

Oracle® WebCenter Sites

Installation Guide for the Site Capture Application

11g Release 1 (11.1.1)

April 2012

Oracle® WebCenter Sites: Installation Guide for the Site Capture Application, 11g Release 1 (11.1.1)

Copyright © 2012 Oracle and/or its affiliates. All rights reserved.

Primary Author: Tatiana Kolubayev

Contributor: Ankush Chadha, Amit Kumar, Gaurang Mavadiya

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation shall be subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License (December 2007). Oracle America, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information on content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

Table of Contents

Audience	5
Related Documents	5
Conventions	6
Steps in This Guide	6
Third-Party Libraries	6
1 Welcome	7
Introducing Oracle WebCenter Sites: Site Capture	8
Procedures in This Guide	8
Before You Begin	8
Installation Worksheet	12
Next Step	12
2 Configuring the Application Server	13
Configuration Options	14
Configuring Tomcat Application Server	14
Configuring WebLogic Application Server	16
Configuring WebSphere Application Server	17
3 Installing Oracle WebCenter Sites: Site Capture	19
Installation Steps	20
Running the Silent Installer	20
Running the GUI Installer	21
Deploying Site Capture	30
Deploying on Tomcat Application Server	30
Deploying on WebLogic Application Server	30
Deploying on WebSphere Application Server	31
Post-Installation Steps	32
If Site Capture Is Using an Oracle Database	32
Verifying the Site Capture Application	32

Authorizing Users to Work with Site Capture.	33
Next Steps	33
4 Enabling Publishing-Triggered Site Capture.	35
Integrating Site Capture with Oracle WebCenter Sites' Publishing Process	36
Next Steps	37
A Site Capture File System.	39
General Directory Structure	40
Custom Folders	42

About This Guide

This guide describes the process of installing Oracle WebCenter Sites: Site Capture, a web application that is used to download dynamically published websites.

Applications discussed in this guide are former FatWire products. Naming conventions are the following:

- *Oracle WebCenter Sites* is the current name of the application previously known as *FatWire Content Server*. In this guide, *Oracle WebCenter Sites* is also called *WebCenter Sites*.
- *Oracle WebCenter Sites: Site Capture* is the current name of the application previously known as *FatWire Site Capture*. In this guide, *Oracle WebCenter Sites: Site Capture* is also called *Site Capture*.
- *Oracle WebCenter Sites: Web Experience Management Framework* is the current name of the environment previously known as *FatWire Web Experience Management Framework*. In this guide, the environment is also called *Web Experience Management Framework*, or *WEM Framework*.

The Site Capture application integrates with Oracle WebCenter Sites according to specifications in the *Oracle WebCenter Sites 11g Release 1 (11.1.1.x) Certification Matrix*. For additional information, see the release notes for WebCenter Sites. Check the WebCenter Sites documentation site regularly for updates to the *Certification Matrix* and release notes.

Audience

Users of this guide must have experience installing and configuring enterprise-level software, such as application servers and databases. Also required is a general administrator's experience with WebCenter Sites and the WEM Framework

Related Documents

For more information, see the following documents:

- *Oracle WebCenter Sites Administrator's Guide for the Site Capture Application*
- *Oracle WebCenter Sites Administrator's Guide*

Conventions

The following text conventions are used in this guide:

- **Boldface** type indicates graphical user interface elements that you select.
- *Italic* type indicates book titles, emphasis, or variables for which you supply particular values.
- `Monospace` type indicates file names, URLs, sample code, or text that appears on the screen.
- **Monospace bold** type indicates a command.

Steps in This Guide

Some of the steps in this guide are written as “quick steps” to provide you with a quick reference on completing various operations. For example, to view the crawler’s list of archived sites, you would see the following quick steps:

Crawlers > *crawlerName* > Archives

The steps above mean:

Go to the “Crawlers” home page, point to a crawler, and select **Archives** from the pop-up menu.

When concepts, features, and the associated operations require explanation, the steps are written in detail.

Third-Party Libraries

Oracle WebCenter Sites and its applications include third-party libraries. For additional information, see *Oracle WebCenter Sites 11gR1: Third-Party Licenses*.

Chapter 1

Welcome

This chapter presents an overview of the Site Capture application and the installation process.

This chapter contains the following topics:

- [Introducing Oracle WebCenter Sites: Site Capture](#)
- [Procedures in This Guide](#)
- [Before You Begin](#)
- [Installation Worksheet](#)
- [Next Step](#)

Introducing Oracle WebCenter Sites: Site Capture

Oracle WebCenter Sites: Site Capture is a web application that integrates with Oracle WebCenter Sites through the Oracle WebCenter Sites: Web Experience Management (WEM) Framework to capture dynamically published websites for evaluation, compliance purposes, high availability requirements, and other types of scenarios.

Crawls can be initiated manually from the Site Capture interface, or they can be triggered by the completion of a WebCenter Sites RealTime publishing session. In each scenario, the crawler captures the site in one of the following modes, depending on the user's selections:

- **Static mode:** The site is stored as files that are ready to be served. Only the latest capture is kept.
- **Archive mode:** The site is stored in a zip file. A pointer in the Site Capture database enables archive management from the Site Capture interface.

Procedures in This Guide

This guide contains procedures for installing and configuring Site Capture to support:

- Static and archive capture initiated manually, from the Site Capture interface.
- Static and archive capture triggered by the completion of a WebCenter Sites RealTime publishing session. Setting up publishing-triggered site capture is an option.
- Administrative users and developers. The Site Capture application is designed for general administrators of the WebCenter Sites system on which the Site Capture application will be running. Developers will write advanced crawler configuration code – for example, code that triggers Site Capture to execute a post-crawl command such as copying statically captured sites to a web server's doc base.

Before You Begin

Users of this guide must have experience installing and configuring enterprise-level software, such as application servers and databases. Also required is a general administrator's experience with WebCenter Sites and the WEM Admin interface.

- To complete the procedures in this guide, you must be a WebCenter Sites general administrator who belongs to the `RESTAdmin` security group.
- Download the *Oracle WebCenter Sites 11g Release 1 (11.1.1.x) Certification Matrix* for information about supported operating systems, application servers, databases, and browsers. Also, download the Oracle WebCenter Sites release notes for information about Site Capture.
- Read this guide to acquire an understanding of the Site Capture installation process. The basic steps are configuring the Site Capture application server, running the installer to create the Site Capture `war` file, deploying Site Capture, completing post-installation and verification steps, and if necessary, setting up publishing-triggered site capture.
- On all systems, set the `JAVA_HOME` variable to point to a valid installation of JDK 1.6.

- Prepare the Site Capture installation components:
 - Ensure you have a dedicated and fully functional Oracle WebCenter Sites installation running in development or content management mode. Site Capture must communicate with this WebCenter Sites system in order to run.
 - Ensure you have a dedicated host machine on which to install the Site Capture application server and the application itself. During the installation process, you will configure Site Capture to communicate with the WebCenter Sites system (described above), which runs on its own host machine.
 - Decide whether your Site Capture application will be running as a single application or in clustered mode. For a diagram of a single-server installation, see [Figure 1, on page 10](#). For a diagram of a clustered installation, see [Figure 2, on page 10](#).
 - Install or reuse the following components:
 - Install a supported Site Capture application server on the dedicated host machine. For clustered installations, install the application server for each Site Capture instance.
 - For a clustered installation, install a load balancer on a host machine of your choice. **The Site Capture installation directory must be a shared directory, accessible to all cluster members.**
 - To store archived sites, you can either reuse WebCenter Sites' database or install a dedicated, supported Site Capture database on a host machine of your choice.
- Decide whether you will be running publishing-triggered site capture. If so, you will need a WebCenter Sites source and target system:
 - The target system provides the REST API and WEM SSO API, which enable the target system to communicate with the Site Capture application at the end of the publishing session in order to start the required crawlers. The Site Capture application will then send crawler invocation status to the target WebCenter Sites system, which will, in turn, send the same information to the source WebCenter Sites system. Both WebCenter Sites systems record the status information in their own log files (`futuretense.txt`, by default).
 - You will integrate the source and target WebCenter Sites systems with your Site Capture installation. For some of the possible configurations, see [Figure 3, on page 11](#) and [Figure 4, on page 11](#).
- Decide whether to install sample crawlers (**recommended**). For more information about the sample crawlers, see the first note on [page 20](#).
- You have the option to install Site Capture silently or graphically. The silent installer provides help and sample values for every piece of information that needs to be set.

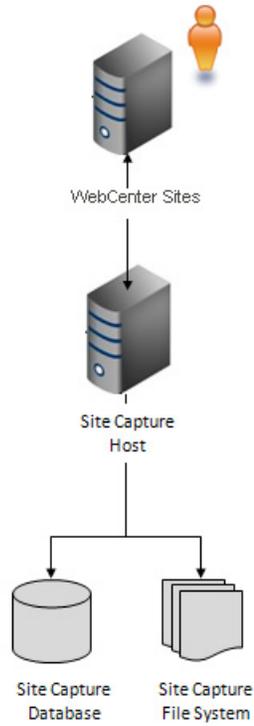
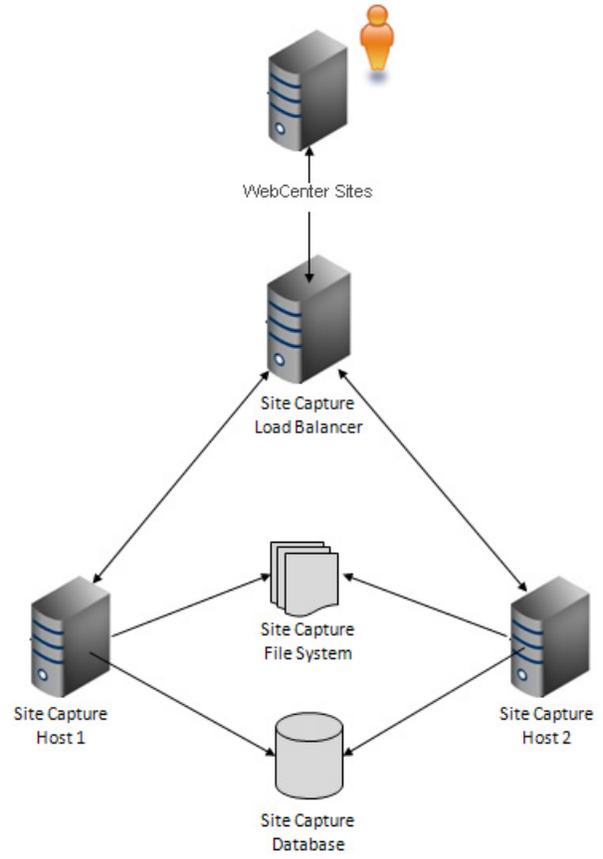
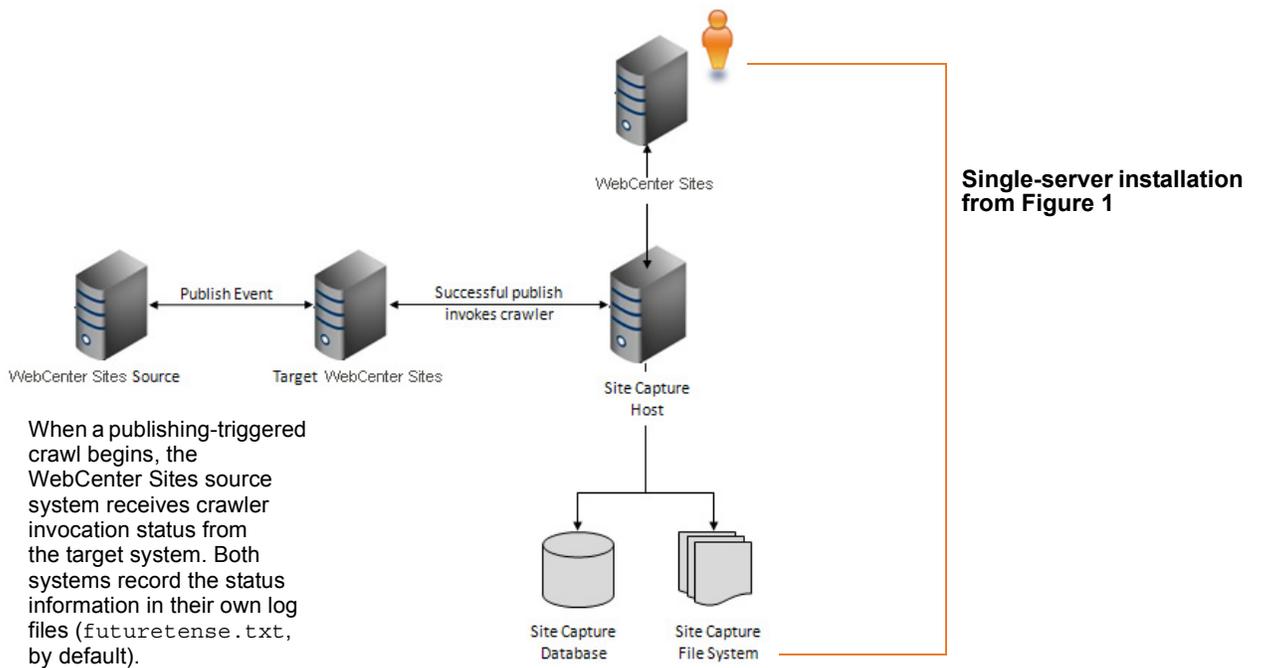
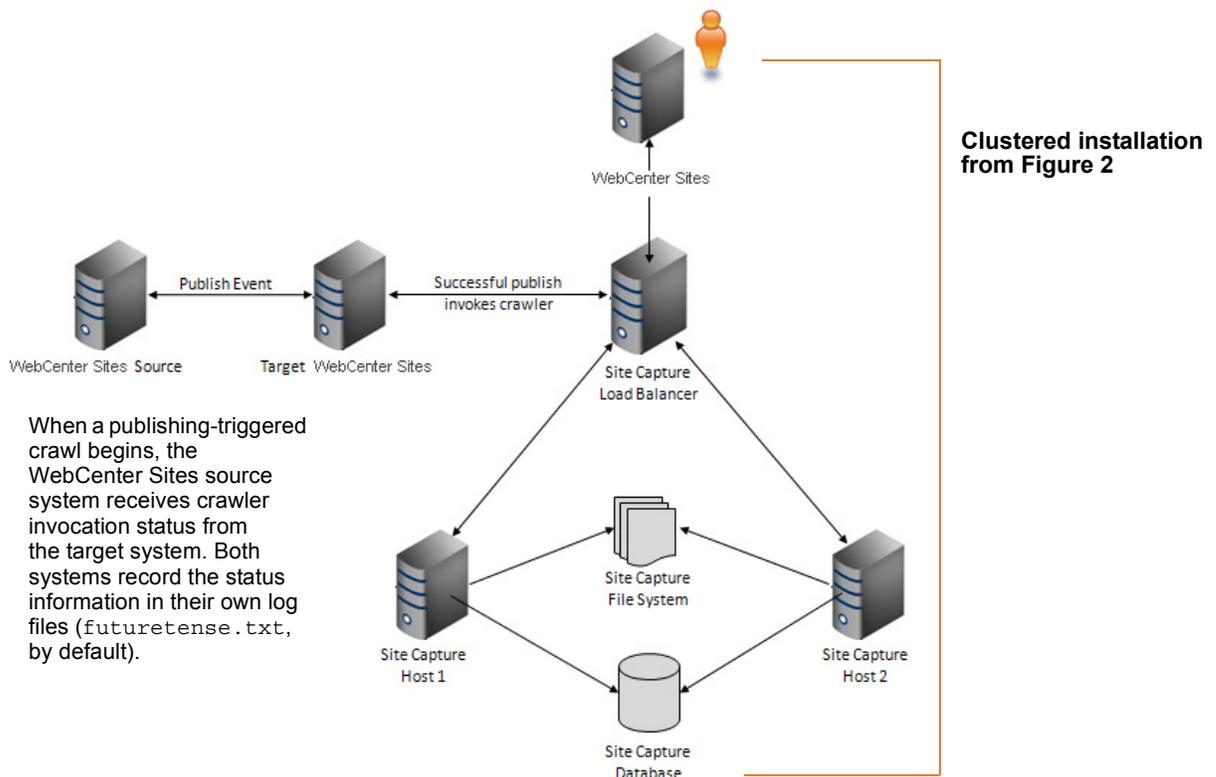
Figure 1: Single-Server Installation**Figure 2: Clustered Installation**

Figure 3: Single-Server Installation Enabled for Publishing-Triggered Capture**Figure 4:** Clustered Installation Enabled for Publishing-Triggered Capture

Installation Worksheet

When configuring the Site Capture application server(s), you will specify parameters that you will also reuse when setting up the installer to create the Site Capture `war` file. Record the parameters in the following table:

Installation Type	Installation Parameter
All types	WebCenter Sites' host name (or IP address). This is the WebCenter Sites on which Site Capture will run as an application:
	WebCenter Sites' port number:
	WebCenter Sites context root:
	WebCenter Sites general administrator's user name:
	Above user's password:
	CAS host name (or IP address):
	CAS port number:
	CAS context root:
Single-Server	Application server's host name (or IP address):
	Application server's port number:
	Application server's datasource name:
Clustered	Load balancer's host name (or IP address):
	Load balancer's port number:

Next Step

Continue to [Chapter 2, "Configuring the Application Server."](#)

Chapter 2

Configuring the Application Server

Site Capture can run as a single application or in a clustered installation.

This chapter contains the following topics:

- [Configuration Options](#)
- [Configuring Tomcat Application Server](#)
- [Configuring WebLogic Application Server](#)
- [Configuring WebSphere Application Server](#)

Configuration Options

In this chapter, you will configure an application server to support Site Capture running as a single instance or in cluster mode. Complete one of the following procedures:

- [Configuring Tomcat Application Server](#)
- [Configuring WebLogic Application Server](#)
- [Configuring WebSphere Application Server](#)

Note

The Site Capture application must run on context root / to support previewing of archives. (Site Capture serves pages from within archives, which can have any context root.)

Configuring Tomcat Application Server

If you are creating a clustered installation, complete the following steps on all cluster members.

To configure Tomcat

1. Open `server.xml` in `<apache-tomcat-6.0.29 directory>\conf` and do the following:
 - a. Replace port 8080 with the Site Capture port (on which the application server is listening):

Note

For clustered installations, specify a unique port number for each cluster member.

```
<Connector port="8080" protocol="HTTP/1.1"
  connectionTimeout="20000" redirectPort="8443" />
```

- b. Configure the data source for Site Capture.

Note

For clustered installations, configure the same data source on each cluster member.

Refer to [Table 1, on page 15](#) for information about configuring the driver class name and URL for your selected driver.

Below is a sample DB2 configuration:

```
<Context path="" docBase="ROOT" debug="5" reloadable="true"
  crossContext="true">
```

```

<Resource name="scDataSource"
  auth="Container"
  type="javax.sql.DataSource"
  url="jdbc:db2://<hostname>:<port>/CLUSTDB"
  driverClassName="com.ibm.db2.jcc.DB2Driver"
  username="xxxxx"
  password="xxxxxxx"
  maxActive="20"
  maxIdle="10"
  maxWait="-1" />
</Context>

```

Table 1: Database Driver Parameters

Database Driver	Parameter	Value
SQL server	DriverClass	net.sourceforge.jtds.jdbcx.JtdsDataSource
	Required .jar files	jtds-1.2.jar
	URL	jdbc:jtds:sqlserver://<server>:<dbport>/<dbname> Ex: jdbc:jtds:sqlserver://10.120.14.22:1433/SC70
DB2	DriverClass	com.ibm.db2.jcc.DB2Driver
	Required .jar files	db2jcc.jar, db2cc_license_cu.jar
	URL	jdbc:db2://<hostname>:<dbport>/<dbname> Ex: jdbc:db2://10.120.16.30:50000/SC70
Oracle	DriverClass	oracle.jdbc.driver.OracleDriver
	Required .jar files	Ojdbc6.jar
	URL	jdbc:oracle:thin:@//<hostname>:1521/<dbname> Ex: jdbc:oracle:thin:@//godzilla.fatwire.com:1521/SC70

2. Copy the required database driver jar file to <apache-tomcat-6.0.29 directory>\lib.
3. Once you have completed configuring the application server(s), continue to [Chapter 3, "Installing Oracle WebCenter Sites: Site Capture."](#)

Configuring WebLogic Application Server

If you are creating a clustered installation, complete the following steps on all cluster members.

To configure WebLogic

1. Create a domain in WebLogic and configure the administrator user name and password for the domain.
2. Start the WebLogic Admin Server from the command prompt:
 - Windows: `startweblogic.cmd`
 - UNIX: `./startweblogic.sh`
3. Log in to the console.
 - a. Create a managed server `<managed_server_name>` in the domain you just created.

Note

For clustered installations, specify a unique port number for each cluster member.

- b. Create a data source and map it to the managed server `<managed_server_name>`.

Note

For clustered installations, configure the same data source on each cluster member.

4. Once you have completed configuring the application server(s), continue to [Chapter 3](#), “[Installing Oracle WebCenter Sites: Site Capture](#).”

Configuring WebSphere Application Server

If you are creating a clustered installation, complete the following steps on all cluster members.

To configure WebSphere

1. Create a WebSphere application server instance.

Note

For clustered installations, specify a unique port number for each cluster member.

2. Create a data source for the newly created server instance.

Note

For clustered installations, configure the same data source on each cluster member.

- a. Copy the required database driver into `<websphere_install_dir>/websphere/AppServer/universalDriver/lib`. (refer to [Table 1](#), on [page 15](#) for the required driver).
 - b. Log in to the WebSphere console and do the following:
 - 1) Add the user name and password of the database by creating an alias (go to **Security > Global security > Java Authentication and Authorization Service > J2C authentication data**).
 - 2) Create a JDBC provider and data source (go to **Resources > JDBC**).
3. Once you have completed configuring the application server(s), continue to [Chapter 3](#), “[Installing Oracle WebCenter Sites: Site Capture](#).”

Chapter 3

Installing Oracle WebCenter Sites: Site Capture

Site Capture can be installed graphically or silently. You will first create the `war` file using the installer, then manually deploy the `war` file and test the installation.

This chapter contains the following topics:

- [Installation Steps](#)
- [Post-Installation Steps](#)
- [Next Steps](#)

Installation Steps

To install Site Capture, you will complete the following basic steps:

1. Run the silent or GUI installer to create the Site Capture war file (ROOT.war).
2. Deploy the Site Capture application.

Detailed steps are provided below.

Note

- If you are creating a clustered installation:
 - Run the installer once. You will deploy the ROOT.war file (and ROOT folder) on all cluster members, as shown in this chapter.
 - The Site Capture installation directory must be a shared directory, accessible to all other cluster members.
- We recommend installing the sample crawlers. Procedures for quickly testing and using the Site Capture installation are based on the **Sample** crawler, described below.
 - **Sample** is a basic crawler that can be easily configured to capture any dynamic site. The configuration step amounts to setting the crawler's start URI in the crawler's configuration file.
 - **FirstSiteII** is a crawler with advanced configuration code for capturing WebCenter Sites' FirstSiteII sample website. The code demonstrates the implementation of various methods and interfaces that are used to control the crawler's site capture process, in this example, downloading a dynamic site (FirstSiteII) as a static site.

Crawlers can be easily deleted should you no longer need them.

Running the Silent Installer

Note

If you wish to run the GUI installer instead of the silent installer, skip to [“Running the GUI Installer,” on page 21](#).

1. Unzip the sitecapture.zip file into the desired location on the server.
2. Configure the omii.ini file (located at the root level in the extracted folder) according to instructions provided in the file, and save the file.

Reminder

We recommend installing the sample crawlers. For information about the crawlers, see the first note on this page.

3. Open the `Install.ini` file (located at the root level in the extracted folder), add the line `loadfile=omii.ini` to the first section, and save the file.
4. Open a command prompt and run one of the following commands:
 - Windows: `scInstall.bat -silent`
 - Linux: `scInstall.sh -silent`You will see the following message in the command window after installation is successful – “Installation Finished Successfully”
5. Wait until the installer process exits normally and shows the command prompt again. The `ROOT` folder and the `ROOT.war` file will be created in the `webapps` folder of the installation directory specified in the `omii.ini` file.
6. Deploy the Site Capture `ROOT.war` file. For instructions, see one of the following sections:
 - [Deploying on Tomcat Application Server](#)
 - [Deploying on WebLogic Application Server](#)
 - [Deploying on WebSphere Application Server](#)

Running the GUI Installer

Note

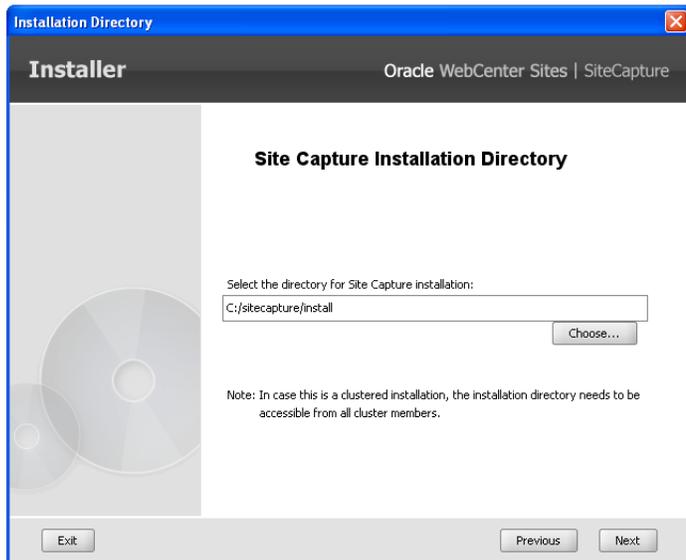
If you wish to run the silent installer instead of the GUI installer, go to “[Running the Silent Installer](#),” on page 20.

1. Download `sitecapture.zip` to a directory on your Site Capture server and extract it into a temporary directory. The extracted folder contains the `scInstall.bat` and `scInstall.sh` files, which are needed to run the Site Capture installer.
2. Execute the following installer script:
 - On Windows: `scInstall.bat`
 - On UNIX: `scInstall.sh`

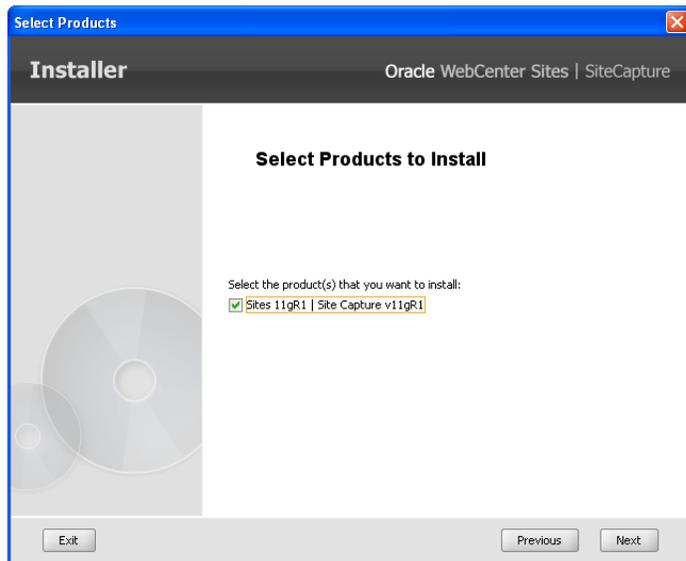
The installer “Welcome” screen is displayed:



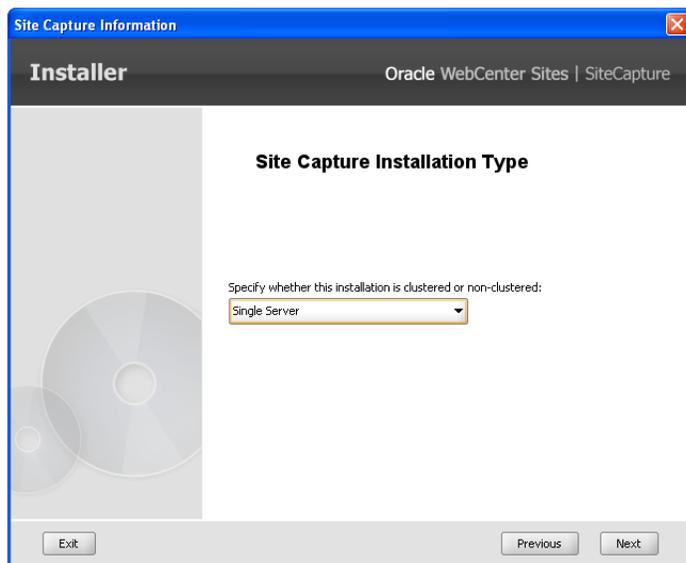
3. Specify the path to the directory where Site Capture will be installed and click **Next**.



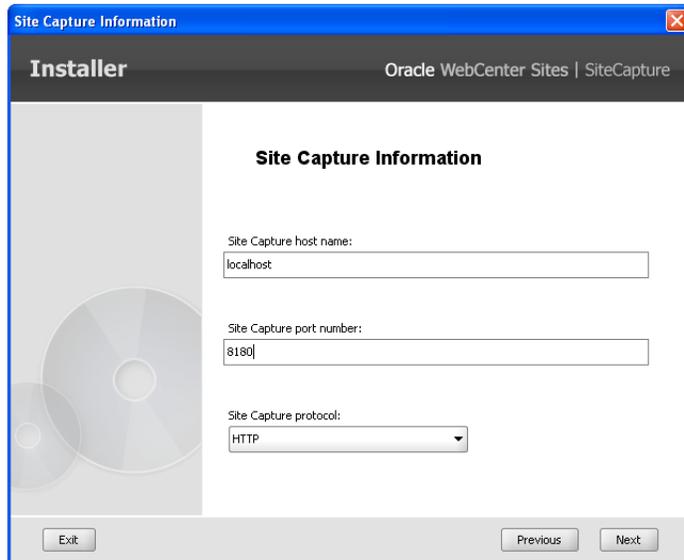
4. Select the product to be installed (Site Capture) and click **Next**.



5. Select either **Single Server** or **Cluster**, and click **Next**.

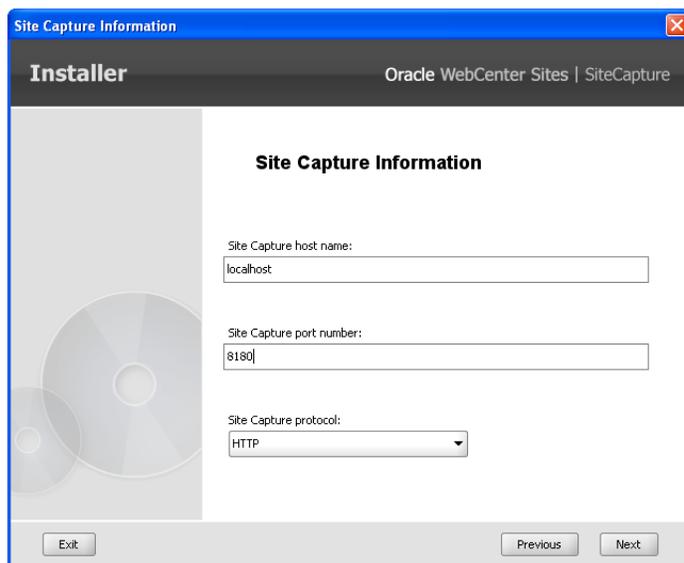


6. Enter information about the Site Capture host machine:
 - Single server installation:
 - Enter the host name (or IP address) of the machine running the Site Capture application server.
 - Enter the port number of the Site Capture application server (the same number that you specified in the application server configuration steps).
 - Specify the protocol to be HTTP or HTTPS.



The screenshot shows a window titled "Site Capture Information" with a blue header bar. Below the header, the word "Installer" is on the left and "Oracle WebCenter Sites | SiteCapture" is on the right. The main content area is titled "Site Capture Information" and contains three input fields: "Site Capture host name:" with the text "localhost", "Site Capture port number:" with the text "8180", and "Site Capture protocol:" with a dropdown menu set to "HTTP". At the bottom of the window, there are three buttons: "Exit", "Previous", and "Next".

- Clustered installation:
 - Enter the host name (or IP address) of the machine running the load balancer.
 - Enter the port number of the load balancer.
 - Specify the protocol to be HTTP or HTTPS.



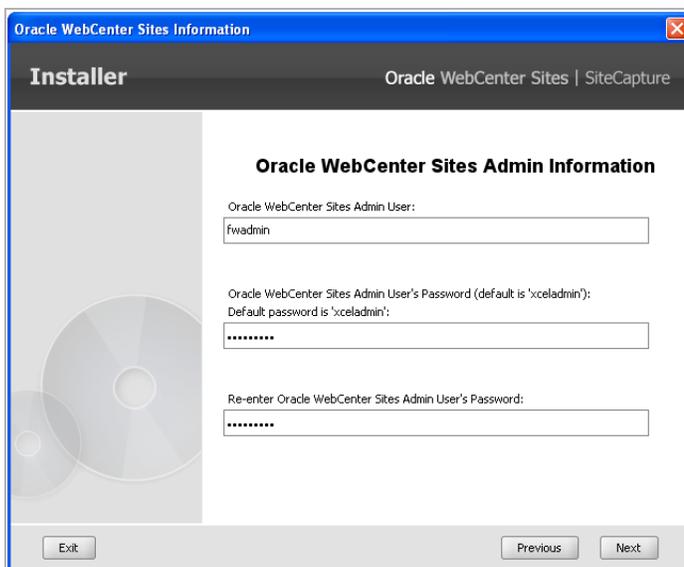
This screenshot is identical to the one above, showing the "Site Capture Information" dialog box. The input fields for "Site Capture host name:", "Site Capture port number:", and "Site Capture protocol:" all contain the same values: "localhost", "8180", and "HTTP" respectively. The window title, header, and footer buttons are also the same.

7. Enter information about the WebCenter Sites system on which Site Capture will run as an application:
 - WebCenter Sites' host name or IP address
 - Port on which WebCenter Sites is listening
 - Protocol of the WebCenter Sites application server (HTTP or HTTPS)
 - Context root of the WebCenter Sites application



The screenshot shows a dialog box titled "Oracle WebCenter Sites Information" with a blue header bar. The main area is titled "Oracle WebCenter Sites Information" and contains four input fields: "Oracle WebCenter Sites host name:" with the value "localhost", "Oracle WebCenter Sites port number:" with the value "8080", "Oracle WebCenter Sites context root:" with the value "/csj", and "Oracle WebCenter Sites protocol:" with a dropdown menu set to "HTTP". At the bottom, there are "Exit", "Previous", and "Next" buttons.

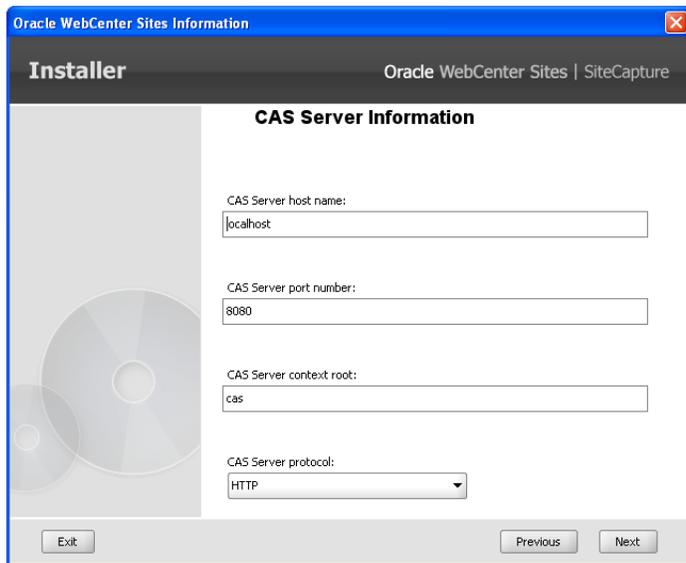
8. Enter the WebCenter Sites administrator's credentials:
 - Current user name of the WebCenter Sites general administrator
 - Current password



The screenshot shows a dialog box titled "Oracle WebCenter Sites Admin Information" with a blue header bar. The main area is titled "Oracle WebCenter Sites Admin Information" and contains three input fields: "Oracle WebCenter Sites Admin User:" with the value "fwadmin", "Oracle WebCenter Sites Admin User's Password (default is 'xceladmin'):" with a masked password "*****", and "Re-enter Oracle WebCenter Sites Admin User's Password:" with a masked password "*****". At the bottom, there are "Exit", "Previous", and "Next" buttons.

9. Enter information about the CAS application:
 - CAS host name (or IP address)

- Port number
- Protocol of the CAS application server (HTTP or HTTPS)
- CAS context root

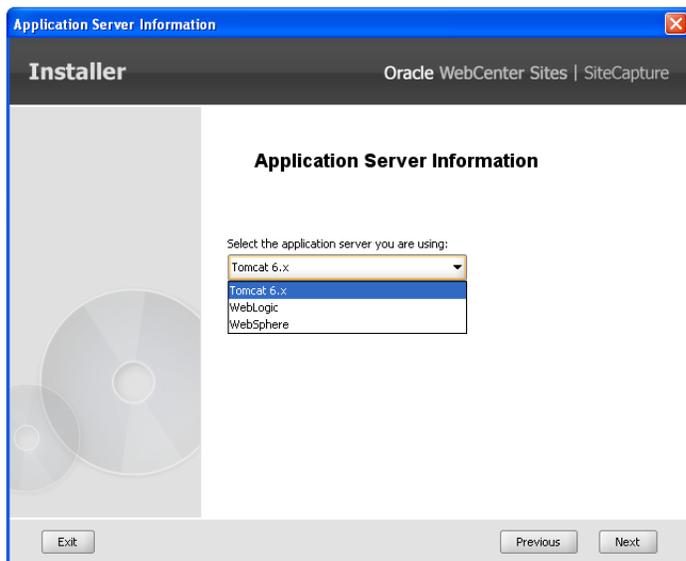


The screenshot shows the 'Oracle WebCenter Sites Information' installer window. The title bar includes 'Oracle WebCenter Sites | SiteCapture'. The main content area is titled 'CAS Server Information' and contains four input fields: 'CAS Server host name' (localhost), 'CAS Server port number' (8080), 'CAS Server context root' (cas), and 'CAS Server protocol' (HTTP). The 'Previous' and 'Next' buttons are visible at the bottom.

10. Select the application server on which Site Capture will be installed.

Note

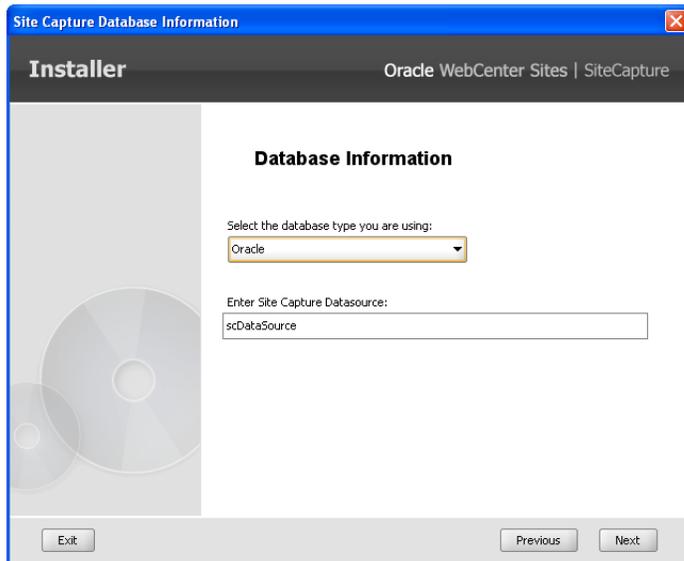
The installer will create a `ROOT.war` file, which you will deploy on the application server.



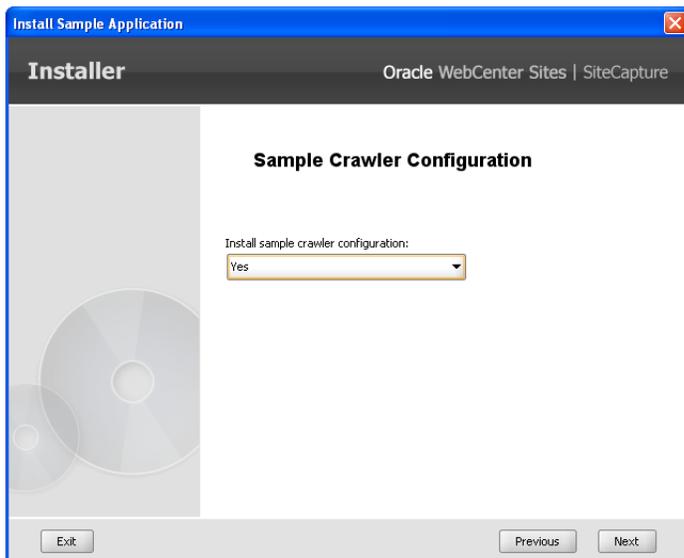
The screenshot shows the 'Application Server Information' section of the installer. It prompts the user to 'Select the application server you are using:' and displays a dropdown menu with the following options: Tomcat 6.x, Tomcat 6.x (highlighted), WebLogic, and WebSphere. The 'Previous' and 'Next' buttons are visible at the bottom.

11. Enter database information:

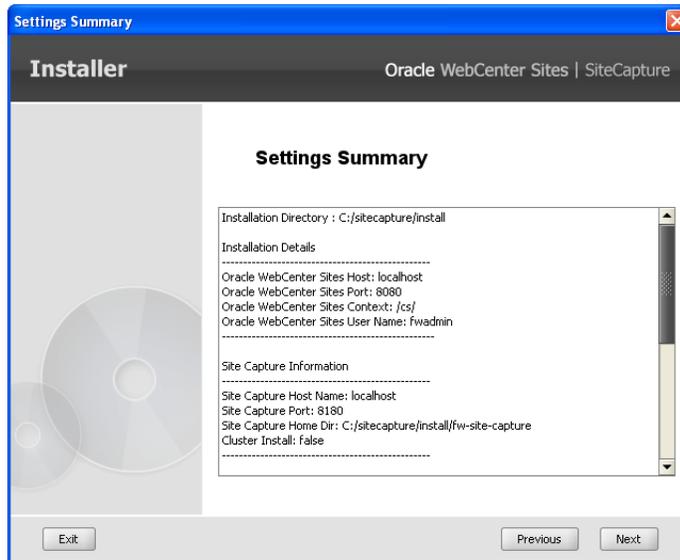
- Select the database to which Site Capture will connect (select either Site Capture's database or WebCenter Sites' database).
- Specify the name of Site Capture's data source exactly as it is configured in the application server (in [Chapter 3](#), "Installing Oracle WebCenter Sites: Site Capture").

**12. Select whether to install the sample crawlers:****Reminder**

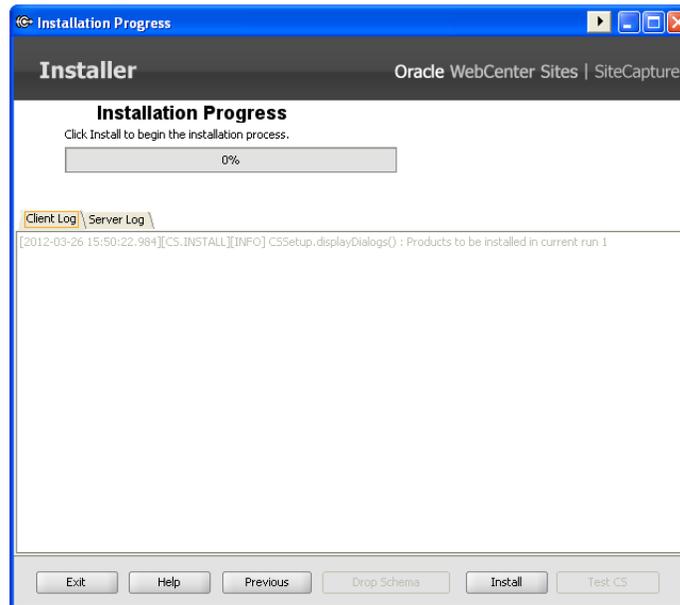
We recommend installing the sample crawlers. For information about the crawlers, see the first note on [page 20](#).



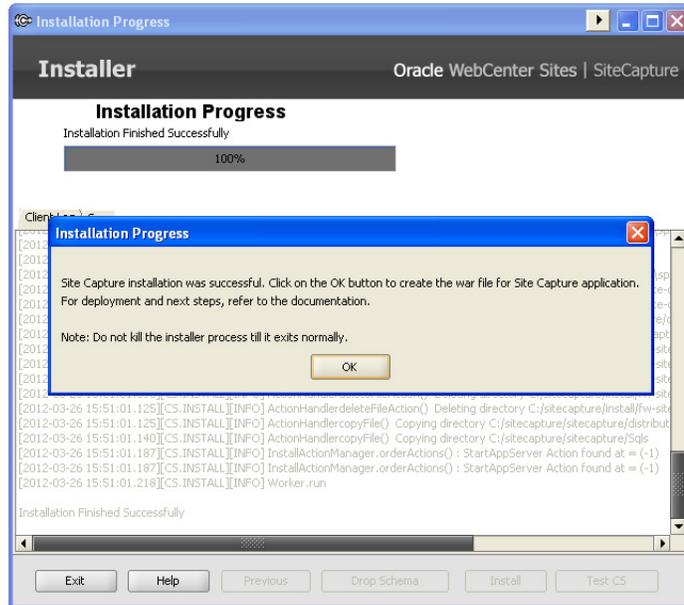
13. Review the settings summary to verify the data you have entered. To edit your settings, click **Previous** to return to the relevant screens.



14. Once you have confirmed your settings, click **Install**.



When the installation process successfully ends, the installer displays the following message: "Site Capture war file was successfully created"



The Site Capture file system was also created. The war file, called `ROOT.war`, is located in the `<SC_INSTALL_DIR>/fw-site-capture/webapps` folder. For information about the Site Capture file system, see [“Site Capture File System,”](#) on page 39.”

15. Deploy the Site Capture `ROOT.war` file. For instructions, see one of the following sections:
 - [Deploying on Tomcat Application Server](#)
 - [Deploying on WebLogic Application Server](#)
 - [Deploying on WebSphere Application Server](#)

Deploying Site Capture

Note

- For clustered installations, complete the deployment steps on each cluster member.
- The Site Capture application must have a context root of /.

Deploying on Tomcat Application Server

1. Copy the ROOT folder in <SC_INSTALL_DIR>\fw-site-capture\webapps\ and overwrite the ROOT folder in the <apache-tomcat-6.0.29 directory>\webapps\ directory.
2. To start Site Capture:
 - a. Ensure that WebCenter Sites is running.
 - b. Start the application server (run the startup.bat or startup.sh file in the <apache-tomcat-6.0.29 directory>\bin folder).
3. Continue to “[Post-Installation Steps](#),” on page 32.

Deploying on WebLogic Application Server

1. Copy the files named antlr-2.7.7.jar and commons-lang-2.5.jar in Site capture’s WEB-INF\lib folder:


```
<SC_INSTALL_DIR>\fw-site_capture\webapps\ROOT\WEB-INF\lib
```

 and paste them into the WebLogic installation path:


```
<Weblogic Home>\wlserver_10.3\common\lib.
```
2. Set the pre-class path in the setdomainEnv.cmd or setdomainEnv.sh file (located in <Weblogic Home>\user_projects\domains\<domain name>\bin) for the Site Capture domain.

Below is an example of where the changes need to be made:

- **On Windows** (in setdomainEnv.cmd):


```
set WL_HOME=E:\<WL_HOME>\Weblogic\wlserver_10.3

set PRE_CLASSPATH=%WL_HOME%\common\lib
\antlr-2.7.7.jar;%WL_HOME
%\common\lib\commons-lang-2.5.jar(in windows machine)

for %%i in ("%WL_HOME%") do set WL_HOME=%%~fsi
```
- **On Linux** (in setdomainEnv.sh):


```
WL_HOME="/root/Oracle/Middleware/wlserver_10.3"

PRE_CLASSPATH=$WL_HOME/common/lib/antlr-2.7.7.jar:
$WL_HOME/common/lib/commons-lang-2.5.jar

export WL_HOME
```

3. Deploy the Site Capture application:
 - a. Go to the deployment section in the console and select the path to the ROOT folder located in `<SC_INSTALL_DIR>\fw-site-capture\webapps`.
 - b. Select the managed server as the deployment target.
4. To start Site Capture:
 - a. Ensure that WebCenter Sites is running.
 - b. Start the managed server `<managed_server_name>`. For example:

```
./startmanagedweblogic.sh <managed_server_name>  
http://<admin_server_hostname>:<admin_server_port>/
```
5. Continue to [“Post-Installation Steps,” on page 32](#).

Deploying on WebSphere Application Server

1. In the WebSphere Admin interface, do the following:
 - a. Go to the path **Application > Application Types > WebSphere enterprise applications** and select the `ROOT.war` file located in `<SC_INSTALL_DIR>\fw-site-capture\webapps\`.
 - b. Complete the deployment process.
2. To start Site Capture:
 - a. Ensure that WebCenter Sites is running.
 - b. Start the application server.
3. Continue to [“Post-Installation Steps,” on page 32](#).

Post-Installation Steps

When Site Capture is installed and the application server is started, Site Capture starts and automatically registers itself as an application enabled on the AdminSite. In addition, Site Capture adds the necessary tables and records to its database, unless the database is Oracle.

To complete and verify your Site Capture installation, complete the steps in the following sections:

- [If Site Capture Is Using an Oracle Database](#)
- [Verifying the Site Capture Application](#)
- [Authorizing Users to Work with Site Capture](#)

If Site Capture Is Using an Oracle Database

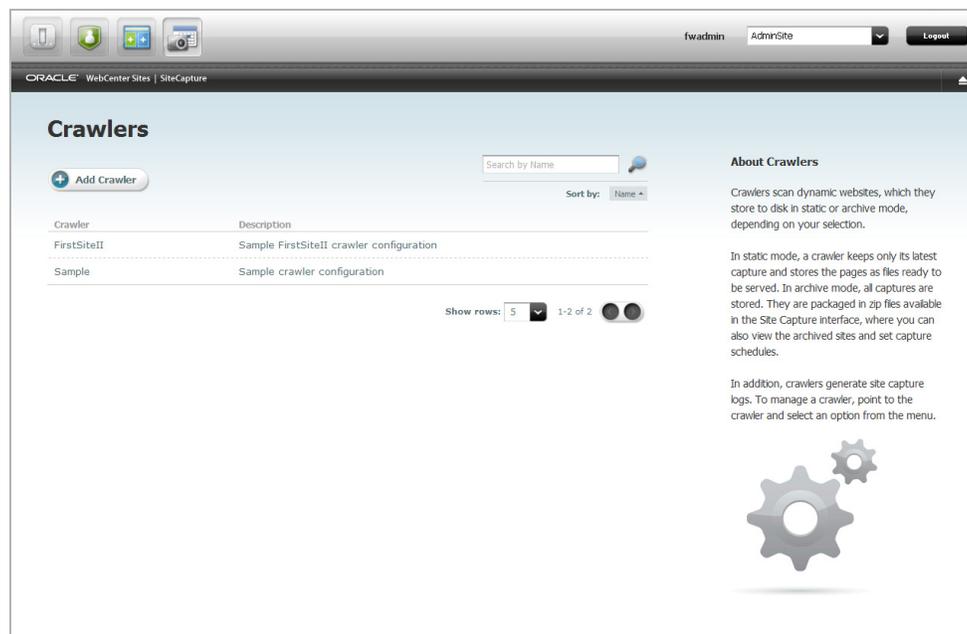
When Site Capture uses the Oracle database, the sql script that creates the schema must be run manually. The script, named `crawler_oracle_db.sql`, is available in the `<SC_INSTALL_DIR>/Sql-Scripts` folder.

Verifying the Site Capture Application

Ensure that the Site Capture application is available in the WEM Admin interface:

1. Log in to WebCenter Sites as a general administrator.
2. Navigate to the AdminSite and select the WEM Admin application.
3. In the menu bar, click **Apps** and verify that **Site Capture** is listed on the “Apps” page.
4. Navigate to the Site Capture application (represented by this icon: ).

The home page opens. If you chose to install the sample crawlers, they are listed on the home page as **Sample** and **FirstSiteII**.



The screenshot displays the 'Crawlers' management interface. At the top, there is a navigation bar with 'fwadmin' and 'AdminSite' tabs, and a 'Logout' button. Below the navigation bar, the page title is 'ORACLE WebCenter Sites | SiteCapture'. The main content area is titled 'Crawlers' and features an 'Add Crawler' button, a search box labeled 'Search by Name', and a 'Sort by: Name' dropdown menu. A table lists the installed crawlers:

Crawler	Description
FirstSiteII	Sample FirstSiteII crawler configuration
Sample	Sample crawler configuration

Below the table, there is a 'Show rows: 5' dropdown and a '1-2 of 2' indicator. To the right of the table, there is an 'About Crawlers' section with the following text:

About Crawlers

Crawlers scan dynamic websites, which they store to disk in static or archive mode, depending on your selection.

In static mode, a crawler keeps only its latest capture and stores the pages as files ready to be served. In archive mode, all captures are stored. They are packaged in zip files available in the Site Capture interface, where you can also view the archived sites and set capture schedules.

In addition, crawlers generate site capture logs. To manage a crawler, point to the crawler and select an option from the menu.

At the bottom right of the page, there is a gear icon representing settings.

Authorizing Users to Work with Site Capture

The Site Capture application is designed for WebCenter Sites general administrators and developers with general administrator rights. All such users have access to the AdminSite, where the Site Capture application is registered and running. (Access is granted via the `GeneralAdmin` role on the AdminSite and the user's membership in the `RestAdmin` security group. For more information about user authorization and the WEM Admin interface, see the *Oracle WebCenter Sites Administrator's Guide for the WEM Framework*.

Site Capture users also require administrative access to its host machine, especially to retrieve statically captured sites and crawler logs.

Next Steps

- If you wish to set up publishing-triggered site capture, see [Chapter 4, “Enabling Publishing-Triggered Site Capture.”](#)
- To get started with Site Capture, see the *Oracle WebCenter Sites Administrator's Guide for the Site Capture Application* for information about navigating the interface and writing crawler configuration code.

Chapter 4

Enabling Publishing-Triggered Site Capture

You have the option to enable publishing-triggered site capture once the Site Capture application is installed.

This chapter contains the following topics:

- [Integrating Site Capture with Oracle WebCenter Sites' Publishing Process](#)
- [Next Steps](#)

Integrating Site Capture with Oracle WebCenter Sites' Publishing Process

Site capture can be triggered by the completion of a RealTime publishing session. If you wish to enable publishing-triggered site capture, complete the steps in this section after you have installed the Site Capture application ([Chapter 3](#)). The steps show you how to integrate WebCenter Sites' publishing system to communicate with your installed Site Capture application. For some of the possible system configurations, see [Figures 3 and 4](#) on [page 11](#).

To integrate Site Capture with WebCenter Sites' publishing process

1. On the WebCenter Sites source system:
 - a. Deploy the `fw-crawler-publish-listener-1.0.jar` file to the `<cs_deploy>/WEB-INF/lib` folder.
 - b. Unzip the `fw-crawler-publish-listener-1.0-elements.zip` file and import `FW_PublishingEventRegistry.html` using `CatalogMover`.

This step creates a `RemoteElementInvokingPublishingEventListener` record in the `FW_PublishingEventRegistry` table in the WebCenter Sites database, which will allow publish events to call the `InvokeCrawler` element on the WebCenter Sites target system.
 - c. Restart the WebCenter Sites source system.
2. On the WebCenter Sites target system:
 - a. Using `CatalogMover`, import `ElementCatalog.html` and `SiteCatalog.html` from the unzipped `fw-crawler-publish-listener-1.0-elements.zip` file extracted in step 1b above (on the WebCenter Sites source system).

This step imports the `InvokeCrawler.jsp`, which is used to start the crawler(s) in the Site Capture application.

Note

The crawler(s) must be defined in the publishing destination definition for Site Capture and in the Site Capture application. For more information, see [“Next Steps,” on page 37](#).

- b. Copy the `crawler.properties` file (in the `<cs_deploy>/WEB-INF/classes` folder) and configure the properties listed below:
 - `sc.url`: Do one of the following:
 - For a single-server installation**, specify the URL of the Site Capture application:
`sc.url=http://<sitecapturehost:sitecaptureport>/__admin`
 - For a clustered installation**, specify the URL of the load balancer:
`sc.url=http://<loadbalancerhost:loadbalancer>/__admin`
 - `cas.url=http://<casHost:casport>/cas`
Specify the CAS application that is pointed to by the Site Capture application:

- `cs.username=<RestAdmin User>`
Specify the user name of the WebCenter Sites general administrator exactly as it was specified during the Site Capture installation process:
 - `cs.password=<Password>`
Specify the above user's password exactly as it was specified during the Site Capture installation process:
 - c. Deploy the `fw-crawler-publish-listener-1.1.jar` file to the `<cs_deploy>/WEB-INF/lib` folder on the target WebCenter Sites system.
3. You have completed the integration process. Continue to “[Next Steps](#)” for a summary on setting up a publishing-triggered site capture operation.

Next Steps

At this point, you have completed the integration process. However, for publishing-triggered site capture to work, the following conditions must also be satisfied:

- A RealTime publishing destination definition must be configured on the source system to name the crawler(s) that will be invoked to capture the newly published site. The definition must also specify the crawlers' capture mode.
- The crawler(s) named in the step above must exist in the Site Capture application. In addition, the `CrawlerConfigurator.groovy` file for each crawler must specify at least a valid starting URI and link extraction logic for the crawler.

Note

Once the above configuration steps are completed, you will publish the site from the source WebCenter Sites system to the target WebCenter Sites system. When publishing ends, site capture begins and proceeds as follows:

1. The source WebCenter Sites system calls the `InvokeCrawler` element on the target system.
2. The target WebCenter Sites system communicates with the Site Capture application and invokes the crawler(s).
3. The WebCenter Sites target system communicates crawler invocation status to the WebCenter Sites source system. Both the source and target systems record the status information in their own log files (`futuretense.txt`, by default).

At the same time, the invoked crawlers capture site resources either statically or in archive mode, depending on your settings in the publishing destination definition.

When enabling publishing-triggered site capture, you can configure as many publishing destination definitions and invoke as many crawlers as necessary. When you are ready to proceed with the configuration steps above, refer to the *Oracle WebCenter Sites Administrator's Guide for the Site Capture Application* for instructions. In the same guide, you will find information about navigating the Site Capture interface, setting up a site capture operation, and coding a crawler's configuration file to control the site capture process.

Appendix A

Site Capture File System

The Site Capture file system is created during the Site Capture installation process to store installation-related files, property files, sample crawlers, and sample code used by the FirstSiteII crawler to control its site capture process. The file system also provides the framework in which Site Capture organizes custom crawlers and their captures.

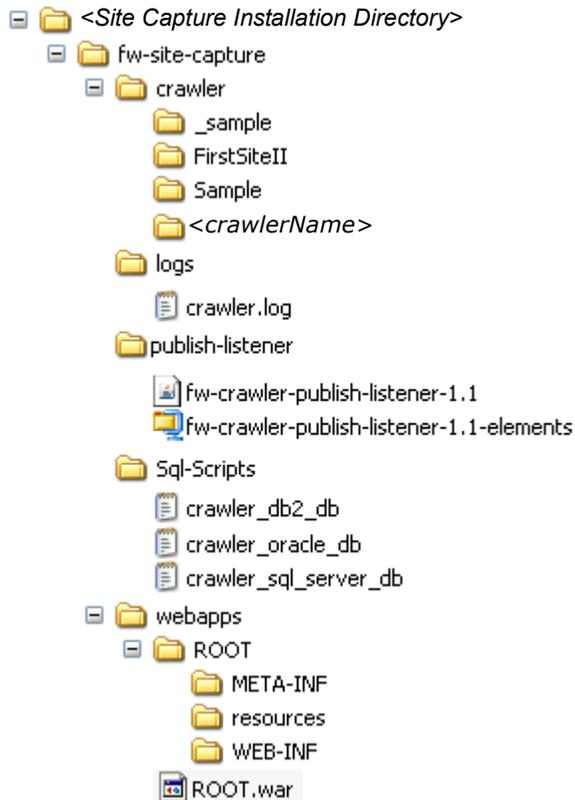
This chapter contains the following topics:

- [General Directory Structure](#)
- [Custom Folders](#)

General Directory Structure

Figure 1 shows Site Capture’s most frequently accessed folders to help administrators find commonly used Site Capture information. All folders, except for <crawlerName>, are created during the Site Capture installation process (for information about <crawlerName> folders, see [page 41](#) and “Custom Folders,” on [page 42](#)).

Figure 1: Site Capture File System



Folder	Description
/fw-site-capture	The parent folder.
/fw-site-capture/crawler	Contains all Site Capture crawlers, each stored in its own crawler-specific folder.
/fw/site-capture/crawler/_sample	Contains the source code for the “FirstSiteII” sample crawler. For information about the code, see the <i>Oracle WebCenter Sites Administrator’s Guide for the Site Capture Application</i> . Note: Folder names beginning with the underscore character (“_”) are not treated as crawlers. They are not displayed in the Site Capture interface.

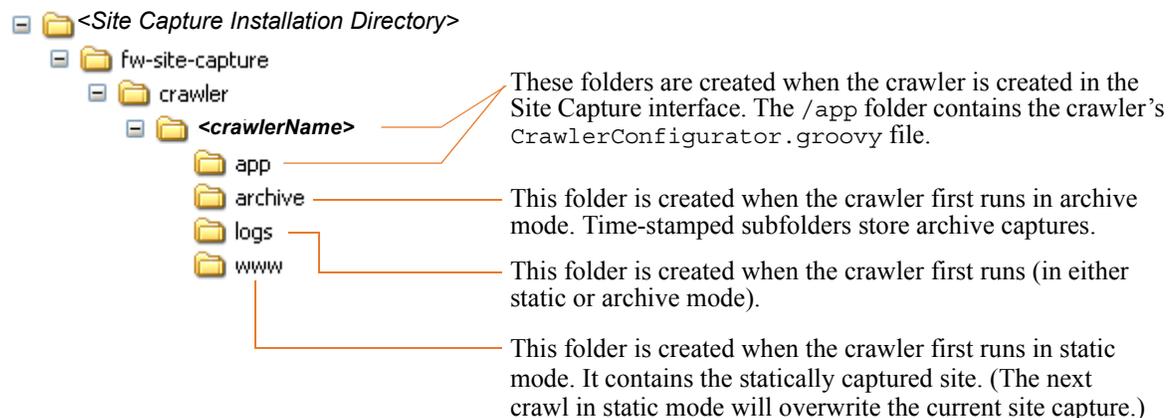
Folder	Description
/fw-site-capture/crawler/ <crawlerName>	<p>Represents a crawler. This custom folder is created by Site Capture for every crawler that a user defines in the Site Capture interface. This folder is used to organize the crawler's configuration file, archive captures, static captures, and logs. For a summary of the <crawlerName> folder structure, see “Custom Folders,” on page 42. For detailed information about the folder structure, see the <i>Oracle WebCenter Sites Administrator's Guide for the Site Capture Application</i>.</p> <p>Note: The FirstSiteII and Sample folders are examples of a <crawlerName> folder, as explained in the next two rows of this table.</p>
/fw-site-capture/crawler/ FirstSiteII	<p>Represents a sample crawler named “FirstSiteII.” This folder is created only if the “FirstSiteII” crawler was installed during the Site Capture installation process.</p> <p>The FirstSiteII folder contains an /app folder, which stores the CrawlerConfiguration.groovy file specific to the “FirstSiteII” crawler. The file contains advanced configuration code for capturing WebCenter Sites' FirstSiteII dynamic sample website as a static website, and in this manner, demonstrating the use of various methods and interfaces in the BaseConfigurator class.</p> <p>When the “FirstSiteII” crawler is invoked in static or archive mode, subfolders are created within the /FirstSiteII folder. For more information about the folder structure, see the <i>Oracle WebCenter Sites Administrator's Guide for the Site Capture Application</i>.</p>
/fw-site-capture/crawler/ Sample	<p>Represents a crawler named “Sample.” This folder is created only if the “Sample” crawler was installed during the Site Capture installation process.</p> <p>The Sample folder contains an /app folder, which stores the CrawlerConfiguration.groovy file specific to the “Sample” crawler. The file contains basic configuration code for capturing any dynamic site. The code demonstrates the use of required methods (such as getStartUri) in the BaseConfigurator class.</p> <p>When the “Sample” crawler is invoked in static or archive mode, subfolders are created within the /Sample folder. For more information about the folder structure, see the <i>Oracle WebCenter Sites Administrator's Guide for the Site Capture Application</i>.</p>
/fw-site-capture/logs	<p>Contains the crawler.log file, a system log for Site Capture.</p>
/fw-site-capture/publish- listener	<p>Contains the following files needed for installing Site Capture for publishing-triggered crawls:</p> <ul style="list-style-type: none"> • fw-crawler-publish-listener-1.1-elements.zip • fw-crawler-publish-listener-1.1.jar

Folder	Description
/fw-site-capture/Sql-Scripts	Contains the following scripts, which create database tables that are needed by Site Capture to store its data: <ul style="list-style-type: none"> • crawler_db2_db.sql • crawler_oracle_db.sql • crawler_sql_server_db.sql
/fw-site-capture/webapps	Contains the ROOT/WEB-INF/ folder.
/fw-site-capture/webapps/ROOT/WEB-INF	Contains the log4j.xml file, used to customize the path to the crawler.log file.
/fw-site-capture/webapps/ROOT/WEB-INF/classes	Contains the following files: <ul style="list-style-type: none"> • sitecapture.properties file, where you can specify information for the WebCenter Sites application on which Site Capture is running. The information includes WebCenter Sites' host machine name (or IP address) and port number. • root-context.xml file, where you can configure the Site Capture database.

Custom Folders

A custom folder is created for every crawler that a user creates in the Site Capture interface. The custom folder, <crawlerName>, is used to organize the crawler's configuration file, captures, and logs, as summarized in [Figure 2](#).

Figure 2: Site Capture's Custom Folders: <crawlerName>



For more information about custom folders, see the *Oracle WebCenter Sites Administrator's Guide for the Site Capture Application*.