server Application Guidelines for Storing Session State</i> and use the Duke's Bookstore and Session Storage sample applications that are bundled with the application server.

Table 4-1 Next Steps

Interest	Next Steps
Learn about administering Sun ONE Application Server 7, including clustering, session persistence, load balancing, and the HADB.	Review the <i>Sun ONE Application Server 7 Administrator's Guide</i> .
Learn more about J2EE by using tutorials.	See the J2EE Tutorial and Java Web Services Tutorials.
Learn how specific J2EE features are implemented on Sun ONE Application Server 7.	See the Sample Applications included in the application server installation.
Learn about general development practices using Sun ONE Application Server 7.	See the <i>Sun ONE Application Server 7 Developer's Guide</i> .

Index

A

- access log [58](#)
- Admin Server [14](#)
 - event log [33](#)
 - password [32](#)
 - port [32](#)
 - username [32](#)
 - verifying start-up [31](#)
- Administration interface [22](#)
 - accessing [31](#)
 - viewing logs [55](#)
- administrative domains [16](#)
- application server
 - starting [29](#)
 - verifying start-up [31](#)
- application server instances [13](#)
 - adding to cluster [38](#)
 - quiescing [72](#)
 - verifying start-up [34](#)
- applications, sample
 - adding to cluster [59](#)
 - copying [48](#)
 - monitoring [58](#)
 - quiescing [60](#)
 - running [53](#)
 - troubleshooting [59](#)
- apply changes to web server [61](#)
- asadmin utility [22, 51](#)
 - configure-session-persistence command [51](#)
 - create-jdbc-connection-pool command [51](#)
 - create-jdbc-resource command [51](#)
 - delete-jdbc-connection-pool command [52](#)

- delete-jdbc-resource command [52](#)
- deploy command [51](#)
- list-instances command [34, 52](#)
- PATH variable set-up [29](#)
- start-instance command [51](#)
- stop-instance command [51](#)

C

- cladmin command [22, 49](#)
 - deploy command [50](#)
 - limitations [52](#)
 - location [49](#)
 - log file [51](#)
 - PATH variable set-up [29](#)
 - requirements [52](#)
 - start-instance [53](#)
 - supported asadmin commands [51](#)
 - syntax [50](#)
- cladmin.log file [51](#)
- clinstance.conf file [49](#)
- clpassword.conf file [50](#)
- clsetup command [22, 27](#)
- Cluster JSP sample application [48](#)
 - running [62](#)
- clusters [16](#)
 - adding applications [59](#)
 - creating [38](#)
 - deploying applications [49](#)
 - example scenario [20](#)
- command-line utilities [22](#)

- configure-session-persistence command [51](#)
- create-jdbc-connection-pool command [51](#)
- create-jdbc-resource command [51](#)
- customer support [12](#)

D

- default_config_dir [12](#)
- delete-jdbc-connection-pool command [52](#)
- delete-jdbc-resource command [52](#)
- deploy command [51](#)
- deployment
 - sample applications to cluster [49](#)
 - troubleshooting [59](#)
 - verifying [53](#)
- directories [11](#)
- disable-timeout-in-minutes [39](#)
- distributable applications [49](#)
- documentation
 - directory conventions [11](#)
 - font conventions [11](#)
 - general conventions [11](#)
 - guide organization [8](#)
 - path formats [11](#)
 - URL formats [11](#)
 - using the [8](#)
- domains, administrative [16](#)

E

- error log, web server [44, 70](#)
- event log
 - Admin Server [33](#)
 - application messages in [58](#)
 - application server instance [34](#)

F

- file session persistence [18](#)
- font conventions [11](#)

H

- ha session persistence [18](#)
- HADB [19](#)
- health checker [41, 70](#)
 - interval-in-seconds [41](#)
 - timeout-in-seconds [41](#)
 - URL [41](#)
- highly-available session persistence [18](#)
- HTTP access log [56](#)
- HTTP listeners [39](#)
 - in RequestExit log file entries [68](#)
- HTTP server
 - listener [35](#)
 - port [35](#)
 - welcome page [36](#)
- http-listener element, server.xml file [35](#)
- HTTPS listeners [39, 44](#)
- HTTPS routing [44](#)

I

- install_config_dir [12](#)
- install_dir [11](#)
- installation requirements [26](#)
- interval-in-seconds [41](#)

L

- list-instances command [34, 52](#)
- load balancer [17, 42](#)
 - algorithm [17](#)
 - dynamic reconfiguration [44](#)
 - health checker [41, 70](#)

- HTTPS routing [44](#)
- properties [43](#)
- response timeout [44](#)
- sample loadbalancer.xml [45](#)
- troubleshooting [68](#)
- loadbalancer.xml file
 - creating [37](#)
 - errors in [69](#)
 - example [37](#)
 - sample [45](#)
- loadbalancer.xml.example file [37](#)
- logs
 - access [58](#)
 - enabling verbose logging in web server [43](#)
 - error, web server [44](#)
 - event, Admin Server [33](#)
 - event, application server instance [58](#)
 - HTTP access [56](#)
 - monitoring using tail command [34, 57](#)
 - viewing [55](#)
 - web server error [44](#)
- LogVerbose [43](#)

M

- memory session persistence [18](#)
- monitoring sample applications [58](#)
- monitoring, load balancer [42](#)

P

- password, Admin Server [32](#)
- PATH environment variable, setting [29](#)
- path formats [11](#)
- ports
 - Admin Server [32](#)
 - HTTP server [35](#)
 - inaccessible [32](#)
 - web server [62](#)

Q

- quiescing
 - applications [60](#)
 - server instances [72](#)

R

- reload poll interval [44](#)
- response timeout [44](#)
- round robin load balancing [17](#)

S

- sample applications
 - adding to cluster [59](#)
 - Cluster JSP [48](#)
 - copying [48](#)
 - deploying to a cluster [49](#)
 - monitoring [58](#)
 - running [53, 62](#)
 - troubleshooting [59](#)
 - verifying deployment [53](#)
 - WAR files [48](#)
- server instances [13](#)
 - adding to cluster [38](#)
 - verifying start-up [34](#)
- session persistence [18](#)
 - configuring [19](#)
 - file [18](#)
 - ha [18](#)
 - memory [18](#)
 - verifying [71](#)
- single sign-on, availability of session information [19](#)
- start-domain command [29](#)
- starting application server instances [53](#)
- start-instance command [51, 53](#)
- stderr [58](#)
- stdout [58](#)
- sticky round robin [17](#)
- stop-instance command [51](#)

sun-loadbalancer.dtd file [38](#)

T

tail command [34](#), [57](#)

timeout-in-seconds [41](#)

troubleshooting

- load balancer [68](#)

- sample applications [59](#)

tutorials

- Cluster JSP sample [48](#)

- enterprise features configuration [29](#)

- preparing to use [25](#)

- steps [27](#)

U

undeploy command [51](#)

URL formats [11](#)

username, Admin Server [32](#)

V

verbose logging [43](#)

verifying

- application deployment [53](#)

- server instance start-up [34](#)

W

web server

- Apache [26](#)

- applying changes [61](#)

- configuring logging [42](#)

- error log [44](#)

web server error log [70](#)

welcome page, HTTP server [36](#)