Clearwell eDiscovery Platform

Distributed Architecture Deployment Guide

7.1.2 Fix Pack 1
Clearwell eDiscovery Platform™: Distributed Architecture Deployment Guide

The software described in this book is furnished under a license agreement and may be used only in accordance with the terms of the agreement.

Last updated: 2012-11-5.

Legal Notice

Copyright © 2012 Symantec Corporation. All rights reserved.

Symantec, the Symantec Logo, Clearwell and the Clearwell eDiscovery Platform are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

The product described in this document is distributed under licenses restricting its use, copying, distribution, and decompilation/reverse engineering. No part of this document may be reproduced in any form by any means without prior written authorization of Symantec Corporation and its licensors, if any.

THE 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. SYMANTEC CORPORATION SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

The Licensed Software and Documentation are deemed to be commercial computer software as defined in FAR 12.212 and subject to restricted rights as defined in FAR Section 52.227-19 "Commercial Computer Software - Restricted Rights" and DFARS 227.7202, et seq. "Commercial Computer Software and Commercial Computer Software Documentation", as applicable, and any successor regulations. Any use, modification, reproduction release, performance, display or disclosure of the Licensed Software and Documentation by the U.S. Government shall be solely in accordance with the terms of this Agreement.

Symantec Corporation
350 Ellis Street
Mountain View, CA 94043
http://www.symantec.com
Step 1: Add Appliances to the Cluster 30
Step 2: Assign Roles 31
Step 3: Provision Appliances (Review Role only) 32

Distribution Enabled 34
Reassigning or Unassigning Roles 34

Appliance Management 35

Managing Appliances 35
Managing Cases (Processing Role Only) 35
  Processing Performance Considerations 36
Managing Cases (Review Role Only) 37
  Review Performance Considerations 37

Check Status Before Backups 38
  About Case Backups 38
  About Node (Appliance) Backups 38
  Prepare Nodes for Backup, Restore, or Removal 38
  Restoring a Node 40

Review Cache Process 41
  File Caching 41
  Review Node Jobs 41

Tips and Troubleshooting 43

Changing Appliance Settings 43
  Adjust Appliance Capacity (R710) 43
  Change the Backup Directory (Required) 44
  Set System-Level Security and Indexing Properties 44

Changing Processing Settings (Required) 46
  Change to Network Location (Email Files) 46
  Copy Files After Upgrading 47
  Install PKI Certificates and Enable Support for Signed and Encrypted Messages 47
  Change Threadder Performance in a Distributed Setup 50
  Disable Retriever Service on Cluster Master 51

Changing Distributed Review Settings 52
  Disable File Conversions on the Cluster Master 52
  Automate Load Balancing 53
  Turn Off Load Balancing 53

Configuring Job Output for Exports 54

Addressing Fragmented Nodes 54
  Prepare to Backup, Restore, or Remove Nodes 55
  Precautions for Node Backups with Fragmented Cases 56
  Precautions for Node Restore or Removal 56

Addressing Node Failures, and Ensuring System Reliability 57
  About Backups 57
  Cluster Backups 57
  Cluster Master Failure 57
  Case Home Failure 57
  Processing Node Failure 58

Contacting Technical Support 58
Distributed Architecture Deployment Guide

Welcome to the Clearwell eDiscovery Platform Distributed Architecture Deployment Guide. This guide provides configuration information for customers with a shared remote database architecture deployment to take advantage of the Review and Processing scale-out capabilities, or Distributed Review and Distributed Processing in Clearwell 7.x.

This section contains the following sections:

• “About This Guide” in the next section
• “Revision History” on page 5
• “Obtaining More Product Information” on page 6
• “Technical Support” on page 6
• “Product Documentation” on page 8

About This Guide

The Distributed Architecture Deployment Guide is intended to assist system administrators through the installation and configuration of their clustered environment, plus the administration tasks necessary to enable the Review and Processing scalability features, known as Distributed Review and Distributed Processing in the Clearwell eDiscovery Platform.

Revision History

The following table lists the information that has been revised or added since the initial release of this document. The table also lists the revision date for these changes.

<table>
<thead>
<tr>
<th>Revision Date</th>
<th>New Information</th>
</tr>
</thead>
</table>
| Nov 2012      | • Minor updates throughout, with changes in Setup and Troubleshooting sections for 7.1.2 Fix Pack 1:  
|               |   – Appliance compatibility and capacity adjustments. See “Setup Requirements” on page 21.  
|               |   – Threader property now enabled by default. See “Change Threader Performance in a Distributed Setup” on page 50. |
| Sep 2012      | • Added FAQ in Setup Requirements section, providing more detailed explanation and actual values for the number of cases, appliances in a cluster, and documents supported, plus key port configuration information. |
| May 2012      | • Updated to include Distributed Processing capabilities, available in Clearwell 7.1.1. |
| Feb 2012      | • New guide for distributed architecture deployments of Clearwell 7.0 enabling the Distributed Review feature. |
Obtaining More Product Information

To obtain more information, refer to:

- **Symantec Products Web site** — Go to [http://www.symantec.com/business/products](http://www.symantec.com/business/products)

- **Documentation link** — To obtain the most current online and PDF versions of the documentation, click the **Documentation** link at the bottom of any page in the Clearwell eDiscovery Platform.

- **Online help** — Click the **Help** link in the Clearwell user interface to access online help.

Documentation Comments/Feedback?

Got questions or comments on this guide, or other user documentation? We appreciate your feedback! Feel free to contact us at ClearwellInfoDev@symantec.com.

Technical Support

The primary role of the Clearwell Technical Support team is to respond to our customers to help resolve technical issues affecting features or functionality of the Clearwell eDiscovery Platform. This includes configuration, workflow, and internal system issues, but does not include assistance with legal-related or case-specific queries. For additional help with technical issues, refer also to Clearwell’s Knowledge Base content available online at:


For information about Symantec's support offerings, visit our Web site at:


Contacting Technical Support

Before contacting Technical Support, make sure you have satisfied the system requirements that are listed in your product documentation. Also, you should be at the system on which the problem occurred, in case it is necessary to replicate the problem.

When you contact Technical Support, please have the following information available:

- Product version and patch level
- Service tag number
- Available memory, disk space, and NIC information
- Operating system
- Problem description:
  - Affected case name
  - Screenshots demonstrating behavior
  - Accurate reproduction steps
  - Error messages and log files
– Troubleshooting that was performed before contacting Symantec
– Recent software configuration changes and network changes

**Note:** You must have administrative access to the system in order to perform most troubleshooting. Be sure to have this available, as well as a system/network administrator resource in the event access is required.

To contact Clearwell Technical Support, use any of the following methods:

- **Customer Support Portal** — (Clearwell Products)
- **Phone** — Toll-Free (North America):
  - 1-800-342-0652

**Licensing Portal**

If your Clearwell product requires registration or a license key, access our Licensing Portal:

- **Symantec Licensing Portal**
  - [https://licensing.symantec.com/acctmgmt/index.jsp](https://licensing.symantec.com/acctmgmt/index.jsp)

**Customer Service**

To contact Symantec Customer Care, use any of the following methods:

- **Customer Care Portal** — (Non-Technical Support) Enterprise Products
  - [http://customercare.symantec.com](http://customercare.symantec.com)
- **Phone** — Toll-Free (North America):
  - 1-800-721-3934
    - For other regions: [http://www.symantec.com/business/support/assistance_care.jsp](http://www.symantec.com/business/support/assistance_care.jsp)

Customer Service is available to assist with non-technical questions, such as the following types of issues:

- Questions regarding product licensing or serialization
- Product registration updates, such as address or name changes
- General product information (features, language availability, local dealers)
- Latest information about product updates and upgrades
- Information about upgrade assurance and support contracts
- Information about the Symantec Buying Programs
- Advice about Symantec's technical support options
- Non-technical pre-sales questions
- Issues that are related to manuals
Product Documentation

The table below lists the end-user documentation that is available for the Clearwell eDiscovery Platform product.

<table>
<thead>
<tr>
<th>Document</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Installation and Configuration</strong></td>
<td></td>
</tr>
<tr>
<td>Installation Guide</td>
<td>Describes prerequisites, and how to perform a full install of the Clearwell software application</td>
</tr>
<tr>
<td>Upgrade Overview Guide</td>
<td>Provides critical upgrade information, by version, useful prior to upgrading an appliance to the current product release</td>
</tr>
<tr>
<td>Upgrade Guide</td>
<td>Describes prerequisites and upgrade information for the current customers with a previous version of the Clearwell software application</td>
</tr>
<tr>
<td>Utility Node Guide</td>
<td>For customers using utility nodes, describes how to install and configure appliances as utility nodes for use with an existing Clearwell software setup</td>
</tr>
<tr>
<td>Native Viewer Installation Guide</td>
<td>Describes how to install and configure the Brava Client for native document rendering and redaction for use during analysis and review in Clearwell</td>
</tr>
<tr>
<td>Distributed Architecture Deployment Guide</td>
<td>Provides installation and configuration information for the Review/Processing Scalability feature (7.x) in a distributed architecture deployment</td>
</tr>
<tr>
<td><strong>Getting Started</strong></td>
<td></td>
</tr>
<tr>
<td>Navigation Reference Card</td>
<td>Provides a mapping of the Clearwell user interface (7.x) compared to 6.x</td>
</tr>
<tr>
<td>Clearwell QuickStart Guide</td>
<td>Describes basic appliance and case configuration</td>
</tr>
<tr>
<td>Reviewer QuickStart Guide</td>
<td>A reviewer’s reference to using the Analysis &amp; Review module in Clearwell</td>
</tr>
<tr>
<td><strong>User and Administration</strong></td>
<td></td>
</tr>
<tr>
<td>Legal Hold Guide</td>
<td>Describes how to set up and configure a Clearwell appliance for Legal Holds, and use the Legal Hold module as an administrator in Clearwell</td>
</tr>
<tr>
<td>Identification and Collection Guide</td>
<td>Describes how to prepare and collect data for processing, using the Identification and Collection module</td>
</tr>
<tr>
<td>Case Administration Guide</td>
<td>Describes case setup, processing, and management, plus pre-processing navigation, tips, and recommendations. Includes processing exceptions reference and associated reports, plus file handling information for multiple languages, and supported file types and file type mapping</td>
</tr>
<tr>
<td>System Administration Guide</td>
<td>Includes system backup, restore, and support features, configuration, and anti-virus scanning guidelines for use with Clearwell</td>
</tr>
<tr>
<td>Load File Import Guide</td>
<td>Describes how to import load file sources into Clearwell</td>
</tr>
<tr>
<td>User Guide</td>
<td>Describes how to perform searches, analysis, and review, including detailed information and syntax examples for performing advanced searches</td>
</tr>
<tr>
<td>Export and Production Guide</td>
<td>Describes how to use, produce, and troubleshoot exports</td>
</tr>
<tr>
<td>Transparent Predictive Coding Guide</td>
<td>Describes how to use the Transparent Predictive Coding feature in Clearwell to train the system to predict results from control data and tag settings</td>
</tr>
<tr>
<td>Document</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>Reference and Support</td>
<td></td>
</tr>
<tr>
<td>Collection</td>
<td>A quick reference card of how to collect data in Clearwell</td>
</tr>
<tr>
<td>OnSite Collection</td>
<td>A quick reference for performing OnSite collection tasks</td>
</tr>
<tr>
<td>Review and Redaction</td>
<td>Reviewer's reference card of all redaction functions</td>
</tr>
<tr>
<td>Keyboard Shortcuts</td>
<td>A quick reference card listing all supported shortcuts</td>
</tr>
<tr>
<td>Production</td>
<td>Administrator's reference card for production exports</td>
</tr>
<tr>
<td>User Rights Management</td>
<td>A quick reference card for managing user accounts</td>
</tr>
<tr>
<td>Online Help</td>
<td>Includes all the above documentation (excluding Installation and Configuration) to enable search across all topics. To access this information from within the Clearwell user interface, click Help.</td>
</tr>
<tr>
<td>Release</td>
<td></td>
</tr>
<tr>
<td>Release Notes</td>
<td>Provides latest updated information specific to the current product release</td>
</tr>
</tbody>
</table>

For the latest product information: http://www.symantec.com/business/products
Overview

This section provides an introduction to the Distributed Review and Processing features of the Clearwell eDiscovery Platform, available in Clearwell 7.x.

This section also provides the pre-requisites necessary and critical setup information needed (before installing and configuring your Clearwell appliances) to take advantage of these features.

Refer to the following topics in this section:

- “Introduction to the Distributed Architecture” on page 11
  - “Basic Concepts” on page 12
  - “Features of a Distributed Architecture” on page 13
- “How Does a Distributed Architecture Work?” on page 14
  - “How are Cases Processed using Distributed Processing?” on page 15
  - “How Does Distributed Review Work?” on page 16
- “Configuration Considerations” on page 18

Introduction to the Distributed Architecture

Clearwell 7.x introduces a distributed architecture for scaling processing and review capacity to meet increasing demands of case-specific, and eDiscovery project requirements. A distributed architecture deployment provides enhanced performance with the flexibility for handling large volume, complex cases by enabling:

- **Distributed Processing and Review.** Support for large volume cases to be processed, and multiple reviews to be distributed across one or more appliances.
  - **Parallel Processing.** Optimized processing via on-demand added throughput.
  - **Reviewer Scalability.** Dynamic increased number of supported users

Clearwell’s distributed architecture is a clustered environment configuration which is set up to communicate with a shared remote database that enables the Distributed Review and/or Distributed Processing features.

How they are employed depends on how the appliances in the cluster are assigned. For example, assigning multiple appliances the “Review role” allows case review to be distributed to multiple reviewers across more than one appliance. Thus, Distributed Review allows for more concurrent reviewers than one appliance alone could support.

**Distributed Processing** allows administrators to add nodes as needed at any point during case processing to optimize processing performance.

Specifically, Distributed Processing and Distributed Review features are capable of higher scalability wherein administrators can manage throughput, response time, and other performance metrics by adding and assigning additional hardware appliances to perform these tasks.
Basic Concepts

The following table describes some basic terms and concepts used when discussing a distributed architecture or clustered environment in relation to the Clearwell eDiscovery Platform.

**Distributed Architecture Basic Concepts**

<table>
<thead>
<tr>
<th>Term or Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appliance / Node</td>
<td>A Clearwell, or Clearwell-supported appliance, referring to an appliance which has the 7.x version of Clearwell software installed. <strong>Note:</strong> In this guide, the term <em>appliance</em> and <em>node</em> are used interchangeably.</td>
</tr>
<tr>
<td>Case Home</td>
<td>The primary location, or node selected when the case was created. The Case Home node is, by default, a Review node. This node contains all the case-specific information, such as the search index files. The Case Home can be a worker node of the Cluster Master, or the Cluster Master itself. See also Cluster Master, Review Node, and Worker Node.</td>
</tr>
<tr>
<td>Cluster / Clustered Environment</td>
<td>A group of appliances/nodes in a distributed architecture, or clustered environment.</td>
</tr>
<tr>
<td>Cluster Master (CM) / Master</td>
<td>The Cluster Master (CM) is a node that manages the cluster information of the distributed architecture. It is the nerve center of the clustered environment, which contains all the information about each node. The Cluster Master is light weight and can be quickly replicated. The Cluster Master’s configurations are persisted in two parts: One part is stored in a MySQL database, and the other is stored in a file system serving as the configuration and license files. The Master is system-critical and must be continuously up and running. If the Master is shut down, it automatically shuts down all other worker nodes in the cluster. See Worker Node.</td>
</tr>
<tr>
<td>Database Node</td>
<td>In a distributed processing setup, we requires the client to have a shared remote database that can be accessed by all the nodes (worker nodes and Cluster Master) on the cluster.</td>
</tr>
<tr>
<td>Processing Node</td>
<td>The appliance or node assigned for processing. Processing typically requires a high power appliance. Once processing is completed, the data that was on the processing node is merged with the data from other processing nodes. The final, merged data is then copied to the Case Home and other Review nodes, and the intermediate processing data is deleted. The processing node is typically one of the worker nodes of the Cluster Master. While the Cluster Master can be a processing node, it should be kept light weight. Thus, assigning the Processing Role to the Cluster Master is not recommended. The processing node can also be assigned to different cases. However, processing nodes assigned to more than one case can slow processing throughput if all cases are processed simultaneously. To optimize performance in this use case, one could process each case at different times on the same processing node.</td>
</tr>
<tr>
<td>Review Node</td>
<td>The appliance or node assigned and provisioned for review. This node usually contains search index files so that searches can be performed locally. The Case Home is always a review node (by default), and can be a worker node of the Cluster Master. Like processing nodes, review nodes can also be assigned to different cases. However, review nodes assigned to more than one case can slow searches if all cases are reviewed simultaneously. To optimize performance in this use case, reviewers could review one case at different times on the same review node.</td>
</tr>
</tbody>
</table>
Features of a Distributed Architecture

Configuring your environment to use Clearwell’s Distributed Review and Processing capabilities in a distributed architecture deployment enables the following additional features of the Clearwell eDiscovery Platform:

- Dynamically increases the scalability (number of reviewers, and processing speed), to maximize the benefits of the review process using the Clearwell appliance, with:
  - No impact on review or processing capacity of a given appliance
  - Search, analysis and review operations occurring on multiple nodes (as configured by an administrator), called load balancing. (See “Configuration Considerations” on page 18.)

- Data-Optimized Export:
  - Optimizes system resource usage and achieves higher scalability.
  - Exported documents are analyzed up front, optimizing retrieval and maximum efficiency of resource usage on multiple appliances, when available.
  - System no longer waits for large files in one subset to complete before exporting the next subset.

  For more information about utility nodes, how to add utility node appliances to the cluster to improve export performance, refer to the Utility Node Guide.

- Case Review Provisioning:
  - Any node in a cluster can be provisioned and de-provisioned for case review without disrupting current users provided that the specified node is not currently in use for review.

- Case Processing Assignment:
  - Distributes the processing workload among several processing nodes with maximized scaling.
    - New hardware resources (either processing nodes or utility nodes) can be added, or removed to help quickly respond and manage processing throughput needs.
How Does a Distributed Architecture Work?

Clearwell’s Distributed Architecture employs a data distribution model that can scale various operations by allocating portions of work to appliances dedicated for the operation.

The Case Home

In a distributed architecture deployment, the case is managed through a single point of entry, the Case Home. The Case Home is an appliance, or node, on which the case was first created, and where the main components of a case are maintained. The Case Home is the central location where any information is checked/updated in the case for all the allocated processing and review nodes. Case home presents a unified and consistent view of the case regardless of the current state of processing, review, and exports. The Case Home and all its processing nodes must belong to a single Clearwell cluster. (See “Configuration Considerations” on page 18 for a comprehensive list of key terms.)

Using a Shared Remote Database Server

Clearwell appliances are configured to use a shared remote database server as part of the cluster. (To see an example of a clustered environment, see “Setting up a Cluster Environment” on page 19.) In particular, the shared remote database server maintains all metadata information about cases, appliances, documents, processing state, review state, work product and other critical data. The remote database server is fully sharable across many cases, their processing jobs and their review work product.
How are Cases Processed using Distributed Processing?

**Processing Cases across Multiple Appliances**

Case processing starts with the Case Home discovering various case sources and building a master list of all sources to be processed. In the Pre-Processing phase, case administrators can filter this source list to handle a portion of the discovered data.

**Distribution of Data among Processing Nodes**

For PST, NSF, and Exchange mailboxes, distribution of data occurs by pulling the information from the Case Home. Each processing node requests the set of PST/NSF files to process from the Case Home. Upon processing the assigned files, the processing nodes will request from the Case home any further files to process, and continues until all PST/NSF files are processed. By contrast, all loose files are distributed equally among all the processing nodes.

**Processing Nodes**

Each processing node processes the data after de-duplicating it against the Case Home. This ensures that only one copy of the data is processed even if the copies are distributed across the various processing nodes. For items that are not duplicated, Clearwell creates appropriate entries in the database. In addition, Clearwell maintains searchable text indexes locally, on each node assigned and provisioned for review.

**Post-Processing and Index Consolidation**

When Clearwell completes processing of all sources (in which data is de-duplicated and indexed), the Post-Processing phase begins. Post-Processing includes index validation, threading, search analytics, statistics, distributed merge, and centralized merge and consolidation builds. Specifically, Clearwell’s distributed architecture ensures task consistency and fully complete builds to present a single view of all data. The following illustrates the index consolidation process.

![Distributed Processing: Index Consolidation Process](image)

During the consolidation phase of Post-Processing, all partial search indexes from each processing node into the Case Home is consolidated and pushed to all nodes provisioned for review. The new content then becomes available for search and review. After Post-Processing is complete, administrators can remove any processing nodes and re-purpose them for other cases.
How Does Distributed Review Work?

Scaling Reviewers across Multiple Appliances

Once processing is completed, review scalability components can be enabled for increasing the number of reviewers that perform searches and create work product (such as tagging, redactions, document comments, review comments). To increase review scalability, administrators can add new appliances, assign them the Review role, and provision them for review for a case. Review load, while spread across many appliances, presents a single consistent view of data to all reviewers.

The following diagram is an example of Review nodes and how they interact with the Case Home.

The Distributed Review Workflow

When you provision the node for Review, the Case Home replicates its view of all its searchable indexes to dedicated Review nodes. Each Review node presents a read-only copy of the search index to all reviewers. When a reviewer performs a search, the search results are based on the copy of the index that is local to the appliance. However, the metadata for search results come from the shared remote database server. Additionally, any retrieval of items for display purposes originate from the shared pool of sources. Local search results are then consolidated with both the metadata and raw content into one transparent view. As a result, users have complete visibility to both the locally-created work product as well as that created by other users on other Review nodes. Since work product references are stored in the database with normal database-level locking and synchronization, users see a single consistent view of their search results.
Using Utility Nodes to Increase Throughput

In the review phase, utility nodes can help increase throughput of retrievals from PST/NSF, and file conversions needed to generate the HTML and image view of any document. As a result, utility nodes improve performance when users perform text/native review, eview cache, metadata, or production, or production export jobs.

The following diagram is an example of utility nodes being used for distributed review:

Optimized Export

Using utility nodes helps improve export performance since retrievals and conversion are parallelized. The documents to be exported are analyzed up front, allowing the system to optimize their retrieval and use resources on multiple appliances, when available, most efficiently. This allows the system to continue exporting larger files in a subset while continuing to export the next subset. For more information about using utility nodes, refer to the Utility Node Guide.
Configuration Considerations

You may want to use a Distributed Architecture configuration if the following requirements apply:

1. You want to scale your current review capacity to increase the number of reviewers. This is for large reviews occurring on a single node (with no other concurrent tasks running).
   
   **Note:** Current review capacity is equal to 100 users on an appliance with 48-Core CPU/128 GB RAM, and 25 users for an appliance with 16-Core CPU/32 GB RAM.

2. You have large volume cases, or large number of cases and want to scale your environment to increase processing throughput.

3. Your standard Web application deployment model requires having a database in its own tier.
   
   **Note:** Processing of new batches does not affect current search and review context of any data already processed into the case.

In the standard configuration, Symantec ships Clearwell 7.x with the Oracle MySQL® database server software embedded in the product.

**Note:** You may also see some performance enhancements as a result of the database having its own memory and disk. Clearwell recommends using the same hardware configuration for all appliances used in a distributed architecture. For hardware-specific considerations, see “Hardware Configuration” on page 21.

For help determining required components for your architecture, or for help with configuring your appliances to use a shared remote database server, contact your Clearwell Solutions Consultant, or Clearwell Customer Support for details.
Setting Up your Environment

This section provides the pre-requisites necessary and critical setup information needed before installing and configuring your Clearwell appliances.

Refer to the following topics in this section:

- “Setting up a Cluster Environment” on page 19
- “Sample Distributed Architecture” on page 20
- “Setup Requirements” on page 21
  - “Hardware Configuration” on page 21
  - “Appliance Capacity” on page 22
  - “Changing Settings” on page 22
- “Frequently Asked Questions” on page 22

Setting up a Cluster Environment

A cluster environment is an architecture consisting of multiple appliances, in which one is designated as the cluster master and the other member appliances are the worker nodes and if utilized, the utility nodes.

**Note:** Utility nodes can be deployed on commoditized hardware and do not necessarily reflect Clearwell appliances.

This configuration allows Clearwell to balance the case review workload among multiple nodes, used to optimize performance and improve review efficiency. Additionally, each node can be connected to one or more utility nodes. The Utility Resource Manager on the Cluster Master controls all utility nodes. Each node can also have its own set of utility nodes. For information on how to use and configure utility nodes, refer to the Utility Node Guide.

**Note:** In a clustered environment enabled for Distributed Review, once a case restore job is started on a worker node, it cannot be stopped. For information about node backups and restore, refer to the "Backup and Restore" in the System Administration Guide.
Sample Distributed Architecture

The following is a sample distributed architecture configuration with a Cluster Master, and three worker nodes, with Node 1 as the Case Home. (This architecture also uses utility nodes for optimized performance.) Each node has been assigned a role of either processing, or review, or both. Nodes with the Review role are provisioned to distribute or load balance reviewers across multiple appliances.

In this example, all three nodes have been configured for Review and/or Processing scalability. Nodes 1 and 3 are provisioned for Distributed Review, (with Node 1 as the Case Home). These nodes are enabled for load balancing of multiple reviewers for a single case on each appliance. Nodes 2 and 3 are assigned the Processing Role (node 3 has both roles), allowing cases to be processed on more than one appliance. See “How are Cases Processed using Distributed Processing?” on page 15 for Processing scalability details.

In the distributed review configuration, automatic redirection distributes the review load onto each provisioned appliance, one at a time, until the appliance reaches its set capacity for the number of concurrent reviewers, or the load balancing threshold. Then, subsequent reviewers are distributed to the next available configured appliance, and so on (depending on the case size, and number of configured nodes, and concurrent reviewers). For more information about the load balance setting, see “Changing Distributed Review Settings” on page 52.
Setup Requirements

A distributed architecture requires:

• Shared remote database

• Nodes must be part of the same cluster, and communicating to the same shared remote database

• All data sources must be visible to all review and/or processing nodes, including contained PSTs/NSFs and converted files

Hardware Configuration

The following table lists the supported appliance configuration for the MySQL server needed for a distributed architecture deployment.

Supported Hardware Configurations

<table>
<thead>
<tr>
<th>Operating System</th>
<th>Memory</th>
<th>CPUs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>64-bit Windows Server 2008 or</td>
<td>32 GB</td>
<td>16-Core</td>
<td></td>
</tr>
<tr>
<td>Windows Server 2008 R2 Enterprise</td>
<td>128 GB</td>
<td>48-Core</td>
<td></td>
</tr>
</tbody>
</table>

**IMPORTANT!** Existing Clearwell appliances can be repurposed for a distributed architecture deployment. However, it is recommended that all appliance models (new or existing) are using the same configuration. For Dell appliance specification details, refer to the appropriate document: Clearwell Appliance Specifications for Dell [R710 or R810 (for 7.0, 7.1, and 7.1.1), and 8100 or 8200 for 7.1.2)] available on the Customer Support Portal. For information about other appliance types, contact your Clearwell Solutions Consultant, or Technical Support.

Hardware Compatibility

Clearwell supports the following types of appliances for the corresponding functions, and minimum requirements needed for hardware (and Virtual Machines if used) in a distributed architecture.

Appliance Compatibility Options

<table>
<thead>
<tr>
<th>Appliance Type</th>
<th>Function</th>
<th>Minimum Hardware</th>
<th>Cluster Required</th>
<th>VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td>Caching, Retrieval, Export</td>
<td>4-Core CPU x 4 GB RAM</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Review</td>
<td>Review</td>
<td>16-Core CPU x 32 GB RAM</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Processing</td>
<td>Processing</td>
<td>16-Core CPU x 32 GB RAM</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Stand-Alone Master</td>
<td>Cluster Master</td>
<td>4-Core CPU x 8 GB RAM</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Shared Remote Database</td>
<td>Database Server</td>
<td>8-Core CPU x 32 GB RAM</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>
Appliance Capacity

Depending on the Clearwell appliance, or size of your case, the following capacities apply in a distributed architecture:

### Appliance Capacity

<table>
<thead>
<tr>
<th>Capacity</th>
<th>Detail</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Size</strong></td>
<td>10 million reviewable</td>
<td>• Can be in a single case or split across multiple cases</td>
</tr>
<tr>
<td></td>
<td>items per case</td>
<td>• “10 million reviewable items” means less than 10 million unique documents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Data characteristics also matter. Cases with numerous unique terms (typically found in attachments) have larger overhead</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Note:</strong> “Reviewable items” includes email, attachments, or loose files, but do not include duplicates. For a detailed explanation, see “Frequently Asked Questions” on page 22.</td>
</tr>
<tr>
<td><strong>Concurrent Reviewers</strong></td>
<td>16-Core CPU/32 GB RAM = 20-25</td>
<td>• Capacity of 25 concurrent reviewers on an R710 appliance is made possible by changing a property setting. See “Adjust Appliance Capacity (R710)” on page 43.</td>
</tr>
<tr>
<td></td>
<td>16-Core CPU/128 GB RAM = 75-100</td>
<td></td>
</tr>
</tbody>
</table>

Changing Settings

You may want to change the default settings for load balance thresholds (for Distributed Review nodes), or disable load balancing depending on the number of reviewers, or individual requirements.

**About Load Balancing Thresholds (for Review)**

During distributed review, when a node must balance the workload, it determines the load percentage on each node based on its hardware configuration and the number of users on each node.

Clearwell calculates the percentage of additional load that each node will balance before it will redirect a user to another node. Clearwell redirects users to various appliances based on the number of users on each appliance. The default threshold value is set to 5.

**Disabling Load Balancing**

Administrators can disable load balancing by allocating nodes to reviewers, and requiring them to log in directly. See “Turn Off Load Balancing” on page 53 in the Tips and Troubleshooting section.

**Frequently Asked Questions**

This section covers the following topics:

- *What is meant by the “10 Million Reviewable Items” limit?*
- *Can an appliance have more than 10 million documents?*
• What is the limit for the total number of documents, users, and cases with an R810 appliance?

• What is the maximum number of supported appliances in a cluster?

• What is the total number of cases one cluster can have?

• What network ports do I need to open between Review appliances, Case Homes, Database, and shared storage?

**What is meant by the “10 Million Reviewable Items” limit?**

The 10 million items supported in a Distributed Architecture deployment using Clearwell includes all unique, concurrently reviewable documents, which can be in a single case or split across multiple cases on any appliance. This is true for single-server deployments as well as clustered environments.

**Can an appliance have more than 10 million documents?**

Yes. An appliance can have more than 10 million documents, however only 10 million unique documents can be accessed at a time. Having a remote database server does not increase the supported limit. The limit is tied to the maximum memory on an appliance that dictates search performance. Having a remote database server allows the nodes in the cluster to synchronize with each other by operating with the same metadata information across the cluster. Further, the remote database server does not increase the memory on an appliance that is used for search, and therefore does not increase the supported limit for the total number of reviewable items.

**What is the limit for the total number of documents, users, and cases with an R810 appliance?**

The physical limit on the number of documents, users and cases on a R810 depends on the disk space of the machine and database. In terms of user activity, Symantec Clearwell supports 10 million unique, (concurrently and actively) reviewable documents, and 100 concurrent reviewers on a Dell R810 appliance.

**What is the maximum number of supported appliances in a cluster?**

In a Distributed Architecture deployment, Symantec Clearwell supports a maximum of 25 appliances per cluster.

**What is the total number of cases one cluster can have?**

The maximum number of cases depends on the physical limit of the desk space on the hardware used in the cluster. Symantec Clearwell supports 10 millions unique (concurrently and actively) reviewable documents on a single appliance.

**What network ports do I need to open between Review appliances, Case Homes, Database, and shared storage?**

Clearwell appliances in a clustered deployment communicate with each other using RMI protocols that utilize dynamic ports. These ports are where the server and clients negotiate on which ports to use. Similar to a single-server deployment, all Clearwell appliances utilize the native Windows Firewall to limit exposure to potential network security threats.
Firewall is configured to allow only the minimal set of ports/applications required for Clearwell operations. Windows Firewall configurations for clustered deployments are no different than a single-server deployment. To see a list of applications that are allowed through the Windows Firewall, go to **System > Support Features**, and select **Firewall Browser**. Enter the Firewall Command "CONFIG". Clearwell also provides a command line utility, `b firewall`, that can customize the configuration of the firewall.
Installation and Configuration

This section describes the overall steps for installing the MySQL server, then migrating and configuring the Clearwell appliance in preparation for a distributed architecture deployment:

- “Install the MySQL Server Setup for the Shared Remote Database” in the next section
- “Clearwell Appliance Migration” in the next section
- “Other Required Cluster Configurations” in the next section

IMPORTANT! It is important to download and install the correct version of MySQL. Although it is possible to download a free copy of MySQL, customers are responsible for ensuring licensing compliance with Oracle. Once MySQL is installed on the database server you can run the Clearwell shared remote database configuration utility which will configure MySQL to take advantage of the machine's memory. The configuration utility will automatically configure MySQL to take advantage of 8, 12, 16, 32, 64 or 128 GB of RAM. Administrators can fine tune MySQL as desired. Refer to the MySQL documentation for details.

Install the MySQL Server Setup for the Shared Remote Database

Prepare the shared remote database server

1. Download and run the MySQL-5.5.8-winx64 (Windows Version) installer, either from the MySQL Web site:
   - http://downloads.mysql.com/archives.php?p=mysql-5.5&v=5.5.8
   or from the Clearwell installer folder:
   - \packages\mysqlx64\n
2. Next, locate and run the DBMSConfiguration.exe configuration utility to configure the Database, from the Clearwell folder:
   \V711\utilities\DBMS

3. Follow the on-screen instructions to install MySQL Server Setup 5.5 on the database server, noting the following:
   A. For Setup Type, select the Custom option.
   B. For Choose a Destination, change the local drive to D:\ or a drive with the most space.

   Note: By default, MySQL is installed in C:\Program files, however space on the C:\ drive may be insufficient. Thus, Clearwell recommends installing the MySQL server on another drive with at least 400 GB of free space.
C. At the end of the MySQL installation, the MySQL installer (by default) selects the option to *Launch the MySQL Instance Configuration Wizard*. **Clear this option**, as the Clearwell MySQL configuration utility already automatically configures MySQL.

![MySQL Server 5.5 Setup](image)

**Note:** Be sure to clear this default option before proceeding.

4. Ensure that the Clearwell appliance and MySQL remote database server show the same date/time, and time zone.

**Note:** If the time zone on the appliance, MySQL remote database, and local database are not all set correctly, issues can occur during migration of the local database data to the shared remote database.

5. Run the Clearwell shared remote database configuration utility. This utility also comes with the installer for a shared remote database configuration.

---

### Clearwell Appliance Migration

Follow the steps in this section to configure a Clearwell appliance for the shared remote database, and migrate all cases to the reconfigured appliance. Configuring an appliance to use the shared remote database invokes two main processes:

A. Moves an appliance's local databases to the shared remote database and configures Clearwell to use the new shared remote database server.

B. Migrates all cases on an appliance at once.

**Note:** The following commands include an internal database backup. Thus, performing a separate node backup on the appliance is not necessary. Also note that 'b' commands can only be run from the installation directory before migration, to use a shared remote database.

**To configure an appliance and migrate all cases:**

6. Run the `b stop-services` command.

7. Run the `b migrate-db` command.
This moves the appliance’s local databases to the shared remote database, and configures the application to use the shared remote database server instead of the local ones.

The “b migrate-db” command also migrates all cases on an appliance at once. The entire migration is atomic and either succeeds or fails. If it fails, Clearwell is still usable with the local MySQL database on the Clearwell appliance.

**Note:** If there are multiple cases or large cases on an appliance, migration may take a while. You can expect the migration to take approximately 40 minutes per one million documents on an appliance. (Times may vary depending on network configuration and utilization.)

8. When prompted, type the **Fully Qualified Domain Name** of the database server.

   Example: `clearwelldb.myhost.com`

9. When prompted, accept the default database name, or type a new database name if there might be a naming conflict on the database server.

10. Run the `b start-services` command.

### Other Required Cluster Configurations

Before adding machines to the cluster, the distributed system requires certain database, system, and case-level configurations:

- Configure Shared Remote Database in a Cluster
- Change Backup Directory Location

See also “Changing Processing Settings (Required)” on page 46 for additional required settings.

#### Configure Shared Remote Database in a Cluster

All appliances in a cluster must be configured to use the same shared remote database. In this configuration, all appliances use the same database server even though multiple databases are created. Follow the same steps as listed in the previous section (“Clearwell Appliance Migration”), but note the following:

**IMPORTANT:** To ensure all machines in the cluster use the same database server and configuration, first migrate all machines to the same database server, and then cluster them. It is important that you use the **fully qualified domain name** of the database server when running the “b migrate-db” command on each machine.

#### Configure Shared Directory

In a clustered environment, enabled for Review and/or Processing scalability, administrators must take the following additional actions:

1. Make the directory `D:\cwshared` a shared directory so that all appliances in the cluster can access the share as, for example: `\machinename\cwshared`. This also ensures your clustered environment is correctly configured to enable retrievals and native view conversions.

2. Remove the read-only option on this directory for all folders, subfolders and contents. If this is not removed, native view can fail in a distributed architecture setup. (Otherwise, ensure that read/write access is given to all Clearwell service accounts.)
Change Backup Directory Location

In a distributed architecture deployment, the directory location for case backups must be changed. This must be done on all appliances in the cluster. See "Change the Backup Directory (Required)" on page 44.
Appliance Assignment and Provisioning

Follow the steps in this section to add, assign, and provision appliances in your newly-configured distributed architecture from the Clearwell user interface.

Refer to the following topics in this section:

- “Overview” on page 29
- “Administrator Workflow” on page 30
  - “Step 1: Add Appliances to the Cluster” on page 30
  - “Step 2: Assign Roles” on page 31
  - “Step 3: Provision Appliances (Review Role only)” on page 32
- “Distribution Enabled” on page 34
- “Reassigning or Unassigning Roles” on page 34

Overview

After installing and configuring appliances to use a shared remote database, use the steps in the following section to add the machines to the cluster. Log in to the appliance you want to designate the Cluster Master, and then add all other appliances to the same cluster as worker nodes. See “Step 1: Add Appliances to the Cluster” on page 30.

Setting System-Level Properties on all Appliances in the Cluster

Once appliances are added, you can set additional security and indexing properties (not enabled by default in a distributed setup) on each appliance in the cluster, as described in “Set System-Level Security and Indexing Properties” on page 44.

The system administrator can then assign multiple nodes the “Review Role” (for Distributed Review for example), on a case-by-case basis. After a processing job has completed for a case, the indexes are copied to the selected review nodes. This task is called Provisioning and happens automatically after a processing job completes. For more about how to assign and provision appliances depending on your needs, see “Step 2: Assign Roles” on page 31.

Similarly, when multiple nodes are assigned the “Processing Role”, one or more cases can be processed on more than one node, thus reducing processing time and increasing performance. (Nodes with the Processing Role are not provisioned, only assigned.)
Administrator Workflow

Once all appliances are configured, continue with the following steps:

1. **Add any new appliances to the cluster.** Log in to the appliance you want to designate the Cluster Master, and then add all other appliances to the cluster as worker nodes. See “Step 1: Add Appliances to the Cluster” in the next section. (To add and configure utility nodes, refer to the Utility Node Guide.

2. **Assign one or more appliances.** Assign an appliance (node in the cluster) for review and/or processing: (See “Step 2: Assign Roles” on page 31.)
   
   A. Any case(s) on the appliance(s) being assigned for the Processing Role that are currently in process will not have these changes take effect until the next time processing or post-processing is run on the case. To ensure your changes take effect immediately, first stop any cases that are currently processing. Then re-run processing after assigning the Processing Role. See “Distribution Enabled” on page 34.
   
   B. If an appliance is assigned the Review Role before processing is started, provisioning happens automatically as part of the processing job. (If it was assigned after processing began, the appliance must be manually provisioned; continue to step 3.)

3. **[Review Role only] Provision appliances.** Provision appliances for review if the Review role was assigned after processing began. See “Step 3: Provision Appliances (Review Role only)” on page 32. For processing and performance effects, see “About Case Backups” on page 38.

After assignments are saved, Load Balancing and/or Parallel Processing is enabled. For nodes provisioned for the Review Role, reviewers will be load balanced between the provisioned nodes. For nodes assigned the Processing Role, case processing will be distributed across assigned nodes to achieve parallel processing. See “Distribution Enabled” on page 34.

**Step 1: Add Appliances to the Cluster**

Be sure to add any new appliances (worker nodes) to the cluster before assigning roles. If you are adding utility nodes (for optimized review and export throughput), refer to the Utility Node Guide for steps on how to add utility nodes to a clustered environment.

**To add a new node to the cluster:**

1. Log on to the Clearwell user interface on the appliance you wish to designate as the Cluster Master, and go to System > Appliances.

2. On the Appliances screen, click Add to add a new node.

3. On the **Appliance** tab, enter the name of the appliance, host, and type the port number.

4. Click **Save.** Continue with “Step 2: Assign Roles” in the next section. See also “Set System-Level Security and Indexing Properties” on page 44 to enable certain properties on newly-added appliances.
Step 2: Assign Roles

Depending on your particular case processing and reviewer requirements, determine which appliances will be assigned the Processing and/or Review roles, then assign the roles.

**Assignment Strategy**

While you can assign both roles to the same nodes, you may want to assign only the Review or Processing Role at one time, depending on your case requirements, size, and number of reviewers. It is not necessary to assign both roles at the same time, unless you want to assign both roles to the same appliance. If doing so, assign the roles before processing starts, so that by the time the case completes processing, it will be immediately ready for review.

**Note:** Clearwell recommends doing this before any cases on those appliances have begun processing so that all data is processed. See “Processing Performance Considerations” on page 36 to understand how your assignments on nodes will affect case processing if processing has already begun. Refer to “Appliance Capacity” on page 22 and “About Case Backups” on page 38 to be sure your assignments optimize performance based on your appliance model, case sizes, and/or number of concurrent reviewers.

**To assign roles (from Case Home)**

1. From a selected case, click **Case Home**, and select **Appliance Roles**.

2. Check the box next to one or more appliances you want to assign (for each case) the Review and/or Processing Role.

3. Click **Save Changes**, or to go back to the original settings, click **Reset**.
4. Determine next step:

**If you assigned the Processing Role:**
- Appliances are now enabled for parallel processing. See “Distribution Enabled” on page 34.
- Continue with managing appliances to prepare/restore nodes as necessary for case processing. See “Managing Cases (Processing Role Only)” on page 35.

**If you assigned the Review Role:**
- If you assigned one or more appliances before processing started, you do not need to continue to provision appliances (for the Review Role). Provisioning happens automatically as part of the processing job. See “Distribution Enabled” on page 34.
- If you assigned an appliance (for the Review Role) after processing began, continue to “Step 3: Provision Appliances (Review Role only)” in the next section to manually provision the appliance to perform Review. Manually provisioning an assigned appliance also allows you to use it immediately after assignment.

**Step 3: Provision Appliances (Review Role only)**

When you provision an appliance, Clearwell copies all required files (such as text indexes, and concept search binaries) to the appropriate node (for review and/or processing) to prepare the appliance for its role.

**Use Cases and Considerations for Provisioning**

Administrators can add, remove, and replace nodes in a distributed architecture deployment when provisioning appliances using any one of the following use cases:

Administrators can add, remove, and replace nodes for a case...

- **on which processing has not started.** Once processing starts, the newly-configured nodes will take effect for all phases of processing.
- **after discovery begins but before processing starts.** This provisioning does not require processing to be stopped. The newly-provisioned nodes will take effect for Indexing and Post-Processing phases.
- **after discovery starts, and only after processing starts.** Changes in provisioned nodes will take effect only during the Post-Processing phase. It will not affect Indexing. To have the new configuration take effect during the Indexing phase, then stop the processing job during Indexing and then resume processing using the same processing batch label.
- **after post-processing starts, without stopping post-processing.** This provisioning does not have any effect on current processing jobs. It will only take effect on the next post-processing and/or processing job. To allow for changes to take effect, stop post-processing and rerun the job.

**Note:** If post-processing is stopped during the Distributed Merge phase, upon re-running the job changes to the node provisioning information for the case will not take effect.
Provisioning appliances

If the review node is assigned before processing, there is no need to run a separate job; it is part of the processing job itself. Verify that all selected review nodes are provisioned to ensure the consistency of search results.

Example 1:

<table>
<thead>
<tr>
<th>Appliance Name</th>
<th>Assignment</th>
<th>Current Assignments</th>
<th>Review Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>cleanwell-test.local</td>
<td>✔️</td>
<td>15 Cases</td>
<td>Not Provisioned</td>
</tr>
<tr>
<td>Kirkwood-test.local</td>
<td>✔️</td>
<td>7 Cases</td>
<td>Provisioned</td>
</tr>
</tbody>
</table>

Example 2:

To provision appliances for the Review Role

1. Consider the current state of the cases on the appliance(s) you intend to provision. See “Use Cases and Considerations for Provisioning” on page 32.) If a node is assigned after processing begins, continue to step 2 to manually provision the appliance.

2. With the case selected, from Case Home > Appliance Roles, select the appliance you want to provision for review.

3. Check the box next to one or more cases to be provisioned for the Review role.

   **Note:** If you did not select this node for the Review role at the time of processing, click the (Provision) link to manually provision the selected node for review.

4. Click Save Changes, or to go back to the original settings, click Reset.
Distribution Enabled

**Distributed Processing Enabled (Processing Role)**

Each node that is assigned the Processing Role for a case is automatically enabled for, and ready to receive processing instructions from the Case Home. Once processing is started, the Case Home assigns the source to the first, then to each of the next available nodes. At the same time, idle processing nodes are performing self-checks so that when ready, they too will share the processing workload until all available nodes are simultaneously processing case source data.

To manage appliances before, during, and after case processing, see “Managing Cases (Processing Role Only)” on page 35.

**Load Balancing (Review Role) Enabled**

When a node is provisioned with the Review Role, reviewers can log in to any node in the cluster. Based on case selection, reviewers are redirected to the appropriate node. If multiple review nodes exist, the system uses a simple load balancing policy to choose a review node automatically for the reviewer.

**IMPORTANT!** Scalability largely depends on your hardware configuration. Refer to “Setup Requirements” on page 21 to check the capacity for your appliance models (memory and CPU), as well as compatibility.

**Reassigning or Unassigning Roles**

As your configuration or requirements change, you can unassign a node, or reassign an existing node for a different role, as well as replace a node and assign it a new role. To assign, reassign, or unassign roles on a node (such as if new nodes are added to the cluster), follow the same steps as described in “Step 2: Assign Roles” on page 31.
Appliance Management

Follow the steps in this section to manage your newly-assigned/provisioned nodes within the distributed architecture.

Refer to the following main topics in this section:

- “Managing Appliances” in the next section
- “Managing Cases (Processing Role Only)” on page 35
- “Managing Cases (Review Role Only)” on page 37
- “Check Status Before Backups” on page 38
  - “Prepare Nodes for Backup, Restore, or Removal” on page 38
- “Review Cache Process” on page 41

Managing Appliances

This section covers how to control the throughput and performance of your appliances, primarily for case processing. Use case scenarios also provide an understanding of how distribution is handled when case processing has started on nodes that were later added, reassigned, or unassigned, with, or without stopping the processing job.

If you have assigned one or more nodes the Processing Role, you can prepare nodes for backup, restore if necessary, or even remove nodes (if other than the Case Home) when no longer in use.

Managing Cases (Processing Role Only)

Note: This section applies only when you want to perform a node backup, node restore, or node removal in the current cluster, and you have cases that are either homed on the node to be backed up, restored, or removed; or the node was used as a processing node for cases homed on other nodes.

To manage case processing on nodes assigned for the Processing Role, you can check for any actions to be performed on the nodes before backing up. However, if processing is incomplete (such as due to job failure, or stops for any reason before finishing), then you must first perform a “Prepare” task (see “Prepare Nodes for Backup, Restore, or Removal” on page 38) before doing a node backup.

If you have cases homed on this node, or this node was used as a processing node for cases homed on other nodes and the case encountered a failed or stopped job, then the case is considered to be in a fragmented state. See “Addressing Node Failures, and Ensuring System Reliability” on page 57.

If the node backup, restore, or removal fails, then the node may have cases that are fragmented. See “Check Status Before Backups” on page 38.
Processing Performance Considerations

The following effects on operations should be considered when monitoring jobs and managing appliances (assigned the Processing Role) in a distributed architecture. (For nodes assigned the Review role, see also “Review Performance Considerations” on page 37.)

**Distributed Processing Considerations**

- Administrators can assign, unassign, and replace nodes in the following cases:
  - on which *processing has not started*.
  - after discovery starts, but *before Indexing begins*.
  - after discovery starts but *after Indexer begins*.
  - after post processing starts, without stopping post processing.

  For details, see “Use Cases and Considerations for Provisioning” on page 32.

- Administrators can stop at any point during Processing (Indexing and Post-Processing), and then assign nodes. The new assignments *will take effect at restart*.

- If nodes have cases in a “Fragmented” state:
  - Those cases will be considered in an Invalid state. If processing the case after restore, processing will fail due to its Invalid state.

  **CAUTION:** Once the case is in an Invalid state, it cannot be processed. At that point, the case can only be restored from a previous backup, before re-starting processing.

  - nodes can still be restored. Similar to restoring a case, previously fragmented cases are flagged as Invalid (shown from the System > All Processing screen). If processing the case after restore, processing will fail due to its Invalid state.

- In general, any case with inconsistent data fragments (such as missing or corrupted data, or stopped or failed processing job) due to offline node or backup with force option, will be considered as unusable/Invalid. As such, those cases cannot be further processed, and cannot be backed up if additional review is done on processed data. Therefore, you should perform a “Prepare” task (for removal, or restore) to ensure nodes are in a “ready” state first, before doing the actual node backup, removal or restore.
Managing Cases (Review Role Only)

If you have cases on one or more nodes assigned the Review role, keep in mind these general guidelines and rules for review and export performance.

Review Performance Considerations

The following effects on operations should be considered when monitoring jobs and managing appliances (assigned the Review Role) in a distributed architecture.

Distributed Review Considerations

• Processing running on the review node may have some affect on the review throughput
• The following rules apply with tagged case data:
  – Most of the case data is in the shared remote database
  – Tag events have a full text index. This index is centralized at Case Home, through which all operations pass.
  – Clearwell uses distributed locks to protect tag operations from race conditions, which means concurrent tagging operations on multiple nodes still have to yield to each other (at a very low level)
  – If every reviewer in the system uses the Document Note feature and tag-specific comments, tagging performance decreases.

Note: For large case review, Clearwell recommends dedicating the appliances as Review nodes especially for significantly large cases with a large number of reviewers during peak review times.

Export Considerations

• For exports running on the Case Home review node, exports may affect the review, such that exports will export the data from the Case Home.
Check Status Before Backups

As a best practice, you should perform periodic case backups, if not nightly, but as often as necessary to ensure your case data can be restored if needed in the event of a failure, such as a disk, database, or network failure. Refer to this section to perform the necessary status checks on all cases and nodes before backing up or starting any processing jobs.

About Case Backups

Case backups contain all the index and database information related to the selected case, including user-generated tags and notes. Perform a case backup when you want to checkpoint a case, that is, restore a case to a previous state. Case backups can also be used as a tool to transfer cases to different appliances.

**Note:** Backing up a case after restore loses all its provisioning information and must be provisioned again. See “Step 3: Provision Appliances (Review Role only)” on page 32.

About Node (Appliance) Backups

Node backups include all index and database information for all cases on the appliance and create a single appliance backup package, available for restore if needed.

**CAUTION:** Before performing a node backup, ensure that all cases on the node are not in a fragmented state, and that the node is ready for backup.

When cases on the node are ready to be backed up, (indicated with “Ready for Backup” in the Appliance), you can continue with the backing up the node. To perform an on-demand or scheduled node backup, refer to the section “Creating Appliance Backups” in the System Administration Guide.

Prepare Nodes for Backup, Restore, or Removal

Before performing a backup, restore, or removal of a node, first check that all cases on the node are in a non-fragmented state, or the “Ready” for backup, restore, or removal status is enabled. Follow the steps in this section to check each node in your cluster in preparation for one or more of these tasks.
To prepare a node for backup/restore/removal

1. In Clearwell, click System, and select Settings > Appliances.
2. Click the appliance you want to check then click the Appliance tab.
3. On the Appliance tab, check the section “Prepare the appliance for the following actions” for the following valid (non-fragmented) status indicators:

   - **Ready for Backup.** The node is in a valid state and ready to be backed up if needed. See “About Node (Appliance) Backups” on page 38.
   - **Ready for Restore.** The node is in a valid state and ready to be restored if needed.
   - **Ready for Removal.** (worker nodes only). The node is in a valid state and ready to be removed if needed. If the node is the Cluster Master, this button will appear unavailable as shown:

   ![Image](image1.png)

   CAUTION: If any one of these statuses do not appear, and instead show a “Prepare” option to restore, back up or remove the node, then the node is in a fragmented state. The Prepare option will not appear if any processing, post-processing, case backup, or case restore jobs are currently running on the node. These jobs must be completed or stopped first before attempting to back up, restore, or remove the node. Be aware however, if you stop a processing job before it completes, the case will be fragmented. In this case, the “Prepare” option appears.
What about backing up the case?

Nodes cannot be backed up, removed, or restored if they are currently running any unfinished processing jobs. Partially-processed cases may have data spread across multiple nodes, such that cases on those nodes will be fragmented.

To ensure the node is ready for backup, restore, or removal, cases on the node must first be defragmented. See “Addressing Fragmented Nodes” on page 54.

4. If the node indicates it is ready for backup, you can perform a normal node backup as directed in the System Administration Guide. See also “About Node (Appliance) Backups” on page 38.

Restoring a Node

An node- or cluster-level restoration restores all cases in the appliance backup package. However, specific cases cannot be restored from an appliance backup package. For more information, see “Addressing Fragmented Nodes” on page 54.

When you restore from a node from a backup you restore the entire appliance. You cannot restore specific cases from an node backup. Refer to the section “About Restore” in the System Administration Guide.

Note: Upon restoring a node, check that all provisioning information is still valid (as it may need to be re-provisioned) before reusing the node. See “Step 3: Provision Appliances (Review Role only)” on page 32.
Review Cache Process

Supported review scalability assumes that review caching has already been performed on the review set. A Review caching job runs on the review node on which it was initiated. However, the bulk of the tasks (retrievals, HTML/native view conversions) are distributed using the utility nodes. Utility nodes are recommended to be used to speed up review caching. Any Clearwell appliance in the cluster can also act as a utility node. For more information about utility nodes, refer to the Utility Node Guide.

Each review node fetches the cached copy on demand from the Case Home or its peers using an underlying file cache infrastructure. (See “Review Node Jobs” on page 41.)

**Note:** It is possible to plan Review caching on review subsets per node such that no file copying is necessary during actual review.

File Caching

While most file operations such as HTML rendering are supported by any other nodes, some such as native renderings must come from the Case Home. However, extracting the contents of container files (such as .zip files) is always done locally on the appliance on which the container file is used.

Review Node Jobs

Review node jobs are executed on the Review node, however, the life cycle of the job is controlled by the Case Home. No state information is stored on the Review node. The worker node can be safely removed at any time.

**Job Execution**

Depending on the type, Jobs will either execute on a review node, processing node, or be redirected to Case Home.

<table>
<thead>
<tr>
<th>Type of Job</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execute on Case Home</td>
<td>Export, Production</td>
</tr>
<tr>
<td>Execute on Review Node</td>
<td>Tagging, Batching, Cache (Review Accelerator), Search, Dashboard Reports</td>
</tr>
</tbody>
</table>
Tips and Troubleshooting

This section provides tips and techniques for resolving issues you may encounter with your Clearwell appliances in a distributed architecture.

Topics in this section:

- “Changing Appliance Settings” on page 43
- “Changing Processing Settings (Required)” on page 46
- “Changing Distributed Review Settings” on page 52
- “Configuring Job Output for Exports” on page 54
- “Addressing Fragmented Nodes” on page 54
- “Addressing Node Failures, and Ensuring System Reliability” on page 57
- “Contacting Technical Support” on page 58

Changing Appliance Settings

With distributed processing and/or review enabled, there are additional appliance, network, and performance enhancement settings, depending on your configuration. Note that some settings in this section are required (as indicated).

Appliance settings in this section:

- “Adjust Appliance Capacity (R710)” on page 43
- “Change the Backup Directory (Required)” on page 44
- “Set System-Level Security and Indexing Properties” on page 44

Adjust Appliance Capacity (R710)

As the system administrator, you can maximize performance of your R710 appliance depending on reviewers and case size. The following steps allow you to increase the capacity of an R710 appliance to ensure handling 25 concurrent reviewers on large cases.

**Note:** Clearwell recommends using this property for large case reviews to avoid delays in the review process.

To increase reviewer capacity on an R710:

1. In Clearwell, go to System > Support Features, and select Property Browser.
2. Enter the property: `esa.common.bitmap.usedirectRef=true`
3. Enable direct reference-based cache by setting this to `false`.
4. Click Submit to save your settings.
Change the Backup Directory (Required)

You must change the directory location for case backups in a distributed architecture deployment, required for all appliances in the cluster.

**To change the backup directory:**
1. In Clearwell, go to **System > Support Features**, and select **Property Browser**.
2. Enter the corresponding property and value for:
   A. Case Backup Directory:
      
      
      esa.case.backupDir=<new backup location>
      
      where `<new backup location>` is the new directory or shared network location where case backups will be stored.
   B. Enable shared Case Backup Directory:
      
      esa.case.sharedBackupDir=true
3. Click **Submit** to save your settings.

Set System-Level Security and Indexing Properties

Use the following properties if you want to enable all appliances in the cluster with SSL, and/or index one or more items (which are disabled by default in a distributed setup).

**Security Setting**

If you want to enable SSL for all appliances in a distributed cluster, then enable HTTPS on each appliance.

**To set system security**
1. In Clearwell, go to **System > Support Features**, and select **Property Browser**.
2. Select the appropriate appliance from the cluster.
3. Enter the corresponding property and value to enable HTTPS:
   
   esa.common.webapp.appserver.transport_guarantee=CONFIDENTIAL
4. Click **Submit** to save your settings.
5. Repeat these steps to enable HTTPS on each appliance in the cluster.

When finished enabling each appliance for HTTPS, import the SSL certificate (as described for a single appliance setup) in "Appendix: Web Services Access Options" in the **System Administration Guide**.

**Indexing**

If you want to index certain types of items (not enabled by default in a distributed setup), then for each appliance, leave the value for the following property blank. Alternatively, to exclude any items from indexing, enter the corresponding value for one or more types of items on each appliance in the cluster.
To include/exclude items to be indexed

1. From the System > Support Features > Property Browser, select the (same or another) appliance in the cluster.

2. Enter the property:

   **esa.crawler.mapi.skipMsgTypes**

   A. To *include* all items, leave the value blank.

   B. To *exclude* one or more items, enter the corresponding values:

   › Contacts:
     IPM.Contact

   › Calendar Item:
     IPM.Appointment

   › Tasks:
     IPM.Task

   › Journal Entries:
     IPM.Activity

   › Post (files):
     IPM.Post

   For two or more items, use comma-separated values. For example, to exclude “Tasks” and “Journal Entries”, enter the property and corresponding values as:

   **esa.crawler.mapi.skipMsgTypes=IPM.Activity,IPM.Task**

3. When finished, click **Submit** to save your settings.

4. Repeat these steps for each appliance in the cluster.

**Note:** You must restart the PST crawler/retriever for any changes to take effect.
Changing Processing Settings (Required)

Note: All properties in this section are required for clustered environments enabled for Distributed Processing in Clearwell.

Distributed Processing settings in this section:

- “Change to Network Location (Email Files)” on page 46
- “Copy Files After Upgrading” on page 47
- “Install PKI Certificates and Enable Support for Signed and Encrypted Messages” on page 47
- “Change Threader Performance in a Distributed Setup” on page 50
- “Disable Retriever Service on Cluster Master” on page 51

Change to Network Location (Email Files)

The location specified for extracted PSTs and NSFs from container files, and converted email files must be a network location that is accessible to all the processing nodes. This setting can be changed at either the case level (Option 1) or System Level (Option 2).

Option 1: Case Level

- Go to Processing > Settings, and under the case settings section “Configure processing parameters and features”, re-type the location to change to a network path for the fields:
  
  A. Extract email files from containers to:
  
  B. Place converted files in:

  This should be a network location that can be accessed by all the processing and review nodes. If a processing node is later added, ensure that the shared location is accessible by the newly-added node. By default, this value points to a location on the “D:” drive.

Option 2: System Level

1. Go to System > Settings, and select the Locations tab.

2. Re-type the location (or click Browse) to change to a network path for both file types:

   A. Extracted Files [Extracted email files (PSTs/NSFs) from containers]

   B. Converted Files

   This should be a network location that can be accessed by all of the processing nodes. Note that the contents of the converted files folder are not automatically backed up or restored by the Clearwell appliance. It is recommended to perform a separate backup.

   CAUTION: For upgraded cases, you must change the value for these properties and then manually copy the files from the previous location to the new network location. Continue with steps “Copy Files After Upgrading” in this section.
Copy Files After Upgrading

**Note:** This applies to distributed architecture environments upon upgrade, only in the following cases:

- You are upgrading from 7.1.1 Fix Packs (to 7.1.2), then to 7.1.2 Fix 1. After upgrading, if you want to move this setup to use distributed architecture, then you will need to copy the files for all cases existing in your setup that used the local D:\ Drive to store contained PST, NSF, and converted files.

- You are upgrading your environment or restoring a case into 7.1.2 Fix 1 in a distributed setup, and the case or environment you are starting with has retained all these files locally, to the D:\ drive.

Follow these steps to set the following Extracted Files and Converted Files properties to a new shared location that can be accessed by all appliances in your distributed architecture deployment.

**To copy files after upgrading**

1. In Clearwell, go to **System > Support Features**, and select **Property Browser**.
2. Select the appliance (case home), on which the case was created.
3. Enter the corresponding property for:
   - A. Extracted files from containers:
     
     `esa.case.extractedFilesDir=<new shared location>`

   - B. Converted files from MBOX:
     
     `esa.system.convertedFilesDir=<new shared location>`

   where `<new shared location>` is the new network location to be accessible by all distributed processing nodes in a distributed processing-enabled environment.
4. Click **Submit** to save your settings.
5. Finally, be sure to copy the files from the old location to the new location you specified in step 3.

Install PKI Certificates and Enable Support for Signed and Encrypted Messages

In a distributed processing environment, all of the PKI certificates must be imported for each appliance in the cluster, for each of the following user accounts: `EsaApplicationService`, `EsaPstRetrieverService`, and the Native Reviewer (`IGC Services`). This will enable support for processing and review of signed and encrypted messages.

First, for two of the three user accounts (`EsaApplicationService` and `EsaPstRetrieverService`) you can install the certificates in one of two ways, depending on your preference:

- **Manual Installation.** See “Option A: Manual Installation” on page 49. (This method is required for installing on utility nodes.)

- **Installation using the Clearwell server.** See “Option B: Installation using Clearwell Server” on page 49.
Next, for the third user account: Native Reviewer (IGC Services), installation must be done through the Clearwell user interface, using the “PKI Certificate Installer” property in **Support Features**. See “Installation using Support Feature [for Native Reviewer (IGCService) user]” in this section.
**Option A: Manual Installation**

If you prefer to use command line, use the following command to install the certificates for both the *EsaApplicationService* and *EsaPstRetrieverService* user accounts.

**To manually install PKI certificates**

**Note:** This method (manual installation) is required for utility nodes, and optional for *EsaApplicationService* and *EsaPstRetrieverService* user accounts.

1. Log in to the Clearwell appliance as the *EsaApplicationService* account user, then from a command line, run the command:
   
   ```
   certutil -f -p <password for PKI certificate> -importpfx <location of PKI certificate file>
   ```

2. Repeat this command for the Retriever account. (Log in as the *EsaPstRetrieverService* user account, and enter the same command.)

   The PKI certificate location should be accessible to the *EsaApplicationService*, and *EsaPstRetrieverService* user accounts. For the Native Review user, use the Support Feature from the Clearwell user interface.

3. (For utility nodes only): Repeat this step for each utility node, for each of the *EsaApplicationService*, *EsaPstRetrieverService* user accounts.

4. Continue to “Installation using Support Feature [for Native Reviewer (IGCService) user]” in this section.

**Option B: Installation using Clearwell Server**

If you prefer to use the Clearwell server to install certificates, follow the steps below for both the *EsaApplicationService* and *EsaPstRetrieverService* user accounts.

**Note:** This method (Clearwell Server installation) cannot be used for installing on utility nodes. Use Option A (manual installation).

**To install certificates from the Clearwell server**

1. Log in to the Clearwell appliance, as the *EsaApplicationService* account user.

2. Double-click on the PKI certificate to import, and provide the password.

3. Follow online instructions to install the certificate, keeping all default settings.

4. Repeat steps 1-3 for the Retriever account. (Log in as the *EsaPstRetrieverService* user account, and follow online instructions.)

   The PKI certificate location should be accessible to the *EsaApplicationService*, and *EsaPstRetrieverService* user accounts. For the Native Review user, use the Support Feature from the Clearwell user interface.

5. Continue to “Installation using Support Feature [for Native Reviewer (IGCService) user]” in this section.
Installation using Support Feature [for Native Reviewer (IGCService) user]

After installing certificates for each of the other two user accounts, follow these steps to install them for the Native Reviewer (IGC Services) user account.

**Note:** This method (Clearwell user interface) is required only for the Native Reviewer (IGC Services) user account.

To install certificates from the Clearwell user interface

1. In Clearwell, go to System > Support Features, and select the property **PKI Certificate Installer**.
2. Select an appliance. (This will need to be repeated for each appliance in the cluster.)
3. Enter the PKI certificate location and password.
4. Enter the User Name (and Domain, if desired) for Native Review.
5. Enter the user password for Native Review.
6. Click **Submit** to save your settings.

Change Threader Performance in a Distributed Setup

By default, the threader property to distribute threader-specific converted files, or ITD files is enabled (as of 7.1.2). In a distributed architecture setup, if necessary, you can prevent these ITD files from being distributed. Use the following case-level configuration to ensure that the converted files directory—where the ITD files reside—points to a network shared path. (See also “Change to Network Location (Email Files)” in this section.)

**Note:** It is recommend that these configurations not be changed while the Discussion Threader is running.

To change this threader setting across all appliances in the system, you can set it at the system level; however, the case-level configuration will always override system settings.

To disable ITD Distribution

1. In Clearwell, go to System > Support Features, and select **Property Browser**.
2. Enter the property: `esa.threader.ITD.distribution.enabled=false`
3. Click **Submit** to save your settings.
Disable Retriever Service on Cluster Master

**Note:** This applies only if you are not using the Cluster Master node for any cases, or as a review or processing node.

The Clearwell installer does not disable the retriever service, which is needed for a distributed architecture deployment. To ensure PST and NSF retrievals will not fail, use the following properties to disable this service on the Cluster Master.

**To disable Retriever service on Cluster Master**

1. In Clearwell, go to **System > Support Features**, and select **Property Browser**.
2. Select the System.
3. Enter the corresponding property for:
   A. PSTs: `esa.if.dotmsg.parallel.crawling`
   B. NSFs: `esa.processing.nsf.retrieval_parallelism`
4. Set the new value for both to 0.
5. Select the check box to confirm your change.
6. Click **Submit** to save your settings
Changing Distributed Review Settings

Use the properties in this section to optimize text and native review processes during export and production jobs, or automate (or disable) load balancing.

Distributed Review settings in this section:

- “Disable File Conversions on the Cluster Master” on page 52
- “Automate Load Balancing” on page 53
- “Turn Off Load Balancing” on page 53

Disable File Conversions on the Cluster Master

If you prefer not to use the Cluster Master for any file conversions during text and native review, or during export and production jobs, use these properties set each file type conversion to “0” using the Property Browser.

To disable file conversions on the Cluster Master

1. In Clearwell, go to System > Support Features, and select Property Browser.
2. Select the System.
3. Enter the corresponding property for:
   A. Docs:
      esa.netit.processing.jpthreads.doc
   B. PDFs:
      esa.netit.processing.jpthreads.pdf
   C. Drawings:
      esa.netit.processing.jpthreads.drw
   D. Stellent Resources:
      esa.utility.stellent.resources
4. Set the new value for all to 0.
5. Select the check box to confirm your change.
6. Click Submit to save your settings
Automate Load Balancing

Automate user redirection manages how the review load is distributed to all review nodes. Use the auto-redirection property to setting load balancing to occur automatically. This setting applies to the entire cluster, and is not case-specific.

To automate load balancing

1. In Clearwell, go to System > Support Features, and select Property Browser.
2. Select the appliance (case home), on which the case was created.
3. Enter the property: \texttt{esa.autoredirect=auto}
   (where \texttt{auto} is the default setting.)
4. Click \textbf{Submit} to save your settings.

Turn Off Load Balancing

Use the following property if you want to disable load balancing for users with the Case Admin role or privileges. Once completed, case administrators will always be redirected to the Case Home.

Alternatively, to completely turn off load balancing for all reviewers, set the following property to “never”.

\textbf{Note:} Turning off load balancing completely only applies if you want to give each subset of case users a specific review node to log in to for review, and you do not want load balancing enabled nor users redirected to log in to the Case Home.

To redirect login and disable load balancing

1. In Clearwell, go to System > Support Features, and select Property Browser.
2. Enter the property according to your preference:
   A. To redirect Case Admin users:
      \texttt{esa.autoredirect=onlyadmin}
   B. To turn off completely for all reviewers:
      \texttt{esa.autoredirect=never}
3. Click \textbf{Submit} to save your settings.
Configuring Job Output for Exports

In a distributed setup, to avoid export and print job failures, set the following property to specify the location for job output.

**To set output location for print and export jobs**

1. In Clearwell, go to **System > Support Features**, and select **Property Browser**.
2. Select the appliance (case home), on which the case was created.
3. Enter the property: `esa.export.joboutput.external=<directory path>`
   where `<directory path>` is the UNC path where export job output will be sent (in a distributed architecture deployment).
4. Click **Submit** to save your settings.

Addressing Fragmented Nodes

In a Clearwell distributed architecture, a **Fragmented** node is one in which case processing did not complete or failed on those nodes and the case data is still spread across multiple nodes.

Clearwell provides a one-click action (done in the Clearwell user interface) to prepare all cases “homed” on the selected node. This prompts all processing nodes to return their unfinished processing data to the Case Home appliance. Contact Clearwell Technical Support for guidance with the following procedure, or help with troubleshooting if you have fragmented cases on nodes, or encounter node failures. See also “Addressing Node Failures, and Ensuring System Reliability” on page 57.

Refer to the following topics in this section:

- “Prepare to Backup, Restore, or Remove Nodes” in this section
- “Precautions for Node Backups with Fragmented Cases” on page 56
- “Precautions for Node Restore or Removal” on page 56
Prepare to Backup, Restore, or Remove Nodes

Before attempting to back up the actual node or restore a node, you must first ensure the node is not in a fragmented state, and must be defragmented. (Actual appliance backups are done through the desktop Clearwell Utility or using a command line script. Refer to the section "Backup and Restore" in the System Administration Guide.)

To prepare nodes for backup, restore, or removal
1. In Clearwell, click **System**, and select **Settings > Appliances**.
2. Click the appliance you want to prepare for backup then click the **Appliance** tab.
3. On the **Appliance** tab, under the section “Prepare the appliance for the following actions”, click one of the following actions, depending on its state:
   - **Prepare for Backup** (to back up the node)
   - **Prepare for Restore** (to restore the node)
   - **Prepare to Remove Node** (to remove nodes if other than the Case Home)

**Note:** The Case Home appliance cannot be removed. If the node is the Case Home, the button “Not Ready for Remove Node” appears, but is unavailable (greyed out). See also “Precautions for Node Backups with Fragmented Cases” and “Precautions for Node Restore or Removal” on page 56.

4. Click **Save** (if you changed the appliance name) or to go back to original settings, click **Cancel**.
Precautions for Node Backups with Fragmented Cases

For cases in a fragmented state, manual or scheduled case backup jobs will not fail. However, node backups will fail to complete successfully if the node contains any fragmented cases. Clearwell attempts to defragment the case at the start of the backup job.

If the node being backed up contains fragmented cases, then you must run a "Prepare for Backup" job prior to starting the node backup. (See "Prepare to Backup, Restore, or Remove Nodes" on page 55.) Unlike a case backup, the node backup job does not attempt to defragment. If the “Prepare” job fails, (for example, if the node contains a missing fragment due to an offline node), contact Clearwell Technical Support for further assistance.

Precautions for Node Restore or Removal

For both node restore and node removal, you must run the appropriate task (“Prepare for Restore” or “Prepare Node for Removal”) prior to performing the actual restore or removal operation. Not doing so will cause cases on the node to become invalid, and cannot be further processed. If any of the prepare jobs fails, or cannot be run, contact Clearwell Technical Support for further assistance.

See also “Addressing Node Failures, and Ensuring System Reliability” on page 57. Contact Clearwell Technical Support for assistance regarding a full node or cluster restoration.
Addressing Node Failures, and Ensuring System Reliability

Since disk or appliance failures in any system is unavoidable, administrators should try to minimize the risks and recover from the failure quickly. This section describes a few scenarios, and recommended recovery methods and/or safeguards against these types of failures.

About Backups

Symantec Clearwell provides support in Distributed Architecture backups in the same way as single-node deployments using node backups and case backups. Having a remote database does not introduce any additional maintenance and setup for backing up the cluster. In addition, ensure all cases and nodes are backed up in Clearwell, using periodic on-demand backups or scheduled backups, as often as necessary to be able to restore data back to the stand-alone server.

Cluster Backups

Cluster backups are a combination of backups that enable you to restore the cluster. There are two ways of backing up a cluster:

1. Appliance backups for each appliance
2. Case backups for all cases AND a system backup

(Take appropriate precautions when running scheduled backups. See “Precautions for Node Backups with Fragmented Cases” on page 56.)

Cluster Master Failure

To safeguard against Cluster Master failures, especially if the Cluster Master is a separate node without any cases on it, nor being used for processing and review, it is recommended to perform periodic system and node backups. In the event of a failure, the Cluster Master can then be restored from your most recent system backup (only if no cases are “homed” on this node, or this node was not used for processing and/or review).

Case Home Failure

A Case Home failure will render the case unusable for review or additional processing. To avoid loss of productivity, Clearwell recommends periodically backing up the case. Refer to “Creating Case Backups” in the System Administration Guide. If the Case Home fails, the case can be restored to a different node, allowing review and processing to continue. Some work may not be lost, but this option minimizes the amount of data loss.

If the Case Home is also the Cluster Master, it may be more difficult to recover as many cases. Additionally, the cluster information will take longer to recover. Even performing a case restore from a recent backup could potentially take a long time depending on the size of the case.
Processing Node Failure

If a failure occurs on a processing node (due to a network outage, or stopped services for example), the processing jobs on that node will fail. Once the network or services are restored, you may re-run processing. However, if the processing node failed due to a disk failure, and services cannot be restarted to be able to access the indexed files, the cases on the processing node will be in an Invalid state, and cannot be further processed.

To be able to restore cases, particularly after any unforeseen failures, best practice is to periodically perform a case backup so that processing may be re-started from a newly-restored case. Refer to “Creating Case Backups” in the System Administration Guide.

Contacting Technical Support

See “Contacting Technical Support” on page 6 for specific instructions on what information to have available, and your preferred method for contacting Technical Support.