Veritas eDiscovery Platform™

Distributed Architecture Deployment Guide

10.1

VERITAS™

Veritas 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: 2021-11-20.

Legal Notice

Copyright © 2021 Veritas Technologies LLC. All rights reserved.

Veritas and the Veritas Logo are trademarks or registered trademarks of Veritas Technologies LLC or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

This product may contain third-party software for which Veritas is required to provide attribution to the third party ("Third-Party Programs"). Some of the Third-Party Programs are available under open source or free software licenses. The License Agreement accompanying the Software does not alter any rights or obligations you may have under those open source or free software licenses. Refer to the Third-Party Legal Notices for this product at:

https://www.veritas.com/about/legal/license-agreements

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 Veritas Technologies LLC 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. VERITAS TECHNOLOGIES LLC 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, whether delivered by Veritas as on premises or hosted services. 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.

Veritas Technologies LLC
2625 Augustine Drive
Santa Clara, CA 95054
http://www.veritas.com

Contents

About This Guide 5

Revision History 5
Technical Support 8
Documentation 8
Documentation Feedback 8

Overview 9

Distributed Architecture 9
Basic Concepts 10
Features of a Distributed Architecture 11
How a Distributed Architecture Works 12
How are Cases Processed using Distributed Processing? 13
How Distributed Review Works 15
Using Utility Nodes to Improve Review Performance 16
Configuration Considerations 18

Setting Up your Environment 19

Setting up a Cluster Environment 19
Sample Distributed Architecture 21
Setup Requirements 22
Hardware Requirements 22
Appliance Capacity 25
Changing Settings 26
About Load Balancing Thresholds (for Review) 26
Disabling Load Balancing 26
Adjust Appliance Capacity (128GB RAM System) 26
Set System-Level Security and Indexing Properties 27
Frequently Asked Questions 29

Installation and Configuration 31

Install DBMS for the Shared Remote Database 31
Appliance Migration 32
Other Required Cluster Configurations 33
Configure Shared Remote Database in a Cluster 33
Change Backup Directory Location (Required) 33
Image Helper Document Converter Service Settings 34
Changing Processing Settings (Required) 34
Change to Network Location (Email Files) 35
Copy Files After Upgrading 36
Install PKI Certificates and Enable Support for Signed and Encrypted Messages 36
Change Threader Performance in a Distributed Setup 37

Disable Text Extraction, Imaging, Retrieval, and Classification Sharing on Cluster Primary node **39**

Appliance Assignment and Provisioning **41**

Overview **41**
Setting System-Level Properties on all Appliances in the Cluster **41**
Administrator Workflow **43**
Step 1: Add Appliances to the Cluster **43**
Step 2: Assign Roles **45**
Step 3: Provision Appliances (Review and Processing Roles) **46**
Distribution Enabled **48**
Reassigning or Unassigning Roles **48**

Appliance Management **49**

Managing Appliances **49**
Managing Cases (Processing Role Only) **49**
Processing Performance Considerations **50**
Managing Cases (Review Role Only) **51**
Review Performance Considerations **51**
Check Status Before Backups **52**
About Case Backups **52**
About Node (Appliance) Backups **52**
Prepare Nodes for Backup, Restore, or Removal **52**
Restoring a Node **54**
Review Cache Process **54**
File Caching **55**
Review Node Jobs **55**

Tips and Troubleshooting **57**

Changing Distributed Review Settings **58**
Automate Load Balancing **58**
Turn Off Load Balancing **58**
Configuring Job Output for Exports **59**
Addressing Fragmented Nodes **60**
Prepare to Backup, Restore, or Remove Nodes **61**
Precautions for Node Backups with Fragmented Cases **62**
Precautions for Node Restore or Removal **62**
Addressing Node Failures, and Ensuring System Reliability **63**
About Backups **63**
Cluster Backups **63**
Cluster Primary node Failure **63**
Case Home Failure **63**
Processing Node Failure **64**

Appendix A: Product Documentation **65**

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*.

This section contains the following sections:

- *"About This Guide" in the next section*
- *"Revision History" on page 5*
- *"Technical Support" on page 8*
- *"Documentation" on page 8*
- *"Documentation Feedback" on page 8*

About This Guide

This is intended to assist system administrators perform:

- Installation and configuration of their clustered environment.
- Administration tasks to enable the Review and Processing scalability features: *Distributed Review* and *Distributed Processing*.

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.

Revision Date	New Information
December 2021	• Updated version for release 10.1 • Added a note on removal of Distributed Architecture support for OS having separate third parties. See *"Setup Requirements" on page 22*
July 2021	• Updated the hardware requirements
March 2021	• Updated the hardware requirements • Updated the *Disable Text Extraction, Imaging, Retrieval, and Classification Sharing on Cluster Primary node* section • Minor edits
March 2020	• Updated the hardware requirements • Minor edits
October 2018	• Minor edits
March 2018	• Added information related to IPv6 support

Revision Date	New Information
June 2017	• Update copyright and release number • Clarification of Image Helper Converter Settings recommendations for node with Cluster Primary node.
July 2016	• Modified obsolete references to explicitly enable esa.threader.ITD.distribution.enabled. This is default behavior. Corrected Threader Performance section so that there is no implied association that the location of converted files can be configured by the property. • Updated branding information.
August 2015	• Removed Rights Management Guide. • Changed Case Home description on page 14 to remove "or the Cluster Primary node itself." • Noted Cluster Primary node hardware is given for minimum configuration and VM performance is slightly lower than physical machines on page 25. • Clarified Medium and Large configurations on page 27 and corrected concurrent reviewer numbers on page 28. • Updated "Configuring Job Output for Exports" with Image Helper domain account information. • Added Windows Server 2012 R2 (Standard or Data Center edition) as supported OS.
March 2015	• Image accessibility • Added note that remote database installer must be run even if it is located on the same node as the Cluster Primary node. • Added requirement in 8.0 that when Cluster Primary node and Remote Database are on the same node, it should be upgraded first in the cluster • Branding and minor edits • Sub-node installation title change
October 2014	• Updated graphics and screen shots • Removed statement requiring all appliance models use the same configuration • Added reference to "Deployment Guide for eDiscovery on VMs" white paper • Updated statement about number of concurrent reviewers, limit to number of cases per appliance • Added permissions considerations to "Change the Backup Directory (required)" • Updated "Other Required Cluster Configurations" with Image Helper information.
March 2014	• Branding edits

Revision Date	New Information
June 2013	• Added *"Other Required Cluster Configurations" on page 33*" • Updated DBMS instructions (MySQL):*"Install DBMS for the Shared Remote Database" on page 31* • Added note that all appliances in a cluster should be configured with the same date/time. • Clarified Administrator Workflow for cases with LFI sources: *"Assigning Appliance Roles for Cases with Load File Import Sources" on page 45* • Added explanation of how index consolidation removes duplicate information when review is taking place in multiple nodes: *"Setting System-Level Properties on all Appliances in the Cluster" on page 41* • Added another step to "Adjust Appliance Capacity (R710)": *"Adjust Appliance Capacity (128GB RAM System)" on page 26* • Added "Disable Text Extraction, Imaging, and Retrieval Sharing on Cluster Primary node" (when Cluster Primary node is on the same node as the remote database): *"Disable Text Extraction, Imaging, Retrieval, and Classification Sharing on Cluster Primary node" on page 39*
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 22*. – Threader property now *enabled* by default. See *"Change Threader Performance in a Distributed Setup" on page 37*.
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 Veritas eDiscovery Platform 7.1.1. • Updated 7.1.3 changes to Remote Database Management System installer (and MySQL details). • Clarified assigning appliance roles when a case has a LFI source • Changes added to "Adjust appliance capacity R710" and "Disable Text Extraction, Imaging, and Retrieval Sharing on Cluster Primary node" (7.1.3 for CM not used as a review or processing node)
Feb 2012	• New guide for distributed architecture deployments of Veritas eDiscovery Platform 7.0 enabling the Distributed Review feature.

Technical Support

Technical Support maintains support centers globally. All support services will be delivered in accordance with your support agreement and the then-current enterprise technical support policies.

For information about our support offerings and how to contact Technical Support, visit our website:

https://www.veritas.com/support

You can manage your Veritas account information at the following URL:

https://my.veritas.com

If you have questions regarding an existing support agreement, please email the support agreement administration team for your region as follows:

Worldwide (except Japan)	CustomerCare@veritas.com
Japan	CustomerCare_Japan@veritas.com

Documentation

Make sure that you have the current version of the documentation. The latest documentation is available from:

- **Documentation** link at the bottom of any page in the Veritas eDiscovery Platform landing page.
- **Products Web site:** https://www.veritas.com/product

Documentation Feedback

Your feedback is important to us. Suggest improvements or report errors or omissions to the documentation. Include the document title, document version, chapter title, and section title of the text on which you are reporting. Send feedback to:

eDiscovery.InfoDev@veritas.com

You can also see documentation information or ask a question on the Veritas community site:

https://vox.veritas.com//

Overview

This section provides an introduction to the Distributed Review and Processing features of the eDiscovery Platform.

Refer to the following topics in this section:

- *"Distributed Architecture" in the next section*
  - *"Basic Concepts" on page 10*
  - *"Features of a Distributed Architecture" on page 11*
- *"How a Distributed Architecture Works" on page 12*
  - *"How are Cases Processed using Distributed Processing?" on page 13*
  - *"How Distributed Review Works" on page 15*
- *"Using Utility Nodes to Improve Review Performance" on page 16*
- *"Configuration Considerations" on page 18*

Distributed Architecture

The eDiscovery Platform has 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 Review** allows for more concurrent reviewers than one appliance alone could support. ("Role", in this context, refers to the role assigned to a node, not the roles available to users.)
- **Distributed Processing** allows administrators to add nodes at any point during case processing to optimize processing performance.
- **Parallel Processing**. Optimized processing via on-demand added throughput.
- **Reviewer Scalability**. Dynamic increased number of supported users

Distributed Processing and Review Support for large volume cases to be processed, and multiple reviews to be distributed across one or more appliances. The platform'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.

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.

How they are employed depends on how the appliances in the cluster are assigned. For example, you can assign the "Review role" to multiple appliances which then opens case review to multiple reviewers across more than one appliance.

Basic Concepts

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

Distributed Architecture Basic Concepts

Term or Concept	Description
Appliance / Node	An eDiscovery Platform, or eDiscovery Platform-supported appliance, referring to an appliance which has version 7.x or later of the eDiscovery software installed. All appliances must be running the same version. **Note:** In this guide, the term *appliance* and *node* are used interchangeably.
Case Home	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.
Cluster / Clustered Environment	A group of appliances/nodes in a distributed architecture, or *clustered environment*.
Cluster Primary node / Primary node	The Cluster Primary node 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 Primary node'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. **Note:** it is not recommended to have cases on the Cluster Primary node. The Primary is system-critical and must be continuously up and running. If the Primary is shut down, it automatically shuts down all other Secondary nodes in the cluster. See *Secondary Node.*
Database Node	In a distributed processing setup, the database node serves as a single database instance for the entire cluster.
Processing Node	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 Secondary nodes of the Cluster. While the Cluster Primary node can be a processing node, it should be kept light weight. Thus, assigning the Processing Role to the Cluster Primary node 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, consider processing each case at different times on the same processing node.
Review Node	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 Secondary node of the Cluster. 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.

Distributed Architecture Basic Concepts

Term or Concept	Description
Secondary Node	An individual appliance or node that is managed by the Cluster Primary node in a clustered environment. Secondary nodes can be assigned different roles as a way of optimizing performance by handling case review and/or processing jobs as one of multiple nodes in a cluster.
Utility Node	Any appliance, not part of the cluster, used to perform retrievals, image conversions, or file conversions. Using utility nodes for these tasks helps improve performance when users run jobs such as review cache, metadata export or production, or production exports.

Features of a Distributed Architecture

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

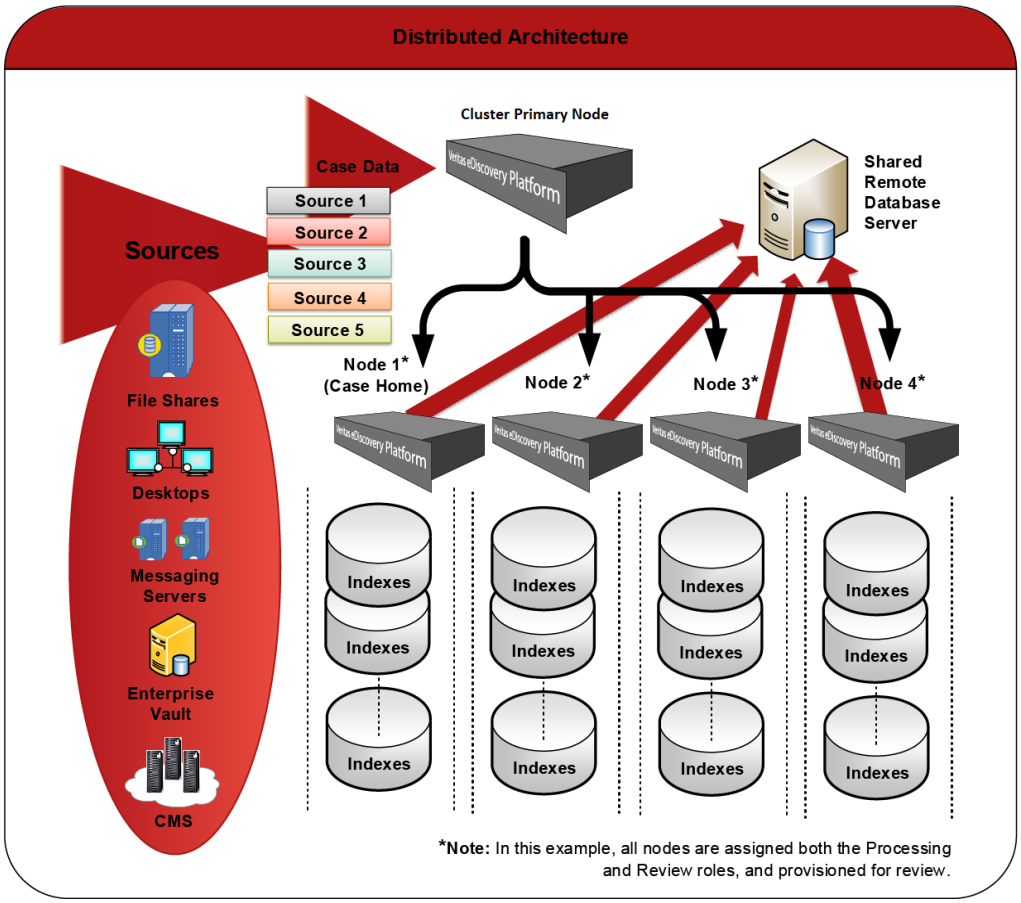
- Dynamically increases the scalability (number of reviewers, and processing speed), to maximize the benefits of the review process using the appliance, with:
  - No impact on review or processing capacity of a given appliance
  - Search, analysis and review operations occurring on multiple nodes (as 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 if 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 a Distributed Architecture Works

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

A Sample Distributed Architecture Environment



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
  - the main components of a case are maintained.
- A central location where any information is checked/updated in the case for all the allocated processing and review nodes.
- A unified, consistent view of the case regardless of the current state of processing, review, and exports.
- A single cluster, along with all its processing nodes. (See *"Configuration Considerations" on page 18* for a comprehensive list of key terms.)

Using a Shared Remote Database Server

The eDiscovery 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
- 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 primary 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, and NSF 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, the system creates appropriate entries in the database. In addition, the system maintains searchable text indexes locally, on each node assigned and provisioned for review.

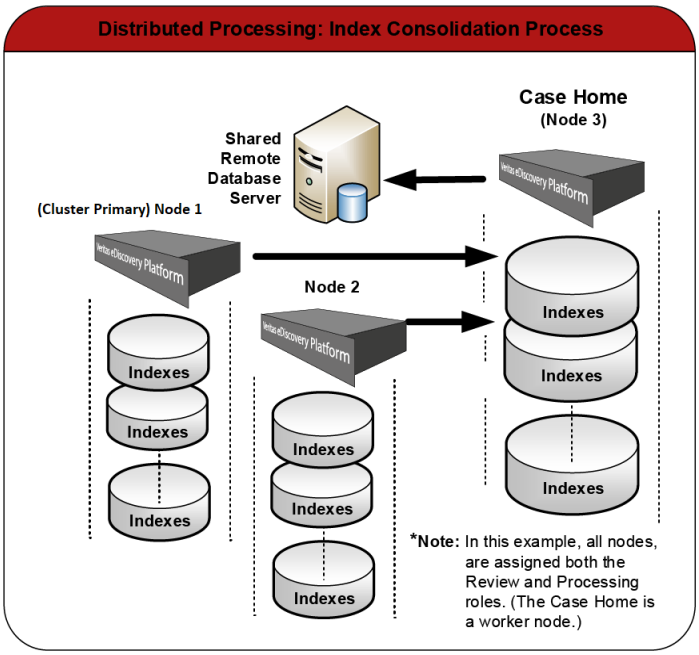
Post-Processing and Index Consolidation

When the platform completes processing of all sources (in which data is de-duplicated and indexed), the Post-Processing phase begins. Post-Processing includes:

- Index validation
- Concept search (if enabled)
- Image analysis.
- Threading
- Search analytics
- Statistics
- Distributed merge
- Centralized merge and consolidation builds

Specifically, the platform'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



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 Distributed Review Works

Scaling Reviewers across Multiple Appliances

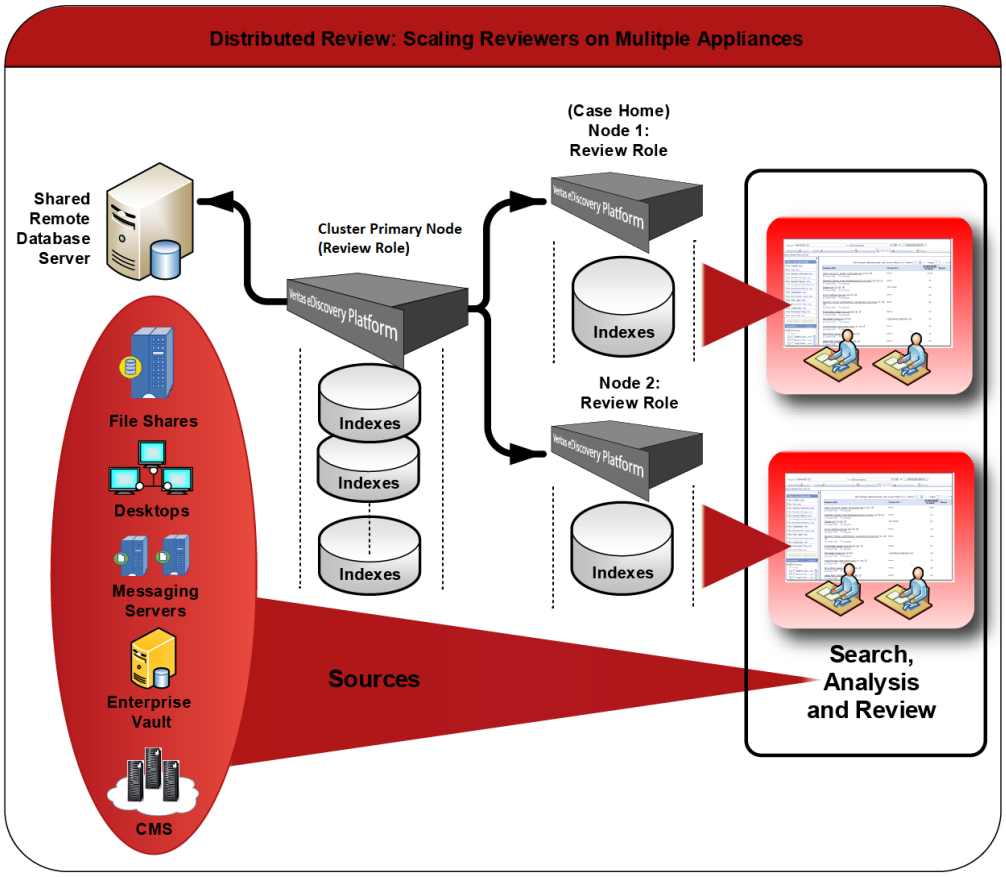
Once processing is completed, review scalability components can be enabled for increasing the number of reviewers that perform searches and create work product (such as tagging, redactions, document comments, review comments). To increase review scalability, administrators can:

- add new appliances
- assign appliances the Review role
- provision appliances 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.

Distributed Review: Scaling Reviewers on Multiple Appliances



The Distributed Review Workflow

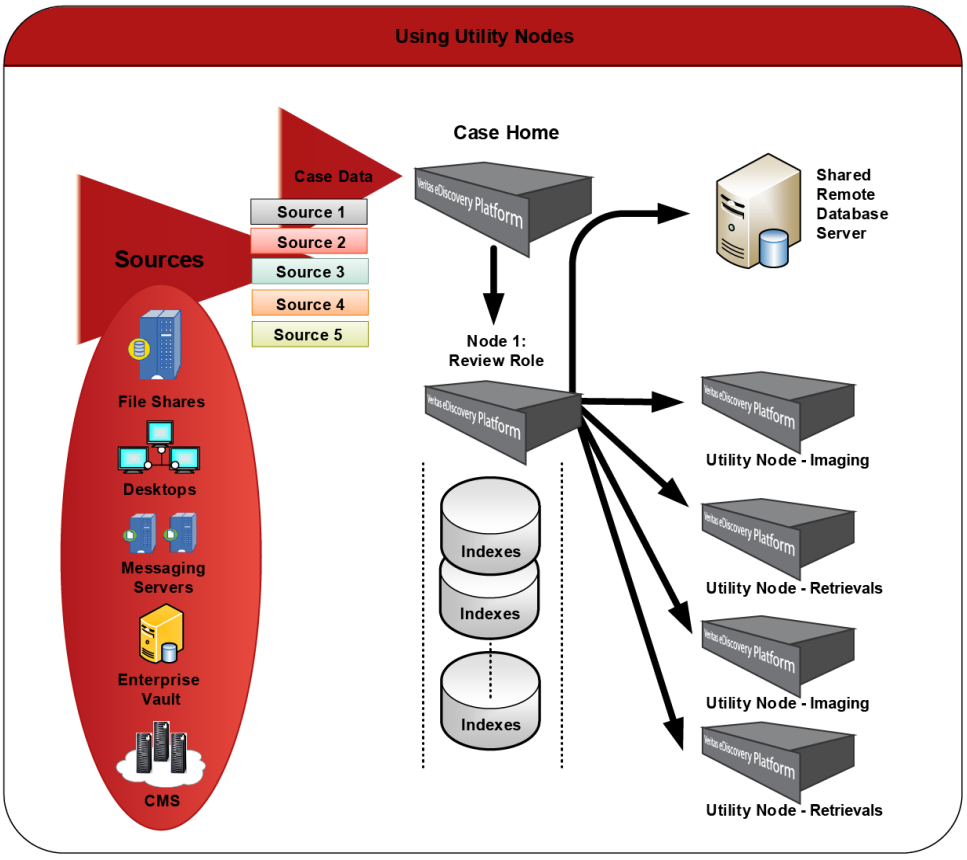
When you provision the node for Review, the Case Home replicates its view of all its searchable indexes to dedicated Review nodes. Each Review node presents a read-only copy of the search index to all reviewers. When a reviewer performs a search, the search results are based on the copy of the index 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 Improve Review Performance

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, review cache, metadata export, or production, or production export jobs.

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

Using Utility Nodes



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:

- 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/192 GB RAM, 50 users for an appliance with 32-Core CPU/128 GB RAM, and 25 users for an appliance with 16-Core CPU/64 GB RAM.

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

- 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, the eDiscovery Platform ships 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. For hardware-specific considerations, see *"Hardware Requirements" on page 22*.

For help determining required components for your architecture, configuring your appliances to use a shared remote database server, contact your Solutions Consultant for details.

Setting Up your Environment

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

Refer to the following topics in this section:

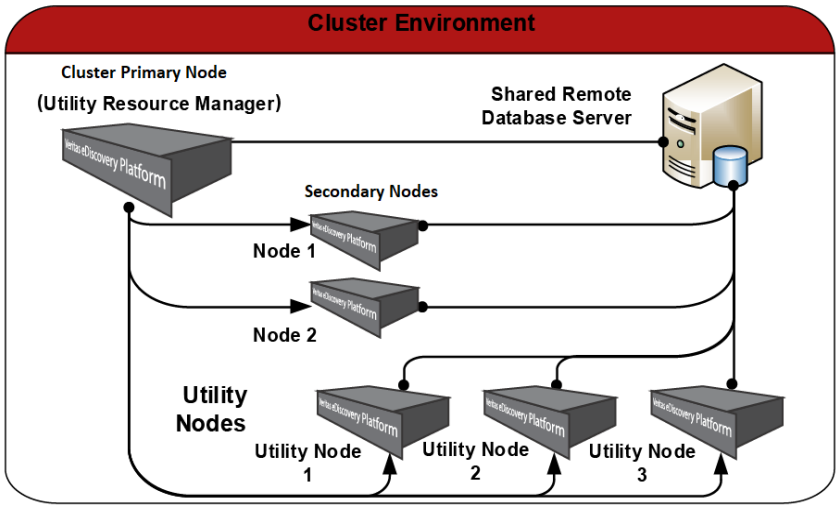
- *"Setting up a Cluster Environment" in the next section*
- *"Sample Distributed Architecture" on page 21*
- *"Setup Requirements" on page 22*
  - *"Hardware Requirements" on page 22*
  - *"Appliance Capacity" on page 25*
- *"Changing Settings" on page 26*
  - *"About Load Balancing Thresholds (for Review)" on page 26*
  - *"Disabling Load Balancing" on page 26*
  - *"Adjust Appliance Capacity (128GB RAM System)" on page 26*
  - *"Set System-Level Security and Indexing Properties" on page 27*
- *"Frequently Asked Questions" on page 29*

Setting up a Cluster Environment

A cluster environment is an architecture consisting of multiple appliances, in which one is designated as the Cluster Primary node and the other member appliances are the *Secondary nodes* and if utilized, the *utility nodes*.

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

Cluster Environment



This configuration allows the system 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 Primary node 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*.

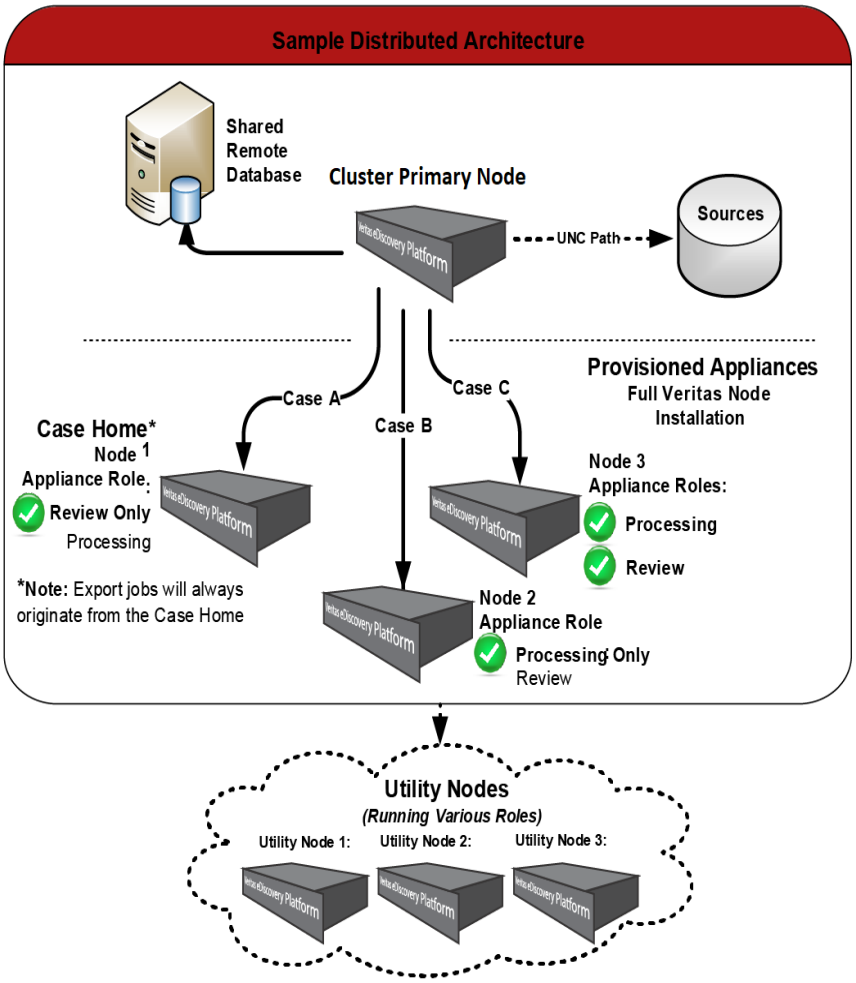
Note: Earlier, eDiscovery Platform only supported IPv4 address format. Starting with 9.0.1, end-to-end workflow of eDiscovery Platform supports IP addresses in IPv4, IPv6, and dual stack. While configuring eDiscovery Platform in Distributed Architecture mode, all appliances including the Cluster Primary node and member Secondary nodes, Utility node, Shared Remote Database, and Legal Hold Confirmation server must be configured with a consistent IP address format. All of them should have either IPv4 or IPv6 or dual stack format. When eDiscovery Platform deployment is done in dual stack, the remote clients can have any of these address formats. When eDiscovery Platform deployment is done in either IPv4 or IPv6 mode, then the IP addresses configured on the remote clients must match with the IP address format used in the eDiscovery Platform deployment.

Note: In a clustered environment enabled for Distributed Review, once a case restore job is started on a Secondary 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 Primary node, and three Secondary 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.

Sample Distributed Architecture Environment



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 13* 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 58*.

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
- Ensure MySQL database server in the distributed architecture deployment meets requirements.

Note: Earlier, eDiscovery Platform used to support hybrid cluster with nodes having Windows Server 2012 and Windows Server 2016. But for Windows Server 2016 and Windows Server 2019, hybrid cluster or distributed architecture is not supported. Because, both OS have separate third parties—Windows Server 2016 has Office 2013 and Windows Server 2019 has Office 2019.

Hardware Requirements

Hardware Compatibility

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

Note: Release 10.0 replaces the IGC Native Viewer with PrizmDoc Viewer, which introduced several performance improvements in review, redaction, annotation, and other imaging-related operations. However, these enhancements brought in the need for higher hardware requirements. The minimum recommended hardware requirements vary based on whether the appliance will be used for any Imaging related jobs with the Imaging role enabled.

Setup with no Imaging Related Operations

Appliance Type	Environment	Appliance Function	CPUs RAM	Disk	VM*
All-In-One (Standalone)	Standalone	*Hosts Cases and Review and performs Processing.* *Includes Legal Holds and Collections*	32-Core CPU x 128 GB RAM	1.5 TB (1500 IOPs)	Y
Utility		*Retrieval and HTML caching*	4-Core CPU x 8 GB RAM	500 GB on D:	Y
Case Home and Processing	Distributed Architecture	*Hosts Cases and Review and performs Processing*	24-Core CPU x 96 GB RAM	1.5 TB (1500 IOPs)	Y

Setup with no Imaging Related Operations

Appliance Type	Environment	Appliance Function	CPUs RAM	Disk	VM*
Review and Processing	Distributed Architecture	*Hosts Review and performs Processing*	24-Core CPU x 96 GB RAM	1.5 TB (1500 IOPs)	Y
Case Home Only	Distributed Architecture	*Hosts Cases without Processing and Review*	16-Core CPU x 64 GB RAM**	1.5 TB (1500 IOPs)	Y
Review Only	Not Applicable				
Processing Only	Distributed Architecture	*Performs Processing*	16-Core CPU x 64 GB RAM**	1.5 TB (1500 IOPs)	Y
Cluster Admin	Distributed Architecture	*Administers the Distributed Architecture cluster and hosts Legal Holds and Collections*	32-Core CPU x 128 GB RAM	1.5 TB (1500 IOPs)	Y
Cluster Admin/ Database	Distributed Architecture	*Hosts the Distributed Architecture shared database, administers the Distributed Architecture cluster and hosts Legal Holds and Collections*	32-Core CPU x 192 GB RAM	1.5 TB (1500 IOPs)	Y
Shared Remote Database MySQL	Distributed Architecture	*Database Server*	24-Core CPU x 128 GB RAM	1.5 TB (1500 IOPs)****	N***

Note: The number of cores in the above table is the total number of logical processors (with hyper-threading enabled) on the machine, and not just the physical cores.

* VM performance is lesser than that of a physical machine.

** This is the minimum required configuration, but the recommended configuration is 32-Core CPU.

*** A physical machine is recommended, but a VM of equivalent performance (on dedicated hardware) can be used if required by local IT policies.

**** You might need 2 TB or higher if the number of cases and items is too high and multiple concurrent activities happen on the system.

IMPORTANT! Existing appliances can be repurposed for a distributed architecture deployment. For information about other appliance types, contact your Solutions Consultant, or Technical Support.

The following specifications are for eDiscovery Platform deployments that are used for native review and jobs such as Image Caching, Production/Export, and Bulk redaction.

Setup Supporting Imaging Operations (With the Imaging role enabled)

Appliance Type	Environment	Appliance Function	CPUs RAM	Disk	VM*
All-In-One (Standalone)	Standalone	*Hosts Cases and Review and performs Processing.* *Also includes Legal Holds and Collections*	32-Core CPU x 128 GB RAM	1.5 TB (1500 IOPs)	Y
Utility		*Caching, Retrieval, Export*	8-Core CPU x 32 GB RAM	500 GB on D:	Y
Case Home and Processing	Distributed Architecture	*Hosts Cases and Review and performs Processing*	24-Core CPU x 96 GB RAM	1.5 TB (1500 IOPs)	Y
Review and Processing	Distributed Architecture	*Hosts Review and performs Processing*	24-Core CPU x 96 GB RAM	1.5 TB (1500 IOPs)	Y
Case Home Only	Distributed Architecture	*Hosts Cases*	24-Core CPU x 96 GB RAM**	1.5 TB (1500 IOPs)	Y
Review Only	Distributed Architecture	*Hosts Review*	24-Core CPU x 96 GB RAM	1.5 TB (1500 IOPs)	Y
Processing Only	Distributed Architecture	*Performs Processing*	16-Core CPU x 64 GB RAM**	1.5 TB (1500 IOPs)	Y
Cluster Admin	Distributed Architecture	*Administers the Distributed Architecture cluster and hosts Legal Holds and Collections*	32-Core CPU x 128 GB RAM	1.5 TB (1500 IOPs)	Y
Cluster Admin/ Database	Distributed Architecture	*Hosts the Distributed Architecture shared database, administers the Distributed Architecture cluster and hosts Legal Holds and Collections*	32-Core CPU x 192 GB RAM	1.5 TB (1500 IOPs)	Y
Shared Remote Database **MySQL**	Distributed Architecture	*Database Server*	24-Core CPU x 128 GB RAM	1.5 TB (1500 IOPs)****	N***

Note: The number of cores in the above table is the total number of logical processors (with hyper-threading enabled) on the machine, and not just the physical cores.

* VM performance is lesser than that of a physical machine.

** This is the minimum required configuration, but the recommended configuration is 32-Core CPU.

*** A physical machine is recommended, but a VM of equivalent performance (on dedicated hardware) can be used if required by local IT policies.

**** You might need 2 TB or higher if the number of cases and items is too high and multiple concurrent activities happen on the system.

IMPORTANT! Existing appliances can be repurposed for a distributed architecture deployment. For information about other appliance types, contact your Solutions Consultant, or Technical Support.

Appliance Capacity

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

Appliance Capacity

Capacity	Detail	Notes
Case Size	*Recommend 5 million reviewable items per case with a total of 10 million items per appliance*	• Can be in a single case or split across multiple cases • Data characteristics also matter. Cases with numerous unique terms (typically found in attachments) have larger overhead **Note:** "Reviewable items" include email, attachments, or loose files, but do not include duplicates. For a detailed explanation, see *"Frequently Asked Questions" on page 29*.
Concurrent Reviewers	*Small appliance:* *16-Core CPU/64 GB RAM* *Medium appliance:* *32-Core CPU/128 GB RAM* *Large appliance:* *48-Core CPU/192 GB RAM*	• Capacity of 25 concurrent reviewers on small appliance, 50 concurrent reviewers on medium appliance, and 100 concurrent reviewers on large appliance **Note:** The capacity of concurrent reviewers varies based on jobs that are running, such as collection, processing, caching, bulk redaction, production, and export.

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.

The system calculates the percentage of additional load that each node will balance before it will redirect a user to another node. The system 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 58* in the Tips and Troubleshooting section.

Adjust Appliance Capacity (128GB RAM System)

As the system administrator, you can maximize performance of your hardware depending on reviewers and case size. The following steps allow you to increase the capacity of a 128GB RAM system to ensure handling 50 concurrent reviewers on large cases.

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

To increase reviewer capacity on a 128GB RAM system:

1. In the user interface, 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 changing this setting to **`false`**.
4. Click **Submit** to save your settings.
5. Once this property has been modified, add the following property (if it is not there already) to the file *default. properties*, located in <installed drive>:\CW\v<installation version>\config\configs\ :
   `esa.bitmap.cache.size.x64m128=6000`

   Note: The property should be reset before and after upgrade.

   A. Stop services.

   B. Run **`b deploy-props`**

   C. Run **`b tomcat-init`**

   D. Restart services.

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 the user interface, 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 A: 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.
5. You must restart the PST crawler/retriever for any changes to take effect.

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 items?*
- *What is the limit for the total number of items, users, and cases on an appliance?*
- *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?

Each Case Home in a Distributed Architecture environment supports a maximum of 10 million reviewable items. This is the number of items after de-duplication as listed on the Processing Statistics page. The 10 million items can be within a single case or multiple cases up to the 100 case limit per Case Home.

Can an appliance have more than 10 million items?

Yes. An appliance can have more than 10 million items, however only 10 million unique items can be accessed at a time. As an example, there could be 20 cases on a Case Home. Each case has 1 million unique items. 10 of the 20 cases can be active and accessed at a time.

What is the limit for the total number of items, users, and cases on an appliance?

The maximum number of cases per Case Home is 100. The supported number of reviewers per case home or Review node depends on the hardware specification as listed below.

Appliance size	CPUs	RAM	Disk	Items	Concurrent Reviewers
Small	*16 CPUs, 2.4GHz*	64 GB	1.5 TB (1500 IOPS)	10 million	25
Medium	*32 CPUs, 2.4GHz*	128 GB	1.5 TB (1500 IOPS)	10 million	50
Large	*48 CPUs, 2.0GHz*	192 GB	1.5 TB (1500 IOPS)	10 million	100

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

The appliances in a clustered deployment communicate with each other using RMI protocols that utilize dynamic ports. The server and client negotiate which ports to use. Similar to a single-server deployment, all appliances utilize the native Windows Firewall to limit exposure to potential network security threats. The Windows Firewall is configured to allow only the minimal set of ports/applications required for eDiscovery 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:

1. Go to **System > Support Features**, and select **Firewall Browser**.
2. Enter the Firewall Command "CONFIG".

The platform 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 or upgrading the database management system (which includes MySQL), migrating and configuring the appliance in preparation for a distributed architecture deployment:

- *"Install DBMS for the Shared Remote Database" in the next section*
- *"Appliance Migration" on page 32*
- *"Other Required Cluster Configurations" on page 33*
- *"Changing Processing Settings (Required)" on page 34*

For release-specific upgrade information, refer to the appropriate Upgrade Overview and Upgrade Guide.

Important:

- Starting in 8.0, if you are upgrading a Distributed Architecture cluster when the Cluster Primary node and Remote Database are configured on the same node, you need to upgrade the database first before upgrading the product. Consult the Upgrade Overview Guide for details.
- You must stop all eDiscovery Platform services using the Clearwell Utility on all secondary nodes before upgrading the primary node. Any secondary node that has eDiscovery Platform services running while upgrading the primary node will not be available for upgrade, which might make the Distributed Architecture environment unreliable.
- Though there is a version of MySQL provided with this application, it is distributed to be used with the application and for no other purpose. Users can obtain additional MySQL licenses at Oracle's downloads page:
  http://www.mysql.com/downloads/

The shared remote database installer will automatically configure MySQL to take advantage of 128, 192, or 256 GB of RAM. Administrators can fine tune MySQL as desired: refer to the MySQL documentation for details.

Note: If your configuration has the shared remote database server and the Cluster Primary node on the same machine, you must manually stop the *EsaApplication service* before running the shared remote database installer.

Install DBMS for the Shared Remote Database

Prepare the shared remote database server

1. Ensure that the appliance and MySQL remote database server are configured with 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.

2. (Optional) If your configuration has the shared remote database server and the Cluster Primary node on the same machine, go to the Windows control panel **> Administrative Tools > Services** and stop *EsaApplicationService*.

3. Locate and run the **DBMSDistArchConfig.exe** remote database installer to install the Database from the eDiscovery Platform folder:
\V<version number>**\utilities\DBMS**
Running this installer will also install MySQL.

   Note: You must run the remote database installer even if the remote database and the primary cluster are located on the same node.

Appliance Migration

Follow the steps in this section to configure an 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:

- Moves an appliance's local databases to the shared remote database and configures the system to use the new shared remote database server.
- 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:

1. Run the **`b stop-services`** command.

2. 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, the system is still usable with the local MySQL database on the 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.)

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

   Example: **`clearwelldb.myhost.com`**

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

5. 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 (Required)*
- *Image Helper Document Converter Service Settings*
- *Changing Processing Settings (Required)*

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 ("*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 eDiscovery service accounts.)

Change Backup Directory Location (Required)

In a distributed architecture deployment, the directory location for case backups must be changed. This must be done on all appliances in the cluster.

Permissions Considerations:

- The Windows service account that runs *EsaApplicationFireDaemon* on all Veritas eDiscovery nodes must have read/write permission for the e*sa.casebackupDir* backup locations.
- All eDiscovery platform cluster servers need to be able to access the share indicated under esa.case.backupDir using the same Windows service account that runs *EsaApplicationFireDaemon*.
- The *esa.case.backupDir* property needs to be the same on all eDiscovery platform nodes.

Note: Do not create any sub-directories under *esa.case.backupDir*. If you create any directories manually, the GUI will display a message about migration being in progress. However, backups will not be visible and will remain that way.

To change the backup directory:

1. In the user interface, 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.

Image Helper Document Converter Service Settings

Introduced in version 8.0, Document Image Helper service expedites document conversion and handling of typical Microsoft Office documents in native view.

Important!

- In a node that is managed by a Cluster Primary node, you must not enable the Document Converter services on the Cluster Primary node. To disable, see *"Disable Text Extraction, Imaging, Retrieval, and Classification Sharing on Cluster Primary node" on page 39*.

To install on **utility nodes** and enable Document Image Helper, follow the instructions in the Veritas *eDiscovery Platform Utility Node Guide*. See Document Image Helper for details.

Changing Processing Settings (Required)

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

Distributed Processing settings in this section:

- *"Change to Network Location (Email Files)" on this page*
- *"Copy Files After Upgrading" on page 36*
- *"Change Threader Performance in a Distributed Setup" on page 37*
- *"Disable Text Extraction, Imaging, Retrieval, and Classification Sharing on Cluster Primary node" on page 39*

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 of 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 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 migrating from a single node setup to distributed architecture, after migration 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.
- Case backup create of a single node and being restored in distributed architecture environment.

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 the user interface, 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 and EsaPstRetrieverService*. This will enable support for processing and review of signed and encrypted messages. Starting with release 10.0, PKI encrypted messages are not supported for native review

For these two user accounts (*EsaApplicationService* and *EsaPstRetrieverService*) you can install the certificates in one of two ways, depending on your preference:

A. **Manual Installation**. See *"Option A: Manual Installation" on page 37*. (This method is required for installing on utility nodes.)
B. **Installation using the eDiscovery Platform server**. See *"Option B: Installation using eDiscovery Platform Server" on page 37*.

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 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.

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

Option B: Installation using eDiscovery Platform Server

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

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

To install certificates from the eDiscovery Platform server

1. Log in to the 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.

Change Threader Performance in a Distributed Setup

By default, in a distributed architecture setup, the distribution of 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.

Note: It is recommended 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 the user interface, 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 Text Extraction, Imaging, Retrieval, and Classification Sharing on Cluster Primary node

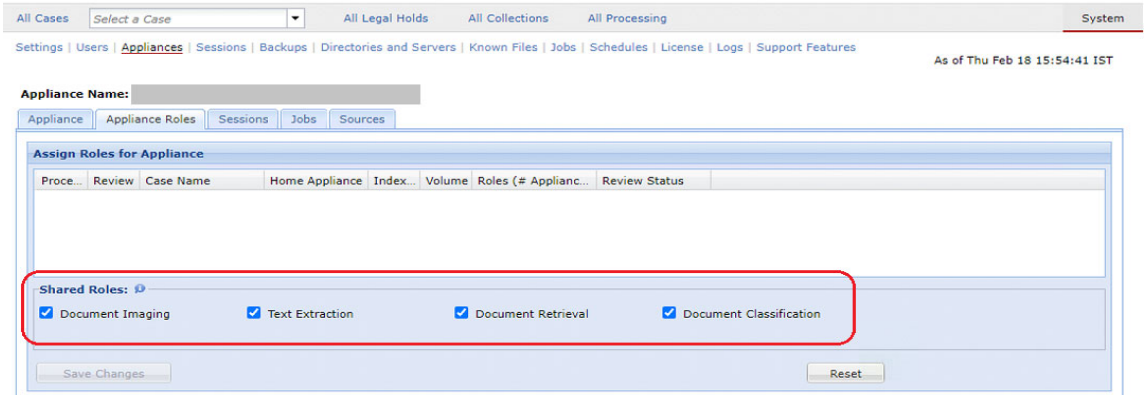
Note: This applies only if you are not using the Cluster Primary node for any cases, or as a review or processing node.

If the Cluster Primary node is on the same node as the remote database, or if it is a lower-capacity machine, you will want to prevent processes that will impact the overall application performance from being shared on that machine.

Removing a shared role will prevent document imaging, text extraction, document retrieval, and document classification processes from being shared on the Cluster Primary node.

To disable document imaging, text extraction, document retrieval, and document classification role sharing on Cluster Primary node

1. In the user interface, go to **System > Appliances > Appliance Roles.**



2. From **Shared Roles** section, clear the check box for the shared role that you want to disable for the appliance. The available options are **Document Imaging**, **Text Extraction**, **Document Retrieval**, and **Document Classification**.

3. Click **Save Changes** to apply the changes.

Appliance Assignment and Provisioning

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

Refer to the following topics in this section:

- *"Overview" in the next section*
- *"Administrator Workflow" on page 43*
  - *"Step 1: Add Appliances to the Cluster" on page 43*
  - *"Step 2: Assign Roles" on page 45*
  - *"Step 3: Provision Appliances (Review and Processing Roles)" on page 46*
- *"Distribution Enabled" on page 48*
- *"Reassigning or Unassigning Roles" on page 48*

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 Primary node, and then add all other appliances to the same cluster as Secondary nodes. See *"Step 1: Add Appliances to the Cluster" on page 43*.

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 27*.

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 45*.

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.)

Note: This setup supports users reviewing files while processing may be taking place on other nodes. Therefore files in the reviewing locations are not updated with the information from newly processed files until consolidation is completed. However, users may see duplicate information (such as location information or email headers) until the shared index is consolidated, which is when deduplication takes place. The users will be notified the next time they do a search or navigate that "The case has been updated or settings have changed. To see

the most current information, refresh the page or clear and re-run your search". For more information about consolidation, see *"How are Cases Processed using Distributed Processing?" on page 13*.

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 Primary node, and then add all other appliances to the cluster as Secondary 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 45*.)

   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 48*.

   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 and Processing Roles)" on page 46*. For processing and performance effects, see *"About Case Backups" on page 52*.

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 48*.

Step 1: Add Appliances to the Cluster

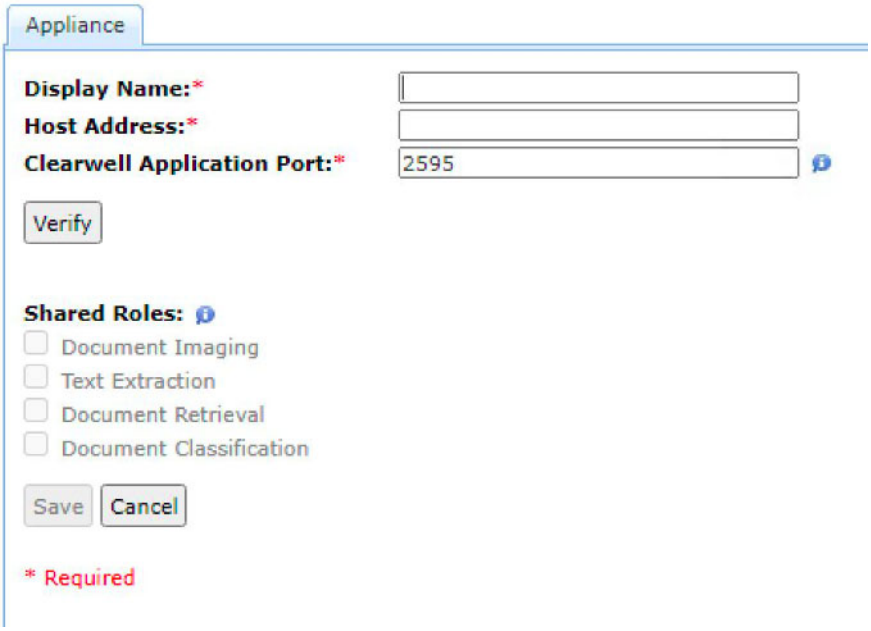
Be sure to add any new appliances (Secondary nodes) to the cluster before assigning roles. If you are adding utility nodes (for optimized review and export throughput), refer to the *Utility Node Guide* for steps on how to add utility nodes to a clustered environment.

Note: Ensure that all the appliances in the cluster are configured with the same date/time, and are in the same time zone.

To add a new node to the cluster:

1. Log on to the eDiscovery Platform user interface on the appliance you wish to designate as the Cluster Primary node, 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. From the **Shared Roles**, select the role that you want the appliance to perform. The available options are:
   - Document Imaging
   - Text Extraction
   - Document Retrieval
   - Document Classification
5. Click **Save**. Continue with *"Step 2: Assign Roles" in the next section*. See also *"Set System-Level Security and Indexing Properties" on page 27* 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.

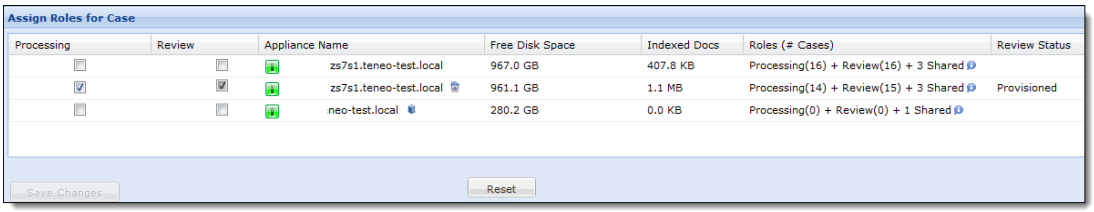
Note: Veritas recommends doing this before any cases on those appliances have begun processing so that all data is processed. See *"Processing Performance Considerations" on page 50* to understand how your assignments on nodes will affect case processing if processing has already begun. Refer to *"Appliance Capacity" on page 25* and *"About Case Backups" on page 52* to be sure your assignments optimize performance based on your appliance model, case sizes, and/or number of concurrent reviewers.

Assigning Appliance Roles for Cases with Load File Import Sources

Distributed processing for LFI sources is not supported. Assign the case home as the only processing node for cases with LFI sources. If you attempt to process a case with an LFI source without a case home provisioned, or with more than one node provisioned for processing, processing will fail. Users will need to re-provision the case home for all processing and re-run processing for the case.

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 48*.

   – Continue with managing appliances to prepare/restore nodes as necessary for case processing. See *"Managing Cases (Processing Role Only)" on page 49*.

   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 48*.

   – If you assigned an appliance (for the Review Role) after processing began, continue to *"Step 3: Provision Appliances (Review and Processing Roles)" 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 and Processing Roles)

When you provision an appliance, the system 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

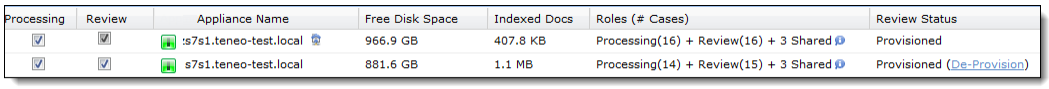
Administrators can add, remove, and replace nodes in a distributed architecture deployment when provisioning appliances using any one of the following use cases:

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 before Consolidation phase and rerun the job.

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.



Processing	Review	Appliance Name	Free Disk Space	Indexed Docs	Roles (# Cases)	Review Status
☑	☑	:s7s1.teneo-test.local	966.9 GB	407.8 KB	Processing(16) + Review(16) + 3 Shared	Provisioned
☑	☑	s7s1.teneo-test.local	881.6 GB	1.1 MB	Processing(14) + Review(15) + 3 Shared	Provisioned (De-Provision)

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 46*.) 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 49*.

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 22* 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 45*.

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 49*
- *"Managing Cases (Review Role Only)" on page 51*
- *"Check Status Before Backups" on page 52*
  - *"Prepare Nodes for Backup, Restore, or Removal" on page 52*
- *"Review Cache Process" on page 54*

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 52*) 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 63*.

If the node backup, restore, or removal fails, then the node may have cases that are fragmented. See *"Check Status Before Backups" on page 52*.

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 51*.)

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 46*.
- 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 effect 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.
  - The system 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, Veritas 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 and Processing Roles)" on page 46*.

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.

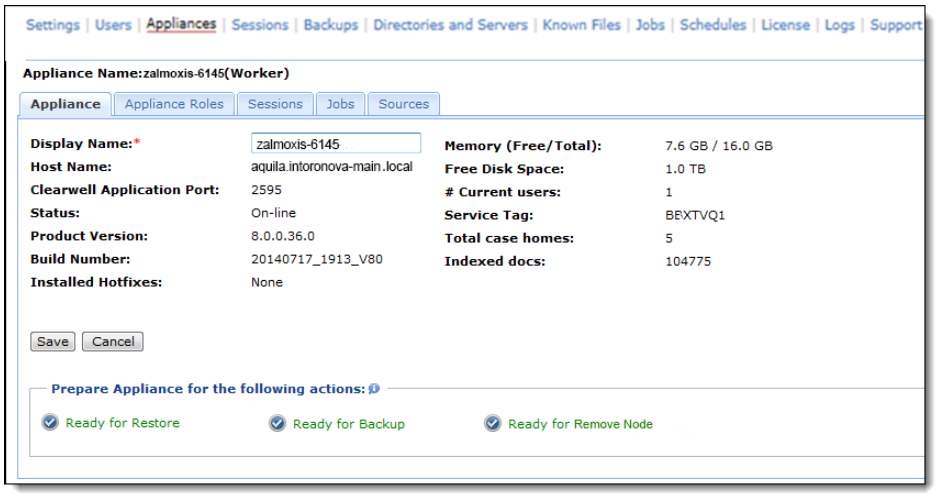
When cases on the node are ready to be backed up, (indicated with "Ready for Backup" in the Appliance), you can continue with the backing up the node. To perform an on-demand or scheduled node backup, refer to the section *"Creating Appliance Backups" in the System Administration Guide*.

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.

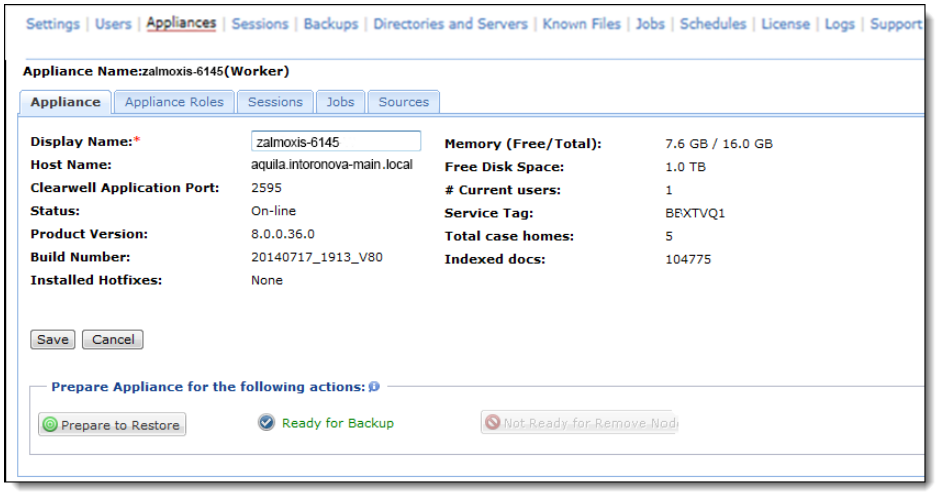
To prepare a node for backup/restore/removal

1. From the user interface, click **System > Appliances**. Click the appliance you want to check.
2. 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 52*.
   - **Ready for Restore**. The node is in a valid state and ready to be restored if needed.
   - **Ready for Removal** (Secondary nodes only). The node is in a valid state and ready to be removed if needed. If the node is the Cluster Primary node, this button will appear unavailable as shown:

Note: 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 60*.

3. 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 52*.

Restoring a Node

A 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 60*.

When you restore from a node from a backup you restore the entire appliance. You cannot restore specific cases from a node backup. Refer to the section *"Backup and 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 and Processing Roles)" on page 46*.

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 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 55*.)

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.

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 Secondary 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.

Job Execution

Type of Job	Examples
Execute on Case Home	Export, Production
Execute on Review Node	Tagging, Batching, Cache (Review Accelerator), Search, Dashboard Reports

Tips and Troubleshooting

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

Topics in this section:

- *"Changing Distributed Review Settings" on this page*
- *"Changing Distributed Review Settings" on page 58*
- *"Configuring Job Output for Exports" on page 59*
- *"Addressing Fragmented Nodes" on page 60*
- *"Addressing Node Failures, and Ensuring System Reliability" on page 63*

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:

- *"Automate Load Balancing" in the next section*
- *"Turn Off Load Balancing" on this page*

Automate Load Balancing

Automatic 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 the user interface, 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.autoredirect=auto`**

   (where **`auto`** is the default setting.)
4. Click **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 ".

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 the user interface, go to **System > Support Features**, and select **Property Browser**.
2. Enter the property according to your preference:

   A. To redirect Case Admin users:
   `esa.autoredirect=onlyadmin`

   B. To turn off completely for all reviewers:
   `esa.autoredirect=never`
3. Click **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.

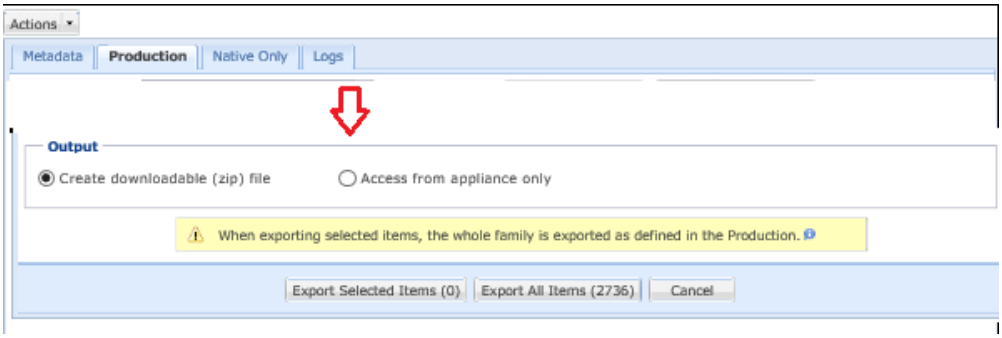
To set output location for print and export jobs

1. From the **All Cases** > **Locations** screen, click **Add**.
2. Specify the following information. An asterisk (*) indicates a required field.

Export Data Location

Field	Description
Account	Enter the name of the source account you created, or click **Browse** to select one from the list of accounts.
Location (\\server\share)*	Enter the path to the location in the UNC format (\\servername\folder) where the collected data will be stored. Click the button to select a File Share or remote directory. Click **Check Free Space** to verify data storage capacity.
Type	Select the type of the location as **Collect and Export** or **Export Only**.
Description	Enter a description for this source account (up to 255 characters).
Access Groups	By default, all groups to which the user has access are listed in the **Included** column which results in the location being added in all groups. Keep only those groups in the **Included** column in which you want to add the location and move all the remaining groups to the **Available** column. If a location is not added to any group, then that location will be available to all users.

3. Click **Save** to submit the location.
4. Use this newly added location in the Export Page to specify the export output path on the **Export** > **Production** > **Output** section > **Access from appliance only**.



Addressing Fragmented Nodes

In the eDiscovery Platform 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.

The platform provides a one-click action (done in the 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 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 63*.

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 62*
- *"Precautions for Node Restore or Removal" on page 62*

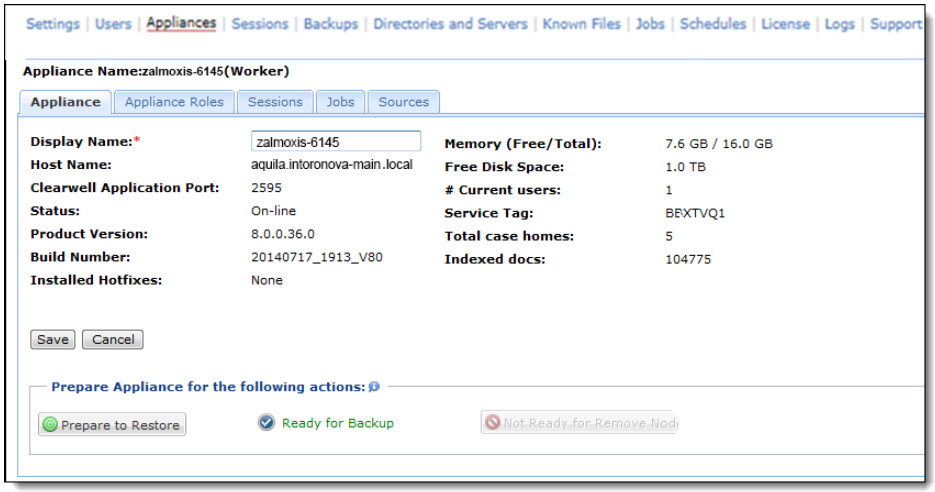
Prepare to Backup, Restore, or Remove Nodes

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 the user interface, 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 Cluster Primary appliance cannot be removed. If the node is the Cluster Primary node, the button "Not Ready for Remove Node" appears, but is unavailable (grayed out). See also *"Precautions for Node Backups with Fragmented Cases"* and *"Precautions for Node Restore or Removal" on page 62*.



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. The system attempts to defragment the case at the start of the backup job for the case.

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 61*.) 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 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 Technical Support for further assistance.

See also *"Addressing Node Failures, and Ensuring System Reliability" on page 63*. Contact 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

Veritas 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 the platform, 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 62*.)

Cluster Primary node Failure

To safeguard against Cluster Primary node failures, especially if the Cluster Primary node 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 Primary node 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, Veritas 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 Primary node, 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*.

Appendix A: Product Documentation

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

Veritas eDiscovery Platform Documentation

Document	Comments
Installation and Configuration	
Installation Guide	Describes prerequisites, and how to perform a full install of the Veritas eDiscovery Platform application
Upgrade Overview Guide	Provides critical upgrade information, by version, useful prior to upgrading an appliance to the current product release
Upgrade Guide	Describes prerequisites and upgrade information for the current customers with a previous version of the software application
Utility Node Guide	For customers using utility nodes, describes how to install and configure appliances as utility nodes for use with an existing software setup
Distributed Architecture Deployment Guide	Provides installation and configuration information for the Review and Processing Scalability feature in a distributed architecture deployment
Getting Started	
Navigation Reference Card	Provides a mapping of review changes from 10.x compared to 9.x, 8.x compared to 7.x and 7.x compared to 6.x
Administrator's QuickStart Guide	Describes basic appliance and case configuration
Reviewer's QuickStart Guide	A reviewer's reference to using the Analysis & Review module
Tagging Reference Card	Describes how tag sets and filter type impact filter counts
User and Administration	
Legal Hold User Guide	Describes how to set up and configure appliance for Legal Holds, and use the Legal Hold module as an administrator
Identification and Collection Guide	Describes how to prepare and collect data for processing, using the Identification and Collection module
Case Administration Guide	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
System Administration Guide	Includes system backup, restore, and support features, configuration, and anti-virus scanning guidelines for use with Veritas eDiscovery Platform
Load File Import Guide	Describes how to import load file sources into Veritas eDiscovery Platform
User Guide	Describes how to perform searches, analysis, and review, including detailed information and syntax examples for performing advanced searches

Veritas eDiscovery Platform Documentation

Document	Comments
Imaging Tool Upgrade Guide	Provides details about the Imaging Tool Upgrade feature and how to perform Imaging Tool Upgrade after the eDiscovery Platform appliance is upgraded to version 10.1, workflows affected when the cases are upgraded or not upgraded, and frequently asked questions (FAQs).
Export and Production Guide	Describes how to use and produce exports, productions, and logs (privilege and redaction logs)
Transparent Predictive Coding User Guide	Describes how to use the Transparent Predictive Coding feature to train the system to predict results from control data and tag settings
Audio Search Guide	Describes how to use the Audio Search feature to process, analyze, search and export search media content
Reference and Support	
Audio Processing	A quick reference card for processing multimedia sources
Audio Search	A quick reference card for performing multimedia search tasks
Legal Hold	A quick reference card of how to create and manage holds and notifications
Collection	A quick reference card of how to collect data
OnSite Collection	A quick reference for performing OnSite collection tasks
Review and Redaction	Reviewer's reference card of all redaction functions
Keyboard Shortcuts	A quick reference card listing all supported shortcuts
Production	Administrator's reference card for production exports
User Rights Management	A quick reference card for managing user accounts
Online Help	
Includes all the above documentation (excluding Installation and Configuration) to enable search across all topics. To access this information from within the user interface, click **Help**.	
Release	
Release Notes	Provides latest updated information specific to the current product release