



System Administration Guide

Version 5.5

CONFIDENTIAL INFORMATION

The information herein is the property of Ex Libris Ltd. or its affiliates and any misuse or abuse will result in economic loss. DO NOT COPY UNLESS YOU HAVE BEEN GIVEN SPECIFIC WRITTEN AUTHORIZATION FROM EX LIBRIS LTD.

This document is provided for limited and restricted purposes in accordance with a binding contract with Ex Libris Ltd. or an affiliate. The information herein includes trade secrets and is confidential.

DISCLAIMER

The information in this document will be subject to periodic change and updating. Please confirm that you have the most current documentation. There are no warranties of any kind, express or implied, provided in this documentation, other than those expressly agreed upon in the applicable Ex Libris contract. This information is provided AS IS. Unless otherwise agreed, Ex Libris shall not be liable for any damages for use of this document, including, without limitation, consequential, punitive, indirect or direct damages.

Any references in this document to third-party material (including third-party Web sites) are provided for convenience only and do not in any manner serve as an endorsement of that third-party material or those Web sites. The third-party materials are not part of the materials for this Ex Libris product and Ex Libris has no liability for such materials.

TRADEMARKS

"Ex Libris," the Ex Libris bridge, Primo, Aleph, Alephino, Voyager, SFX, MetaLib, Verde, DigiTool, Preservation, Rosetta, URM, ENCompass, Endeavor eZConnect, WebVoyage, Citation Server, LinkFinder and LinkFinder Plus, and other marks are trademarks or registered trademarks of Ex Libris Ltd. or its affiliates.

The absence of a name or logo in this list does not constitute a waiver of any and all intellectual property rights that Ex Libris Ltd. or its affiliates have established in any of its products, features, or service names or logos.

Trademarks of various third-party products, which may include the following, are referenced in this documentation. Ex Libris does not claim any rights in these trademarks. Use of these marks does not imply endorsement by Ex Libris of these third-party products, or endorsement by these third parties of Ex Libris products.

Oracle is a registered trademark of Oracle Corporation.

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

Microsoft, the Microsoft logo, MS, MS-DOS, Microsoft PowerPoint, Visual Basic, Visual C++, Win32, Microsoft Windows, the Windows logo, Microsoft Notepad, Microsoft Windows Explorer, Microsoft Internet Explorer, and Windows NT are registered trademarks and ActiveX is a trademark of the Microsoft Corporation in the United States and/or other countries.

Unicode and the Unicode logo are registered trademarks of Unicode, Inc.

Google is a registered trademark of Google Inc.

iPhone is a registered trademark of Apple Inc.

Copyright Ex Libris Limited, 2018. All rights reserved.

Document released: April 2018

Web address: <http://www.exlibrisgroup.com>

Table of Contents

	Updates to This Guide	7
Part I	Managing the Rosetta System	
Chapter 1	Understanding the Rosetta System Architecture	11
	About the Rosetta System Architecture	11
	Rosetta Application Roles	13
	<i>Deposit Role</i>	13
	<i>Repository Application Role</i>	15
	<i>Permanent Application Role Server</i>	18
	<i>Delivery Application Role</i>	19
	<i>Index Application Role</i>	20
	Administering the Solr Server	20
	<i>Index Rank</i>	20
	<i>Index Size and Location</i>	20
	<i>Index Status</i>	21
	<i>Index Shards</i>	22
	Rosetta-Supported Topologies	23
	<i>All-in-One Topology</i>	23
	<i>2-Tier Topology Diagram</i>	24
	Load Balancer Requirements	25
	<i>Purpose</i>	25
	<i>Sticky Session</i>	25
	<i>Forwarding</i>	25
	<i>Load Balancing Web Requests</i>	25
	<i>Worker Pools for Back-End Activities</i>	26
	Managing Application Roles and Other Server Settings	26
	<i>global.properties</i>	26
	<i>2-Tier Topologies</i>	26
	<i>Rosetta Scalability</i>	27
	<i>Managing Workers</i>	27
	<i>SIP Processing Throttling</i>	30
	Managing Plug-Ins	31

	Disk Space Management	34
	<i>Checking Disk Space</i>	34
	<i>Freeing Disk Space</i>	35
	Rosetta Storage Management	35
	<i>Understanding Storage Architecture</i>	35
	<i>Configuring Storage Groups</i>	36
	<i>Storage Space Checker</i>	47
	<i>Configuring Storage Rules</i>	47
	<i>Rule Details Page</i>	48
	<i>Adding a Storage Rule</i>	50
	<i>Configuring External Metadata: SRU/SRW</i>	53
Chapter 2	General Parameters	59
	Backoffice	59
	BIRT	60
	Delivery	64
	Deposit	65
	Format Library	66
	General	67
	Menu	69
	Network	70
	Authentication	70
	Repository	71
	SMTP Email	72
	Search	73
Chapter 3	Updating the Rosetta System	75
	Verifying your Rosetta Server Version.....	75
	Service Pack Installation.....	76
	Hot Fix Installation.....	76
	<i>Installing a Rosetta Hot Fix</i>	76
	<i>Rolling Back the Hot Fix Installation</i>	77
	<i>Restarting Rosetta</i>	77
	Copying Rosetta System Configurations	78
	<i>Items Copied in Copy Configuration</i>	78
	<i>Implementing the Copy Configuration</i>	79

Chapter 4	Working with the Rosetta System	83
	Starting and Stopping the Rosetta Application Servers.....	83
	Verifying the Operation of the Rosetta System.....	85
	Logging On to the Rosetta System	85
	System Checks.....	86
	Description of Aliases	87
	PDS Directory Tree.....	88
	Rosetta System Log Files	88
	System, Background, and Operational Jobs	89
	<i>Monitoring All Jobs</i>	92
	<i>Scheduling a System Job</i>	93
	<i>Canceling a System Job</i>	95
	Shared Directories	96
	<i>Operational and Permanent Repository Definitions</i>	96
Part IV	Managing the Rosetta Database	
Chapter 6	Working with the Rosetta Database.....	101
	Connecting to the Database from UNIX	102
	Changing the Database Schema's Password	102
	Starting and Stopping the Oracle Server.....	102
	Starting and Stopping the Oracle Listener.....	103
	Viewing and Managing Tablespaces	106
	Viewing and Managing Oracle Data File (DBF) Information	108
	Archiving Data.....	110
	Viewing and Managing Oracle Database Memory (SGA)	112
	Viewing Alert Log (Oracle Database Log) File Content	114
	Viewing and Managing Information on Oracle Database Users.....	116
	Creating Database Users with Privileges	118
Part V	Data Backup and Recovery	
Chapter 8	Backing Up and Recovering Data.....	123
	About Backing Up Data.....	123

Rosetta Scripts	124
<i>Deposit Application Role</i>	124
<i>Repository Application Role</i>	125
<i>Permanent Repository</i>	125
Backup Strategies.....	125
<i>About Backup Methods</i>	126
<i>Implementing Backup Strategies</i>	126
<i>Backup Scenarios</i>	127
Recovering Data.....	127

Updates to This Guide

This guide is being reissued due to the following changes:

- **Table 22 Rosetta Jobs** on page **90** was updated.
- The new `anonymize_mets` general parameter was added to **Delivery** on page **64**.
- A note was added to the Find Duplicates job in **System, Background, and Operational Jobs** on page **89** that the maximum number of duplicates to be reported can be configured.
- A note was added to the Clean Up job in **System, Background, and Operational Jobs** on page **89** that Clean Old Events can be configured to include/exclude specific event types.
- A note was added to the Clean Up job in **System, Background, and Operational Jobs** on page **89** that when selecting **Clean Deleted OAI-PMH records**, it is recommended that you set the OAI `deletedRecord` policy to `transient` in the `oaiproviderconfig.xml` configuration file.

Part I

Managing the Rosetta System

This part contains the following:

- **Chapter 1: Understanding the Rosetta System Architecture** on page 11
- **Chapter 2: General Parameters** on page 59
- **Chapter 3: Updating the Rosetta System** on page 75
- **Chapter 4: Working with the Rosetta System** on page 83

1

Understanding the Rosetta System Architecture

This section contains:

- [About the Rosetta System Architecture](#) on page 11
- [Rosetta Application Roles](#) on page 13
- [Administering the Solr Server](#) on page 20
- [Rosetta-Supported Topologies](#) on page 23
- [Load Balancer Requirements](#) on page 25
- [Managing Application Roles and Other Server Settings](#) on page 26
- [Managing Plug-Ins](#) on page 31
- [Disk Space Management](#) on page 34
- [Rosetta Storage Management](#) on page 35

About the Rosetta System Architecture

The Rosetta system architecture is based on the multi-layer concept. Components of each layer can interact with components of other layers, as well as with components of the same layer.

Table 1 describes the layers and their components:

Table 1. The Rosetta System Layers

Layer	Components	Description
Storage	<ul style="list-style-type: none">■ Deposit Storage■ Operational Storage■ Permanent Storage	Provides physical storage of Producer Agent content for all files that are processed and preserved.

Table 1. The Rosetta System Layers

Layer	Components	Description
Application roles	<ul style="list-style-type: none"> ■ Deposit ■ Repository ■ Delivery ■ Index (SOLR) ■ Permanent 	Executes all Rosetta processes and activities - SIP processing, Maintenance tasks, Delivery, Preservation Actions.
Database	Schemas: <ul style="list-style-type: none"> ■ DEP ■ SHR ■ REP ■ RPT ■ PER 	Stores data, configuration items, and is used for operating Rosetta.

Figure 1 illustrates the components that each layer contains, and provides a general overview of the interaction between these layers and their components:

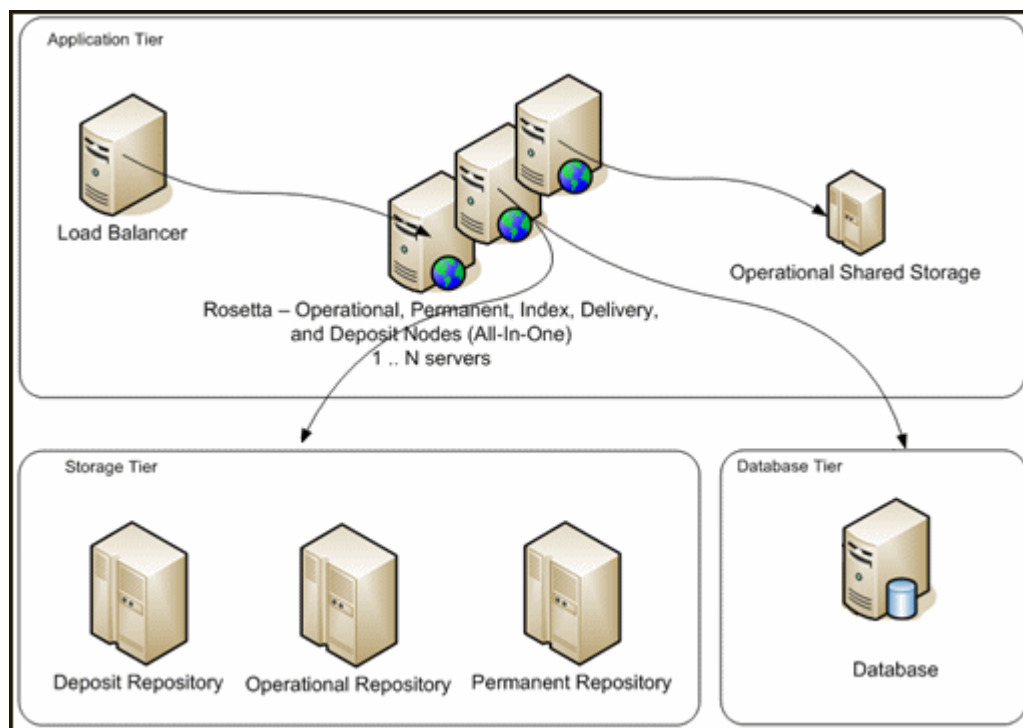


Figure 1: Rosetta System Architecture

The different application roles communicate each other by calling web services (WS) or by using queues managed in the shared database scheme.

The following sections contain detailed descriptions of the processes and components in each stage.

Rosetta Application Roles

Each server used by Rosetta can be assigned to one or more application roles. Each application role is in charge of different activities. For example, all Deposit activities—acquiring the deposited files, converting the input data to METS structure—are performed by servers whose `APP_ROLE = DEP` (Application Role is set to Deposit). If more than one server is assigned to the same application role, a load balancer is required to balance the load between the servers.

This section describes the following Rosetta application roles:

- **Deposit Role** on page 13
- **Repository Application Role** on page 15
- **Permanent Application Role Server** on page 18

Deposit Role

The Deposit role acquires the deposited material and stores this content in the Deposit Storage.

The Deposit role interacts with the following components, as illustrated in **Figure 2**:

- **Patron Directory Service (PDS)** on page 14
- **Database Schemas** on page 14
- **Deposit Storage** on page 15

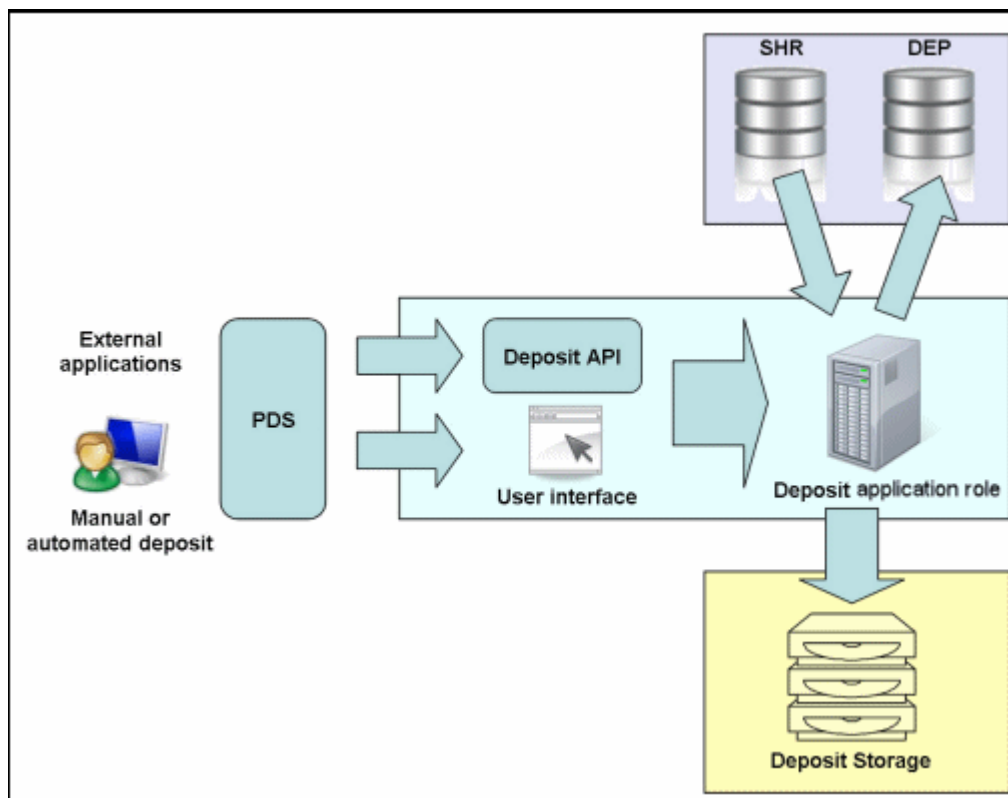


Figure 2: Deposit Application Role Architecture

Patron Directory Service (PDS)

Users can register using either an external legacy application or the Rosetta system. The PDS enables the authentication and login of users regardless of the registration method. For example, the PDS can be configured to work with an external user database, such as an LDAP directory service.

The PDS is configured by a System Administrator.

Database Schemas

The Deposit application role uses the following database schemas:

- **SHR**, which contains configuration information required for processing Producer Agent content (such as material flow and Producer profile configuration).
- **DEP**, which stores information about deposit activities that a Producer Agent submitted

Deposit Storage

The Deposit storage area contains

- deposit activities, which are deposited by Producer Agents, and
- submission information packages (SIPs), which are generated by the Rosetta system.

For each deposit activity, the Rosetta system creates a `Deposit` folder. To differentiate between `Deposit` directories that store different deposit activities, the Rosetta system adds an automatically generated ID to the name of each `Deposit` directory.

Figure 3 illustrates the organization of a `Deposit` directory.

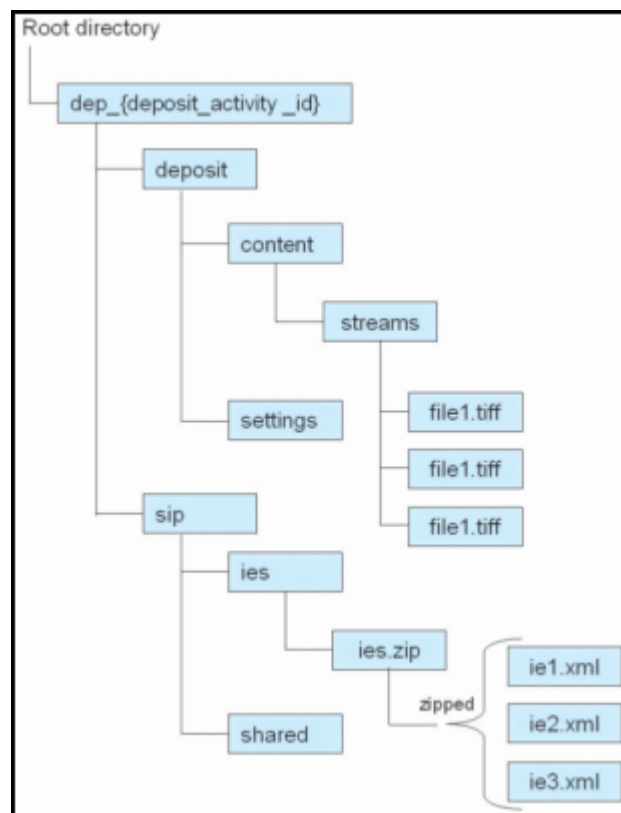


Figure 3: Deposit Directory Organization

Repository Application Role

The repository application role (sometimes known as the staging area) manages all activities associated with SIPs (validation stack, enrichment, 3A/TA, Web editor) and IEs (processes, publishing, Web editor).

The repository application role interacts with the following components, as illustrated in **Figure 4**:

- **Database Schemas** on page 17
- **Deposit Storage** on page 17
- **Operational Storage** on page 17
- **Permanent Storage** on page 17

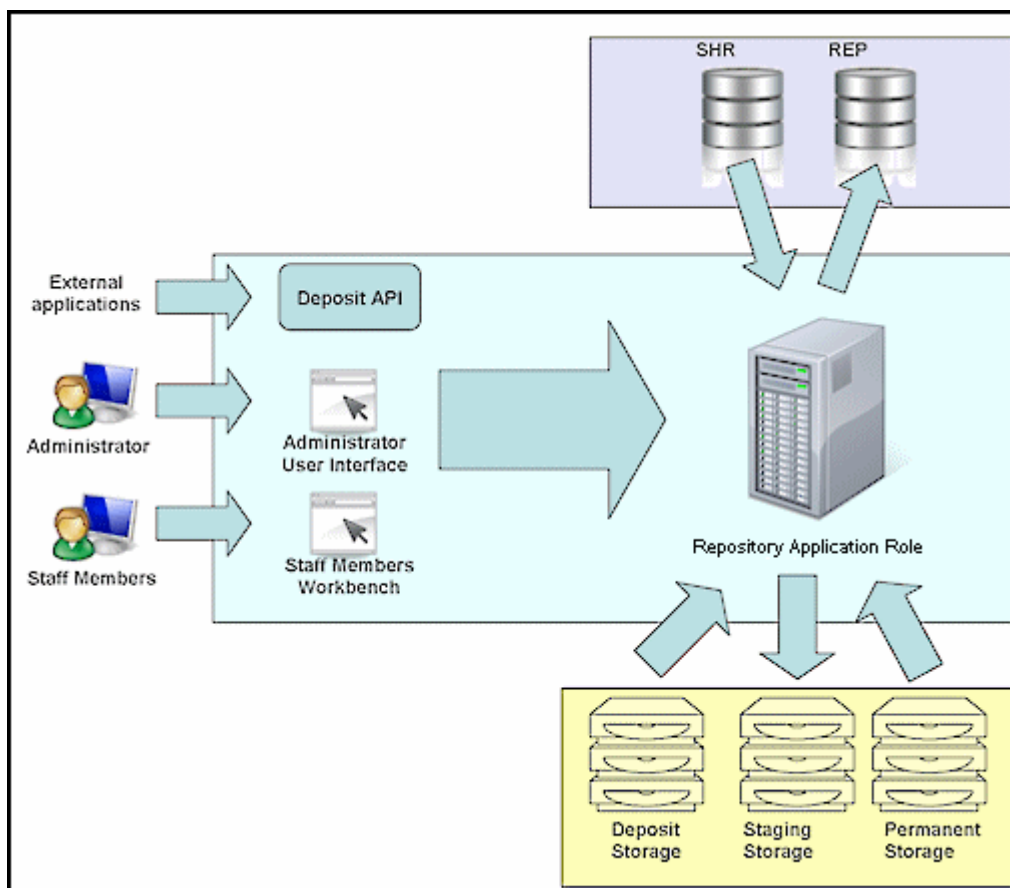


Figure 4: Staging Repository Application Role Architecture

Database Schemas

The repository application role uses the following database schemas:

- **SHR**, which contains configuration information required for processing SIPs and IEs (such as material flow and Producer profile configuration).
- **REP**, which stores information about content stored in the Operational Storage, including intellectual entities, metadata, SIP processing configuration, storage rules, and delivery rules.

Deposit Storage

The repository application role interacts with the Deposit Storage in order to move the stream files and METS files to the Operational Storage for further processing.

For more information about the Deposit Storage, see [Deposit Storage](#) on page 15.

Operational Storage

The Rosetta system stores the stream files, as defined by storage rules. (For more information on storage rules, see the **Configuring Storage Rules** section in **Part III, Configuration Components**, of the *Rosetta Configuration Guide*.)

Information about the physical location of the stream files, along with their identifiers, is stored in the Oracle database. These identifiers are also stored in METS files in order to enable the Rosetta system to establish a relationship between a METS file — which contains metadata about the stream files — and the files themselves.

METS files are grouped into submission information packages (SIPs) and stored as described in [Deposit Storage](#) on page 15.

Permanent Storage

The repository application role interacts with the Permanent Storage when a content object (such as an intellectual entity, representation, or file) that has already been moved to the Permanent Storage must be edited.

Because no changes can be made in the Permanent Storage, the Rosetta system retrieves the content object from the Permanent Storage and stores this object in the Operational Storage.

After the object is edited, the Rosetta system returns this object to the Permanent Storage.

For more information about the Permanent Storage, see [Permanent Storage](#) on page 19.

Permanent Application Role Server

The permanent application role server provides permanent storage for content objects (such as IEs, representations, and files) that have been processed and approved.

The permanent application role server interacts with the following components, as illustrated in **Figure 5**:

- **Database Schema** on page 19
- **Operational Storage** on page 17
- **Permanent Storage** on page 19

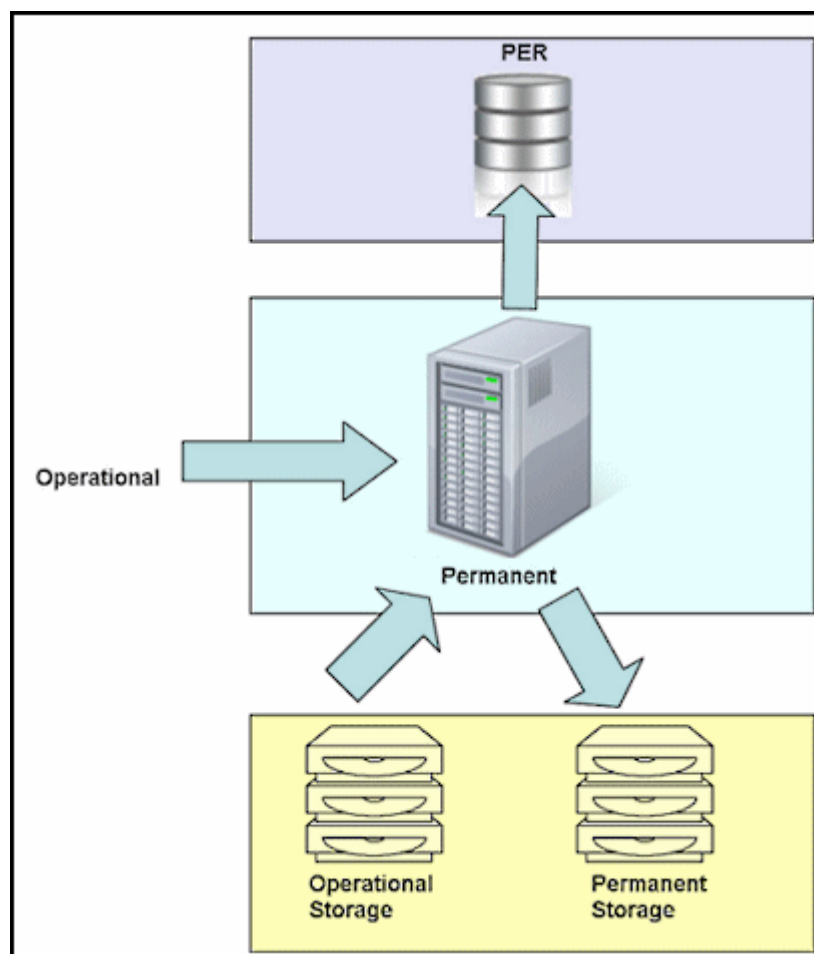


Figure 5: Permanent Application Role Architecture

Database Schema

The permanent application role server uses the PER database schema to store information about content objects (such as intellectual entities, representations, and files) that are permanently stored in the Rosetta system.

Operational Storage

The permanent application role server interacts with the Operational Storage to retrieve content objects — that have been approved by Staff Members — from the Operational Storage, and move these objects to the Permanent Storage.

For more information about the Operational Storage, see [Operational Storage](#) on page 17.

Permanent Storage

The Permanent Storage is designed to store intellectual entities (IEs) that were approved by Staff Members for permanent preservation. As a result, content objects that are stored in the Permanent Storage cannot be updated, deleted, or rearranged.

Whenever a content object must be changed (for example, its metadata requires editing), the Rosetta system moves it back to the Operational Storage. When the editing process is complete, the system creates a new METS XML as the new version of the IE and writes it to the Permanent Repository.

The Rosetta system stores files in the Permanent Storage, as defined by storage rules. (For more information on storage rules, see [Configuring Storage Rules](#) on page 47.)

Information about the physical location of the files, along with their permanent identifiers, is stored in the Oracle database. These permanent identifiers are also stored in METS files in order to enable the Rosetta system to establish a relationship between a METS file — which contains metadata about the permanent files — and the files themselves.

To store METS files, the Rosetta system creates a folder for each IE. The system uses the IE's permanent identifier as the name of the folder.

Delivery Application Role

The delivery application role is in charge of managing the Delivery requests that come from outside of Rosetta, for example from Primo patrons or the Aleph Web-OPAC.

The delivery application role can reside on the same server as other application roles (such as repository) but it can also reside on a dedicated server that can be accessed by external users.

Delivery interacts with both storages, operational and permanent, in order to deliver the requested metadata and files. It also interacts with the SHR scheme where the delivery rules are managed and the REP scheme where information about the files is stored.

Index Application Role

The index application role manages the SOLR indexing that allows the search of objects in Rosetta. The index application role can reside on the same server as other application roles (such as repository) but it can also reside on dedicated server(s) to allow better search performance. All index (IDX) servers must be accessible to all backoffice (REP) servers.

For more details about the SOLR, see the SOLR section of this guide.

Administering the Solr Server

The Solr indexing mechanism uses a properties configuration table that is managed by Ex Libris and updated as needed with every release or service pack. All METS and DNX elements are indexed and searchable.

The Solr server also indexes all attributes, each of which has a corresponding row in the `UI_Labels` code table. Customers can control the label of each indexed attribute through this table.

Index Rank

A new attribute added to the `APPLICATION_VERSION` table controls whether an index server (a server that has the `IDX` role) is active. Setting that attribute to one (1) means the index server is active and will be used by Solr for new/modified indexed information. Setting it to zero (0) means the index server will not be used by Solr for any new index rows. Modified index rows are still maintained in the same index server on which they were initially created.

Index Size and Location

The index size parameter, `solr_server_capacity`, defines the index size for all index servers. The value is defined in KB. Generally, for every 1 MB of files indexed, the system requires 2 GB of space.

The index size is informational only and not enforced by Rosetta. Index size is displayed in the Index Status UI and is also used to calculate the percentage of currently used index size.

The location of the index files is stored in the global properties file (for each server that is defined as `IDX` in its application role). The property name is

`dps.storage.idx`. The default value is `/exlibris/dps/d4_1/profile/solr/data`.

Index Status

A page displaying the indexing progress is available from **Home > Advanced Configuration > Repository > Index Status**. The Permanent Index tab displays the indexing progress for IEs and the operational Index tab displays the indexing progress for SIPs.

Repository / Index Status

Permanent Index Operational Index

Index Instances

Server Name	Health	Index Size (Files)	% Full	Weight	Awaiting Index (IEs)
il-dps04.corp.exlibrisgroup.com	UP	286	<input type="text"/>	1	0
il-dps08.corp.exlibrisgroup.com	UP	301	<input type="text"/>	1	0

Index Errors

IE records in Exception Queue: 0 [Re-index](#)

Index Operations

[Reindex Full Repository](#)

[Back](#) [Refresh](#)

Figure 6: Permanent Index Status Page from the Administration UI

Server Name	Health	Index Size (Files)	% Full	Weight	Awaiting Index (SIPs)
il-dps04.corp.exlibrisgroup.com	UP	352	<input type="text"/>	1	0
il-dps08.corp.exlibrisgroup.com	UP	358	<input type="text"/>	1	0

Figure 7: Operational Index Status Page from the Administration UI

Details about Index Instances appear in the first section of the Index Status page. In the second section, Index Errors, the number of IEs/SIPs that failed during indexing, is shown. An option to resubmit the failed IEs/SIPs and re-activate the index process is also available in this section (**Re-index** button).

If there are records in the Exception Queue, clicking the number opens a list of the PIDs in the IEs/SIPs. These IEs/SIPs are not searchable by their metadata, but they can be reached by the quick search (using their PID).

You can run a full repository re-index from the third section of the page, Index Operations. A full re-index should be run in the following cases:

- New index fields or functionality requiring a global re-index
- Index corruption or hardware failure for an index node
- Addition or removal of an index node

Index Shards

When the SOLR index is managed on multiple servers, and one of the servers is down, the search results may not be returned from this server.

When this happens, objects that are indexed on that server are still available for delivery through a deep link URL (for example, from Primo).

In addition, searching IEs in Rosetta by PID works for all IEs, even the ones that are indexed on the server that is down.

Rosetta-Supported Topologies

The following topologies are supported in Rosetta:

- **All-in-One:** See [All-in-One Topology](#) on page 23.
- **2-Tier:** See [2-Tier Topology Diagram](#) on page 24 and page 26

All-in-One Topology

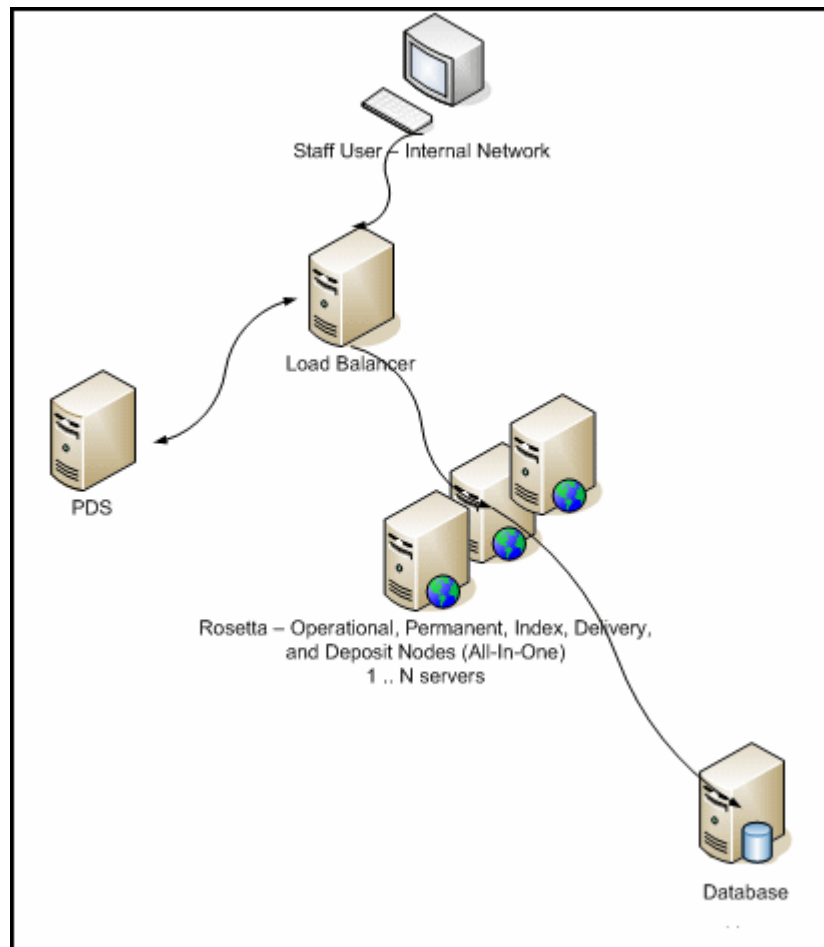


Figure 8: All-in-One Topology

Deposit, Delivery, Staging, and Permanent all reside on the same server. This configuration can be scaled wide to support redundancy and through-put.

2-Tier Topology Diagram

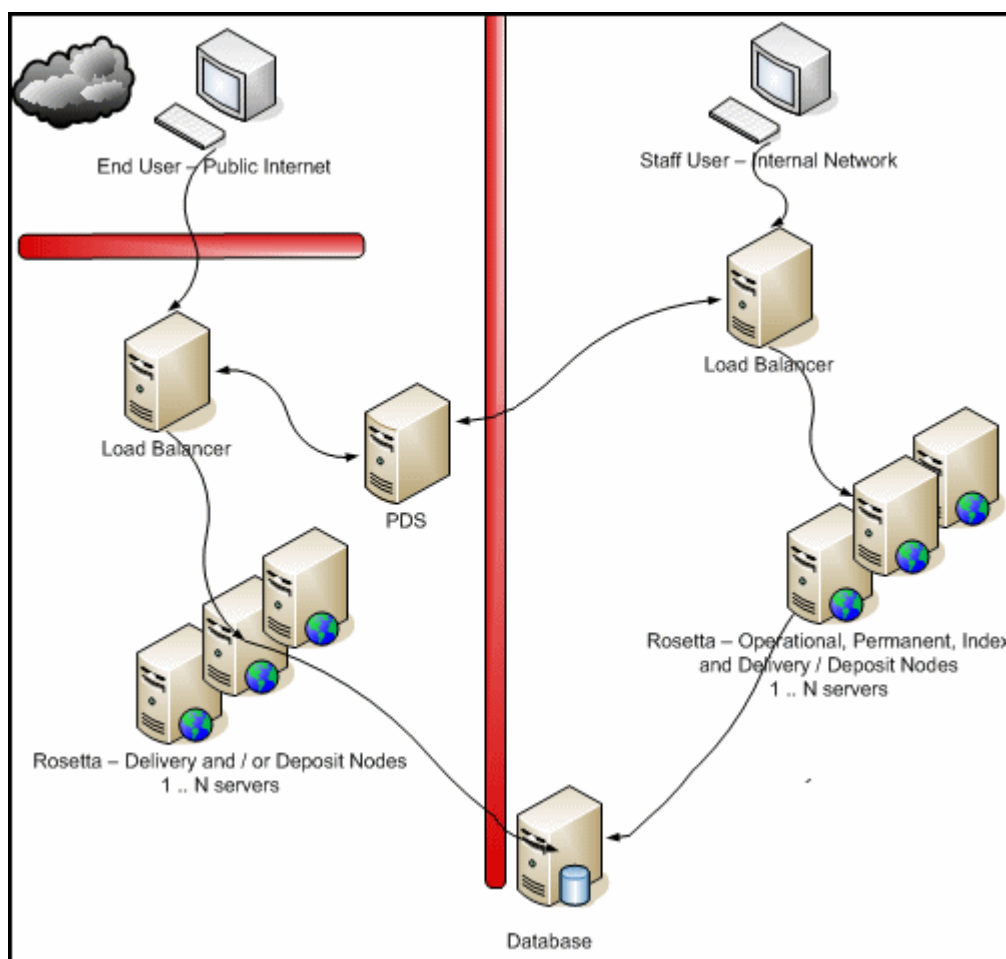


Figure 9: 2-Tier Topology

Deposit and/or Delivery are on a separate tier in a DMZ, and Staging and Permanent reside on an internal tier. This configuration can also be scaled wide, and is only used in environments where deposit or delivery functionality must be available to users outside of the organization.

PDS can be run either on Rosetta application servers with a REP or DEL application role or on an external server, either as a single instance or highly-available with a DB backend (with PDS 2.x). For detailed information and instructions on configuring PDS, see the *Patron Directory Services Guide*, available under Cross-Product Information in the Documentation Center.

Load Balancer Requirements

Purpose

Load-balancing functionality can be achieved by either hardware or software components. The purpose of this section is to describe the functional aspects of the load balancer as used by Rosetta regardless of whether it is a hardware or software solution.

Sticky Session

Session stickiness must be provided by the load balancer.

A load balancer redirects end users, identified by source IPs, to the same physical Deposit server, during a single user session.

Forwarding

Configure the load balancer to pass the client's IP address in the X-Forwarded-For HTTP header value.

Load Balancing Web Requests

In any of the supported topologies, a load balancer should be defined for each Fully Qualified Domain Name (FQDN). A separate FQDN is required for each module/role that must be accessed separately. The following modules can be accessed separately:

- PDS
- Deposit
- Back Office
- Delivery

Configure the load balancer to redirect all requests to the specified domain names to be forwarded to the real server which made up the cluster (for example, port 1801). For example, requests to `http://delivery.rosetta.myorg.org/` should redirect to `http://delivery1:1801` and <http://delivery2:1801>.

NOTE:

If your topology requires that the Web service response includes the load balancer hostname (and not a specific application server hostname), make sure the load balancer forwards the request's original host header (for example, the `mod_proxy ProxyPreserveHost` directive is turned on).

All UI and Web service requests are directed to the appropriate load balancer FQDN as defined in `global.properties`. For further details, see [Managing Application Roles and Other Server Settings](#) on page 26.

Worker Pools for Back-End Activities

Workers are managed by the database direct communication with the application servers. No load balancer is necessary. Users can manage the number of workers (threads) that run on each server in order to optimize the performance of the system. For further details, see [Managing Workers](#) on page 27.

Managing Application Roles and Other Server Settings

The settings of each server, including the assignment of an application role to a server, are defined in the configuration file called `global.properties`, which is located under `$dps_dev/system.dir/conf/`.

`global.properties`

This file holds all the settings that might vary between servers, while all the settings that are relevant for the entire installation are defined in the database's general parameters table (see below).

Changes in the `global.properties` file take effect only after running the script `$dps_dev/system.dir/bin/set_globals.sh`.

IMPORTANT:

Rosetta must be shut down before running `set_globals.sh`.

2-Tier Topologies

In a two-tier topology, before changing the the `APP_ROLE` of a certain server, the System Administrator must update `[module].server` in `global.properties` with the correct FQDN for each module (before running the `set_globals` script). If a load balancer is applied, update `load balancer.[module].host` and `load balancer.[module].port` for each relevant module.

NOTE:

If the changes affect more than just the `global.properties`, the System Administrator must run the `$dps_dev/system.dir/bin/rosettings.csh` script.

Cases that require the Administrator to run the `rosettings.csh` script include:

- Changing FQDN/LB name / port for deposit, operational, delivery application roles
- Changing the PDS URL / port
- Changing the database credentials
- Turning the PDS SSL on or off, or changing the PDS SSL port
- Setting the Rosetta SSL via load balancer

In any of these cases, we strongly recommend that you consult first with the Rosetta Support team.

Rosetta Scalability

To increase the computing power of Rosetta, you can use more than one server with the same application role. In this case, all the servers that work with the same application role should have a load balancer that balances the load between them.

When adding or modifying the load balancer settings, the URL must be changed in the `global.properties` file of each server. To apply the changes, the System Administrator must run the `$dps_dev/system.dir/bin/rosettings.csh` script.

NOTE:

Changing the application role of a server or adding a load balancer should be undertaken only after consulting the Rosetta Support team.

Managing Workers

Rosetta uses worker threads to handle back-end processing. Such back-end processes include:

- Deposit work
- Event processing
- Indexing
- SIP processing
- Permanent processing
- SIP loading

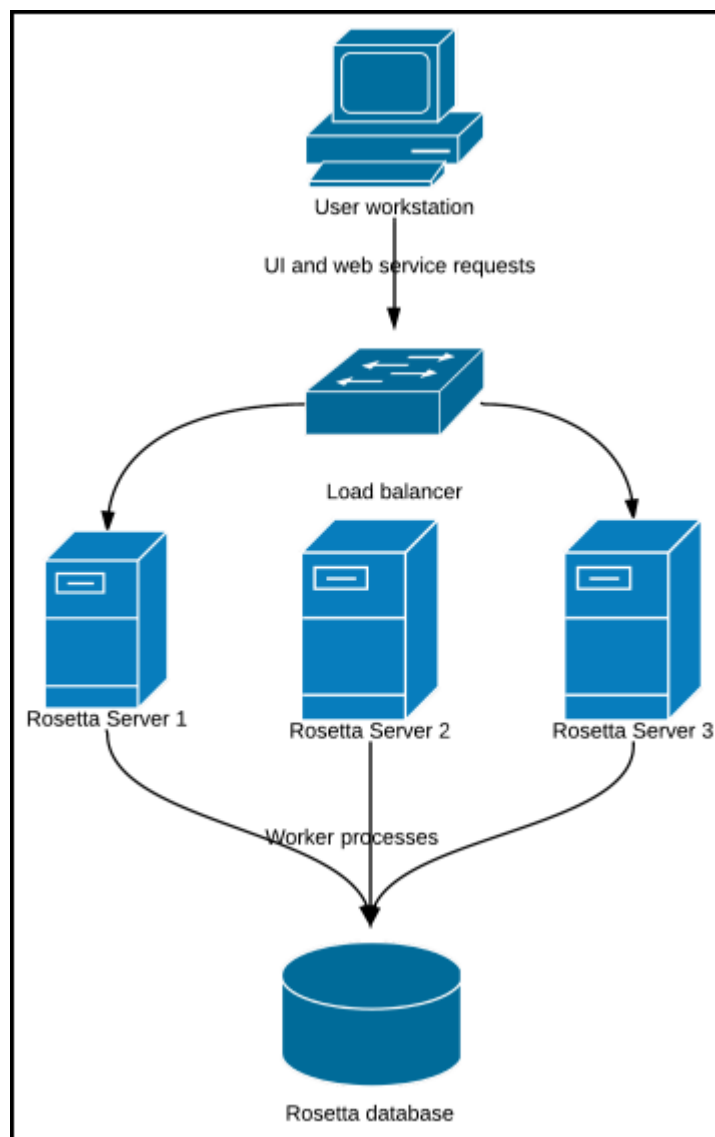


Figure 10: Rosetta Workers

SIP processing resources can be configured on each server in order to optimize the performance of the system. These settings are configured out-of-the-box for a server of typical computing power in line with the Rosetta system requirements. Installations with larger or smaller servers may need to adjust these settings to achieve an optimum utilization of system resources. In addition, worker threads can be configured to create a server that handles only specific functionality within an application role. A typical use case could be the need to allocate enough computing power for front-end activity (such as Web deposit and Delivery) while back-end activities such as maintenance tasks run in the background.

To adjust these settings, log on to the Rosetta Administration site and follow the path from the Advanced Configuration page: **General > SIP Processing Workers**.

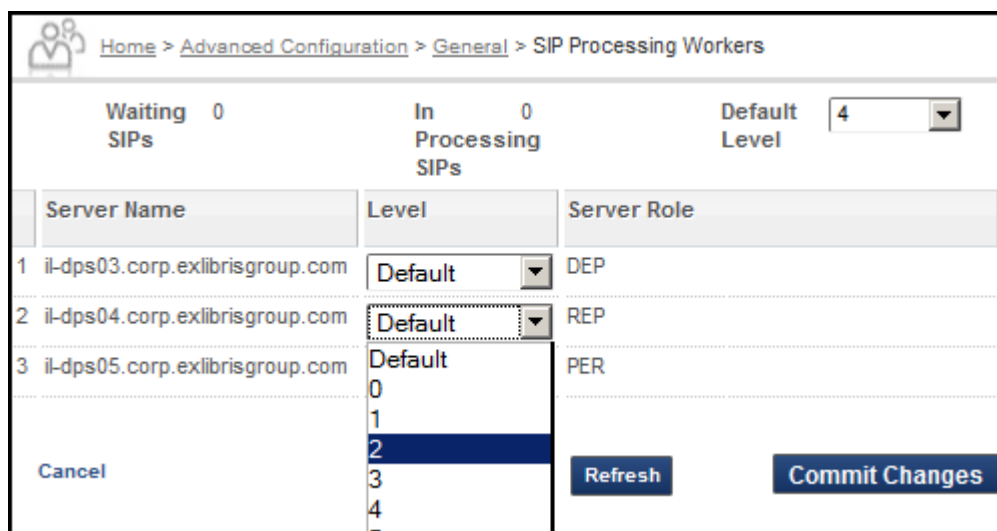


Figure 11: SIP Load Balancing

The Level parameter is a 0-10 value that determines the SIP processing load for the entire environment by setting all servers to **default** and adjusting the default value, or by individually adjusting each SIP processing server (DEP, REP, and PER).

To ensure UI high-availability, lower the level on one or more of your REP servers, or consider reserving one of your REP servers for UI functionality by setting its level to 0.

An indication of a balanced and fully-utilized SIP processing configuration is one in which the number of waiting SIPs is constantly approaching 0 and the number of SIPs in process is steady (the actual number naturally depends on the number of servers). Adjust the **Default Level** to apply changes to all servers set to **Default**, or adjust servers' levels individually. If the number of waiting SIPs is rising and SIP processing level is at maximum (and/or cannot be increased in order to reserve UI resources), you may require additional hardware resources to achieve better throughput. Please consult the Rosetta Support Team and Ex Libris Sizing Manager for further analysis.

Users with heavy SIP loading needs may also want to reserve Operational server resources for end-user-facing modules by detaching the Operational servers that handle SIP processing from the load balancer. The load balancer would direct all requests to servers with a low level of SIP processing workers (or none at all), ensuring high availability.

NOTE:

Currently these servers are not dedicated to end user requests, as they also handle Process Automation.

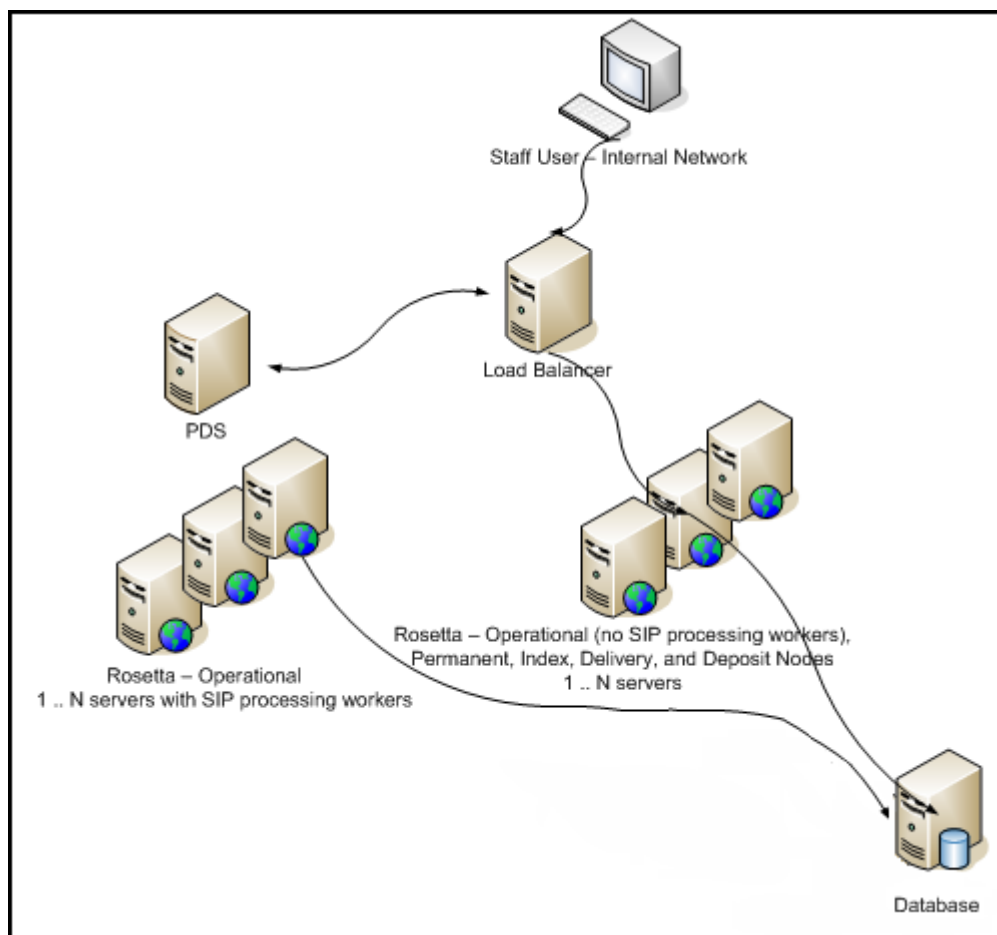


Figure 12: High Availability Server Flow

Libraries that are interested in setting up such a topology should contact Rosetta Support.

SIP Processing Throttling

SIP processing can be configured to refrain from processing additional SIPs when the system load average (during the last minute) is above a certain level. The load average is calculated similar to the top UNIX utility, taking into consideration memory usage and CPU utilization. This setting is useful when unusually large SIPs are deposited sporadically and unpredictably, in order to prevent Rosetta from unexpectedly running out of resources.

Managing Plug-Ins

Managing plug-ins with Rosetta involves accessing file directories on the server and working with the Rosetta Administration interface.

Before you begin:

- Have the JAR file containing your plug-in.

To install and configure the plug-in:

- 1 Upload the JAR file to the application custom plug-in input directory.

The location of the input directory is relative to the `operational_shared` directory (can be found in the `HfrGeneralParameters` table's `operational_shared` parameter). The full path is `.../operational_shared/plugins/custom/`.

- 2 Install the plug-in using the application plug-in management UI.
 - a On the custom Plugin List page (Advanced Configuration > Plug-In Management), click the **Add Plug-In instance** button.

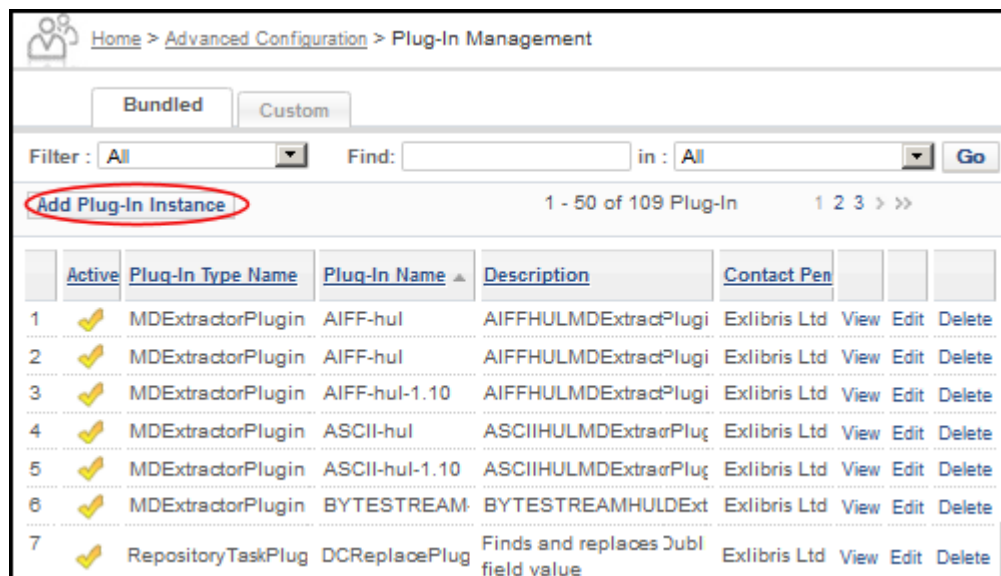
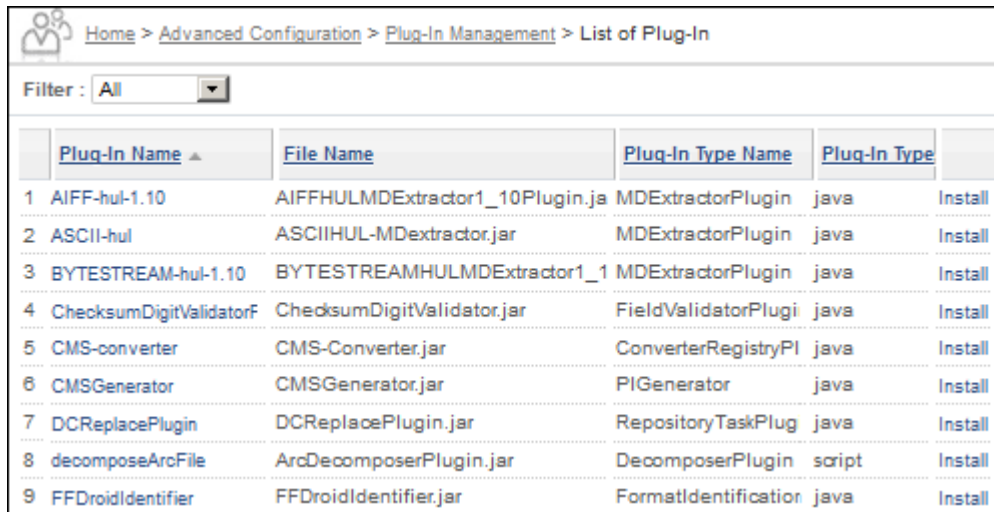


Figure 13: Plug-In Management Page

A list of available plug-ins opens. The plug-in you installed should appear in the list. Scroll or use the Filter or Find feature to help you find it.



The screenshot shows a web interface for managing plug-ins. At the top, there is a breadcrumb trail: Home > Advanced Configuration > Plug-In Management > List of Plug-In. Below this is a filter dropdown menu set to 'All'. The main content is a table with the following columns: Plug-In Name, File Name, Plug-In Type Name, Plug-In Type, and Install. The table lists nine plug-ins, each with a numbered row and an 'Install' link.

	Plug-In Name ▲	File Name	Plug-In Type Name	Plug-In Type	
1	AIFF-hul-1.10	AIFFHULMExtractor1_10Plugin.jar	MExtractorPlugin	java	Install
2	ASCII-hul	ASCIHUL-MDExtractor.jar	MExtractorPlugin	java	Install
3	BYTESTREAM-hul-1.10	BYTESTREAMHULMExtractor1_1	MExtractorPlugin	java	Install
4	ChecksumDigitValidator	ChecksumDigitValidator.jar	FieldValidatorPlugin	java	Install
5	CMS-converter	CMS-Converter.jar	ConverterRegistryPlugin	java	Install
6	CMSGenerator	CMSGenerator.jar	PluginGenerator	java	Install
7	DCReplacePlugin	DCReplacePlugin.jar	RepositoryTaskPlugin	java	Install
8	decomposeArcFile	ArcDecomposerPlugin.jar	DecomposerPlugin	script	Install
9	FFDroidIdentifier	FFDroidIdentifier.jar	FormatIdentifier	java	Install

Figure 14: List of Available Plug-Ins

- b When you find the plug-in you want to install, click the corresponding **Install** link.
- c The Plug-In Information page for that plug-in opens.

The screenshot shows a web-based configuration interface for managing plug-ins. The breadcrumb navigation is: Home > Advanced Configuration > Plug-In Management > Plug-In Information. The page is divided into three main sections: Plug-In Information, Contact Information, and Plug-In Parameters.

Plug-In Information			
Plug-In Id	-	Plug-In Name	DCReplacePlugin
Plug-In Type Name	RepositoryTask	Implementation Type	java
Implementation Name	com.exlibris.d	Interface	-
Owner	-	Material Type	DIGITAL
Family Type	TASK	Description	Finds and replaces a DublinCore
Resource Type	-	Module	Repository
Status	-	Public API	false

Contact Information			
Contact Type	admin	Email address	yaarac@exlibris.co.il
Last Name	Ltd	First Name	Exlibris
Address 1	Agodat Asport2	City	Jerusalem
Telephone 1	(054)485-2328		

Plug-In Parameters	
* Find	<input type="text"/>
* Replace with	<input type="text"/>
* DC field	dc:title

At the bottom of the page, there are 'Back' and 'Cancel' links on the left, and an 'Install' button on the right.

Figure 15: Information for Existing Plug-In

- d Enter a name in the **Plug-In Name** field and a **Description** to reflect it.

NOTE:

Because you are installing and not editing, Rosetta will create a new entry in the plug-in list.

- e Add or change values for all required parameter fields.

NOTE:

If you are installing a variation on an existing plug-in, you must rename the plug-in and enter values in the required parameter fields before Rosetta will allow the installation to be completed.

- f Click install.

The new plug-in is installed in the system and appears on the Plug-In List page, available for use.

To create another plugin instance with different initialization parameters values, repeat step Step **d**, making sure the plug-in name is unique.

To change the initialization parameters of an installed plug-in:

- 1 From the Plug-In Management page (Advanced Configuration > Plug-In Management), find the name of the plug-in whose parameters you want to change and click its corresponding **Edit** link.
- 2 Change the initialization parameter values and click **Save**.

To stop using a given plug-in:

From the Plug-In Management page (Advanced Configuration > Plug-In Management), select the **Delete** link corresponding to the row of the plug-in you want to disable.

To upgrade a (custom) plug-in:

From the Custom tab of the Plug-in List page (Advanced Configuration > Plug-In Management), click the **Upgrade** button.

An upgraded version of the plug-in is deployed.

Disk Space Management

Administrators can configure how disk space is used on the Storage Server, as described in the following sections:

- [Checking Disk Space](#) on page 34
- [Freeing Disk Space](#) on page 35

Checking Disk Space

Administrators can view the following disk space parameters in the directory:

- Total disk space
- Used disk space
- Available disk space

To check disk space in the directory:

Type the following commands:

```
cd /exlibris
df -kh
```

The Rosetta system provides information about disk space, as shown in the following example:

```
dps.corp.exlibrisgroup.com-pl(1) >>cd /exlibris
dps.corp.exlibrisgroup.com-pl(1) >>df -kh
Filesystem          size  used  avail capacity  Mounted on
filer01:/vol/data8/il-dtldev07a
                    30G   26G   3.7G    88%     /exlibris
```

Administrators can run the `df -kh` command for each deposit area in order to see the total amount of disk space used by deposited activities. The location of each deposit area is configured in the `HFrGeneralParameter` table, under the parameter name `logic_deposit_area`.

Freeing Disk Space

Administrators can free an additional amount of disk space by deleting old log files from the log directory (whose alias is `dps_log`).

Rosetta Storage Management

This section contains:

- [Understanding Storage Architecture](#) on page 35
- [Configuring Storage Groups](#) on page 36
- [Configuring Storage Rules](#) on page 47

Understanding Storage Architecture

Administrators can define how Rosetta stores content on the various servers: Deposit, Staging, and Permanent. Administrators configure the following components separately for the Staging Server and the Permanent Repository:

- **Storages**, which are the storage areas. A storage can be a Network File System (NFS) or another type of data storage solution that Rosetta can access through HTTP requests (such as Amazon S3).
- **Storage groups**, which aggregate individual storages that contain similar types of data, such as metadata or intellectual entities (IEs).

Because a storage group can contain multiple storages, Administrators configure storage group policies that define the storage to be used for each content object.

- **Storage rules**, which determine the storage group to be used. Each storage rule consists of the following parameters:
 - Input parameters (such as content object type), with which the Rosetta system compares actual parameters of a content object

- Output parameters, which define the storage group to be used, if actual parameters of a content object match input parameters of the storage rule

The diagram below shows the organization of the storage components in the Rosetta system:

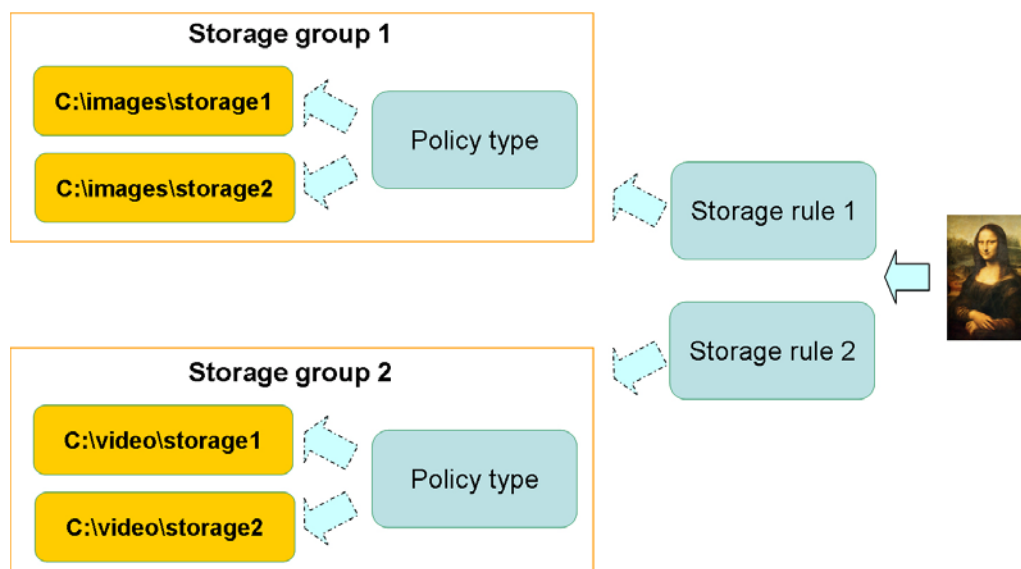


Figure 16: Storage Components

The information flow consists of the following stages:

- 1 After a content object is moved from the Deposit Server to the Staging Server, the Rosetta system sequentially compares the actual content object parameters with the input parameters of the storage rules.
- 2 When the content object parameters match the input parameters of a storage rule, the Rosetta system moves the object to the storage group, as defined in the output parameters of the storage rule.
- 3 Within the storage group, the Rosetta system moves the object to a storage, as defined in the policy type.

Configuring Storage Groups

Administrators work with storage groups using the Storage Groups List page.

#	Name	Description	Policy	Create Date	Update Date	
1	IE Group	Storage media that contains the IE METS files	ORDERED	20/11/2012 21:41:26		Update Delete
2	File Group	Storage media that contains the Stream files	ORDERED	20/11/2012 21:41:26		Update Delete
3	Metadata Storage Permanent Group	The storage media for storing the Metadata in the Perma...	ORDERED	20/11/2012 21:41:26		Update Delete

Figure 17: Storage Groups List page

To access the Storage Groups List page, from the Advanced Configuration page, click **Repository > Storage Rules and Definitions**.

The following actions (on the following pages) can be performed on the Storage Group List page:

- **Adding a Storage Group** on page 37
- **Re-Ordering the List of Storages in a Storage Group** on page 45
- **Updating a Storage Group** on page 45
- **Deleting a Storage in a Storage Group** on page 46

Adding a Storage Group

Administrators can add a new storage group in order to aggregate similar types of data, such as metadata or intellectual entities (IEs). When creating a storage group, Administrators must add at least one storage to save the storage group in the Rosetta system.

Administrators can add as many storage groups as are needed. When adding a storage group, Administrators perform the following actions:

- 1 Define general information about a storage group
- 2 Add at least one storage to the storage group

To add a new storage group:

- 1 On the Storage Group List page (see [Configuring Storage Groups](#) on page 36), click **Add Storage Group**. The Storage Group Edit page opens.

Figure 18: Storage Group Edit Page

- 2 In the **General Information** pane, complete the fields as described in the following table:

Table 2. Storage Group General Information Fields

Field	Description
Name	The name of the storage group.
Description	The description of the storage group.
Policy Type	The criteria that determine which storage is used. The following options are available: <ul style="list-style-type: none"> ■ Random policy - Randomly select a storage that has adequate space available. ■ Orderly policy - Use the first storage (based on the storage order) that has adequate space available.

- 3 In the **Storages** pane, select your storage plug-in.

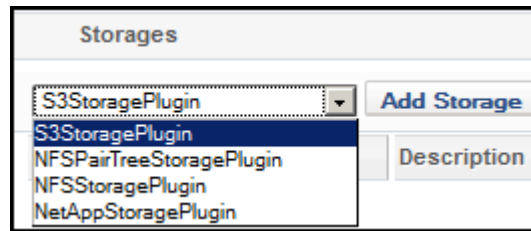


Figure 19: Four Storage Plug-ins

The following options are available:

a S3 Storage Plugin

A storage solution based on the Amazon S3 cloud service. In this situation, Rosetta can be connected to the cloud and the Permanent repository is located on Amazon servers.

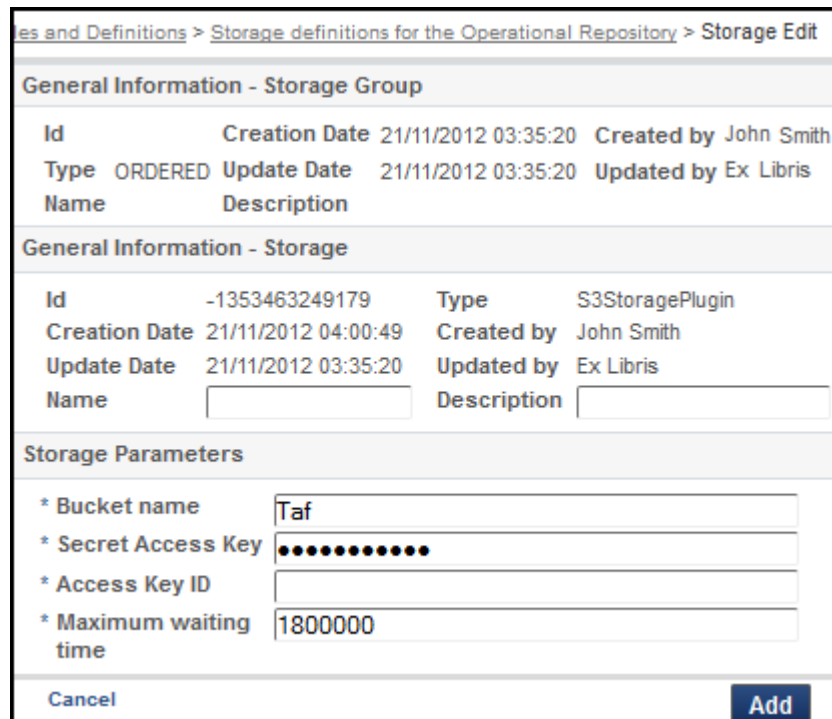


Figure 20: S3StoragePlugin

With this storage plug-in, the following parameters are used:

- Bucket name – This is the term used by Amazon for describing the location of the files in their cloud.
- Secret Access Key – The access key for the bucket.
- Access Key ID – The key ID.

- Maximum waiting time – The time Rosetta will wait for response from Amazon when retrieving files. When the time elapses, Rosetta displays a message to the user that the file is not accessible.

b NFS Pair Tree Storage Plug-in

An NFS plug-in in which the directory structure inside the file system is based on the files' fixity value and not dates.

General Information - Storage			
Id	-1353463294882	Type	NFSPairTreeStoragePlugin
Creation Date	21/11/2012 04:01:34	Created by	John Smith
Update Date	21/11/2012 04:01:22	Updated by	Ex Libris
Name	<input type="text"/>	Description	<input type="text"/>

Storage Parameters	
* Storage Root Directory	<input type="text"/>
* Storage Maximum Capacity in MB	<input type="text"/>
* File Block Size	<input type="text" value="8"/>
* Files Handling Method	<input checked="" type="radio"/> Copy <input type="radio"/> Soft Link <input type="radio"/> Move

Figure 21: NFSPairTreeStoragePlug-in

In the pair tree storage, files are located in a folder that is created based on an MD5 hash of their PID. This increases the speed and efficiency when search algorithms are used to retrieve files.

NOTE:

This plug-in can be used to implement different logic for folders' naming conventions.

c NFS Storage Plugin

The original out-of-the-box plug-in provided by Rosetta. Files are stored conventionally, based on their date of upload.

Rules and Definitions > Storage definitions for the Operational Repository > Storage Edit

General Information - Storage Group			
Id	Creation Date	21/11/2012 03:35:20	Created by John Smith
Type ORDERED	Update Date	21/11/2012 03:35:20	Updated by Ex Libris
Name	Description		

General Information - Storage			
Id	-1353463334860	Type	NFSStoragePlugin
Creation Date	21/11/2012 04:00:49	Created by	John Smith
Update Date	21/11/2012 03:35:20	Updated by	Ex Libris
Name	<input type="text"/>	Description	<input type="text"/>

Storage Parameters	
* Storage Root Directory	<input type="text"/>
* Sub-Directories Prefix	<input type="text"/>
* Amount of Files per Directory	<input type="text"/>
* Storage Maximum Capacity in MB	<input type="text"/>
* File Block Size	<input type="text" value="8"/>
* Files Handling Method	<input checked="" type="radio"/> Copy <input type="radio"/> Soft Link <input type="radio"/> Move

Cancel Add

Figure 22: NFSStoragePlugin

Parameters for the NFS storage plug-in are described below.

Table 3. Storage Parameters for NFS Storage

Field	Description
Storage Parameter Name	The name of the storage parameter.
Root Directory	The root directory of the storage. NOTE: The system automatically adds a suffix to the root directory with a date stamp when creating the directory name. For example, e:\storage1\data\2008\06\10.

Table 3. Storage Parameters for NFS Storage

Field	Description
Directory Prefix	<p>The directory prefix that is part of the directory name. For example, if the directory prefix is <code>file\$</code>, the system creates the following directories:</p> <ul style="list-style-type: none"> ■ <code>e:\storage1\data\2008\06\10\file\$1</code> ■ <code>e:\storage1\data\2008\06\10\file\$2</code> ■ <code>e:\storage1\data\2008\06\10\file\$3</code> <hr/> <p>NOTE: The directory prefix is suffixed with a sequential counter.</p> <hr/>
Files Per Directory	<p>The maximum number of files that can be saved in the storage. When the maximum number of files that can be saved in the directory is reached, the system creates a new directory.</p>
Storage Max Size (MB)	<p>The total amount of data that can be saved in the storage.</p>
Storage Block Size	<p>The storage block size is used for an exact calculation of the available disk space.</p> <hr/> <p>NOTE: It is strongly recommended that you leave this field at its default value.</p> <hr/>

Table 3. Storage Parameters for NFS Storage

Field	Description
Files Handling Method	<p>Select one method from the three available:</p> <ul style="list-style-type: none"> ■ Copy the files from their previous location (Deposit or Operational) to this storage (Operational or Permanent) so that two copies of the processed files exist in the system. ■ Move the files from their previous location (Deposit or Operational) to this storage (Operational or Permanent) so that only one copy of the file exists in the system. ■ Soft Link for Operational storage definitions, useful for handling large files, thus avoiding copying and moving them twice. <hr/> <p>NOTE: Because derivative copy files are not moved and copied to the Permanent repository, they must be moved and copied from the Deposit file system to the Operational repository. Otherwise, if soft-copied, they might get lost when the Deposit file system is cleaned. Therefore, an out-of-the-box rule has been added that makes sure that these files are always copied.</p> <hr/> <p>NOTE: Move policy should only be used where operational and permanent storage are on the same mount; otherwise, there is no performance advantage (since streams are always removed from operational after SIP reaches permanent).</p>

d Net App Storage Plugin

The NetApp plug-in is a storage solution by NetApp, Inc., that integrates with the Rosetta storage system.

General Information - Storage			
Id	-1353463366898	Type	NetAppStoragePlugin
Creation Date	21/11/2012 04:02:46	Created by	John Smith
Update Date	21/11/2012 04:02:40	Updated by	Ex Libris
Name	<input type="text"/>	Description	<input type="text"/>

Storage Parameters	
* Container	<input type="text"/>
* Protocol	<input type="text"/>
* Host	<input type="text"/>
* Port	<input type="text"/>
* User Name	Tf
* Password	●●●●●●●●

Figure 23: Parameters for NetApp Plug-in

The NetApp application manages the storage of files in a combination of tapes and disks. Rosetta interacts only with the management layer, which stores and retrieves files for Rosetta.

4 Click **Add Storage**.

The Storage Edit page for the storage plug-in you chose opens.

5 In the **General Information - Storage** pane, enter a name and description for the storage.

6 In the **Storage Parameters** pane, complete the fields as described in the relevant section above. (For NetApp parameters, see the documentation for the NetApp storage plug-in.)

7 Click **Save**. The new storage is displayed on the Storage Group Edit page.

NOTE:

The storage is not added to the Rosetta system until you click **Save** on the Storage Group Edit page (see Step 8).

8 To add more storages, repeat Step 3 - Step 6.

9 On the Storage Group Edit page, click **Save**.

The storage group and its related storages are saved in the Rosetta system.

Re-Ordering the List of Storages in a Storage Group

To define a storage that must be used for a specific content object, the Rosetta system compares the parameters of the storage with the content object parameters (such as file size), using the policy type.

Storages are analyzed in the same order that they are displayed on the List of Delivery Rules page. The Rosetta system uses the first storage found that matches the parameters of the content object.

An Administrator can re-order storages to change their priority.

To re-order the list of storages:

- 1 On the Storage Groups List page (see [Configuring Storage Groups](#) on page 36), locate the storage group with which you want to work and click **Update**. The Storage Group Edit page opens (see [Figure 24](#)).

Repository > Storage Rules and Definitions > Storage definitions for the Operational Repository > Storage Group Edit

General Information

Id: 1 Updated by: Admin
Creation Date: 20/11/2012 21:41:28 Created by: Admin
Update Date:
Name: IE Group Policy Type: ORDERED
Description: Storage media that contains the IE METS files

Storages

S3StoragePlugin Add Storage

#	Name	Description	Type	Creation Date	Update	Delete
1	IE Storage disk 1	Attributes of disk 1 IEs in ...	NFSSStoragePlugin	20/11/2012 21:41:28	Update	Delete
2	IE Storage disk 2	Attributes of disk 2 IEs in ...	NFSSStoragePlugin	20/11/2012 21:41:28	Update	Delete

Cancel Save

Figure 24: Storage Group Edit page

- 2 In the **Storages** pane, use the up and down arrows to change a storage's priority.
- 3 Click **Save**.

The Rosetta system now analyzes the storages in the updated order.

Updating a Storage Group

Administrators can update a storage group to add more storages, modify the existing storages, or change the descriptive information of the storage group.

To update a storage group:

- 1 On the Storage Group List page (see [Configuring Storage Groups](#) on page 36), locate the group you want to update and click **Update**. The Storage Group Edit page opens.
- 2 Modify the fields you want to update and click **Save**.

The storage group is updated in the Rosetta system.

Deleting a Storage in a Storage Group

Administrators can delete a storage in a storage group.

NOTE:

The Rosetta system does not enable Administrators to delete a storage that contains data.

To delete a storage in a storage group:

- 1 On the Storage Group List page (see [Configuring Storage Groups](#) on page 36), locate the storage group with which you want to work and click **Update**. The Storage Group Edit page opens.
- 2 In the **Storages** pane, locate the storage you want to delete and click **Delete**. The confirmation page opens.
- 3 Click **OK**.

The storage is deleted from the Rosetta system.

Deleting a Storage Group

Administrators can delete a storage group.

NOTE:

The Rosetta system does not enable Administrators to delete a storage group that contains at least one storage item.

To delete a storage group:

- 1 On the Storage Group List page (see [Configuring Storage Groups](#) on page 36), locate the group you want to delete and click **Delete**. The confirmation page opens.
- 2 Click **OK**.

The storage group is deleted from the Rosetta system.

Storage Space Checker

A Storage Space Checker (implemented as a Startup Check plug-in) checks each storage group for available space, as defined by the aggregation of the Storage Max Size values for all storages (with a Storage Max Size) in a given group. Storage groups comprised of one or more storages with no Storage Max Size (for example, Amazon s3 storage) are not checked.

The default thresholds are 5% for a warning and 1% for an error, and can be defined in the UI of the Storage Space Checker plug-in instance (under Plugin Management). Threshold values are global for all (relevant) storage groups.

Warnings and errors appear in the System Check UI with the details of the relevant storage group.

NOTE:

Rosetta storage space definitions are logical and presume the existence of available physical disk space, which must be monitored separately.

Configuring Storage Rules

Administrators configure storage rules to determine which storage group the Rosetta system must use for specific content objects.

Administrators define storage rules from the Rule List page and the Rule Details page.

Enabled	Order	Name	Description	Created	Modification Date			
1	<input checked="" type="checkbox"/>	Derivative Copies Rule	storing Derivative Copies...	21/11/2012 20:44:13		Update	Duplicate	Delete

Default Rule:			
Repository Def...	Default rule for storing files in the Staging repo...	21/11/2012 20:44:13	Update

Figure 25: Rule List page

To access this page from the Advanced Configuration page, click **Repository > Storage Rules and Definitions**, and select one of the options for configuring storage rules (for example, Permanent Repository - IEs, or Operational Repository).

The following actions (on the following pages) can be performed from the Rule List page:

- **Adding a Storage Rule** on page 50
- **Updating a Storage Rule** on page 51
- **Defining the Default Storage Rule** on page 52
- **Re-Ordering the List of Storage Rules** on page 52
- **Deleting a Storage Rule** on page 52
- **Activating and Deactivating a Storage Rule** on page 53

Rule Details Page

If you are adding or editing a specific rule, you will work on the Rule Details page to define the parameters of the rule. See the **Operators Used in Rule Parameters** section for details on how to use operators in parameters.

Operators Used in Rule Parameters

The following operators are used for specific types of parameter data.

String Values

String values are words that are not separated by a comma (,), for example, one Producer name (John Smith), one MIME type (audio/mp3), one error code, one Format ID). String values use the following operators:

- Equal – The string and the input value must match exactly.
- Contains – The string and the input value must match partially with the ‘*’ character.

List of Strings

A list of strings is a list of string values separated by a comma (,) sometimes populated by a widget. Lists of strings use the following operators:

- List Contains – used when each error returned should match exactly a single given error in the rule.
- List Equals - Used when the order of the items in the list and the list itself should match exactly. For example, a rule defined as “Invalid page dictionary object, Invalid object number in cross-reference stream” will match to the actual output from JHOVE – “Invalid page dictionary object, Invalid object number in cross-reference stream.”

Numeric Fields

Numeric fields (for example, file size) use numbers as matching and comparison values.

- Greater Than (>) – The input value should be greater than the parameter value.
- Less Than (<) - The input value should be less than the parameter value.
- Equal (=) - The input value should be equal to the parameter value.
- Not Equal (!=) - The input value should be not equal to the parameter value.

Date Fields

Date fields (such as Creation Date) compare date values with time operators.

- After – The input date should be later than the parameter date value.
- Before - The input date should be earlier than the parameter date value.
- Equal (=) - The input date should be the same as the parameter date value.
- Not Equal (!=) - The input date should not be the same as the date parameter value.

Any

All fields can use this operator for indicating that any input value will be accepted by the rule. For example, if the 'Any' operator is used in the Producer Name field, the rule can match all Producers.

The following table summarizes the possibilities for matching between the rule parameter values and the run-time values:

Table 4. Possible Matches Between Rule Parameter and Run-Time Values

Run-time Value	Operator	Possible Rule Values	Result
Demo Producer	Equal	Demo Producer	Match
Demo Producer	Contains	Demo*	Match
image/tiff or image/bmp	In List	Image/tiff, image/bmp	Match
image/tiff, image/bmp	List Equals	Image/tiff, image/bmp	Match
grey or gray	In List with Regular Expression	gr[ea]y	Match

Table 4. Possible Matches Between Rule Parameter and Run-Time Values

Run-time Value	Operator	Possible Rule Values	Result
12345	<, >, =, !=	10000	< - No match > - Match = - No match != - Match
23/11/2011	Before, After, =, !=	23/11/2011	Before - No match After - No match = - Match != - No match

To define Boolean logic when using multiple conditions, select one of the following options between conditions:

- OR
- AND (default)

NOTE:

The Boolean connector between different types of attributes (for example, IE Attributes and File Attributes) is always AND.

Adding a Storage Rule

Administrators can add a new storage rule to define the automatic selection of a specific storage group.

Administrators can add as many storage rules as needed.

To add a new storage rule:

- 1 On the Rule List page (see [Configuring Storage Rules](#) on page 47), click **Add New Rule**. The Rule Details page opens.

Dnx Section	Dnx Section Key	Operator	Value
webHarvesting	primarySeedURL	Any	

Parameter	Result
Storage Group	Storage Repository Group

Figure 26: Rule Details Page

- 2 Enter a **Name** and a **Description** for the new rule.
- 3 Enter parameters for the rule. Use the [Operators Used in Rule Parameters](#) section to create parameters.
- 4 Click **Save**.

The Rosetta system uses the storage rule in order of priority, as described in [Re-Ordering the List of Storage Rules](#) on page 52.

Updating a Storage Rule

Administrators can update a storage rule in order to modify input and output parameters.

To update a storage rule:

- 1 On the Rule List page (see [Configuring Storage Rules](#) on page 47), locate the rule you want to update and click **Edit**. The Rule Details page opens.
- 2 Modify the fields you want to update, and then click **Save**.

The storage rule is updated in the Rosetta system.

Defining the Default Storage Rule

Administrators must configure a default storage rule. In the event that the Rosetta system cannot find a rule that matches any of the storage rules, the system uses the default storage rule.

To define the default storage rule:

- 1 On the Rule List page (see [Configuring Storage Rules](#) on page 47), in the **Default Rule** pane, click **Update**. The Default Rule Details page opens.
- 2 In the **Output Parameters** pane, in the **Result** drop-down list, select the storage group to be used by default.
- 3 Click **Save**.

The Rosetta system now uses the selected storage rule as the default rule.

Re-Ordering the List of Storage Rules

To define a storage group that must be used for storing an object, the Rosetta system compares the input parameters defined in a storage rule with the parameters of the storage group.

Storage rules are analyzed in the same order in which they are displayed on the Rule List page. The Rosetta system uses the first storage rule found that matches the parameters of the storage group.

To re-order the list of storage rules:

- 1 On the Rule List page (see [Configuring Storage Rules](#) on page 47), select each relevant rule and use the up and down arrows to change the rule's priority.
- 2 Click **Save**.

The Rosetta system now processes the storage rules in the defined order.

Deleting a Storage Rule

Administrators can delete an existing storage rule. After a storage rule is deleted, it is no longer available to the Rosetta system for matching.

To delete a storage rule:

- 1 On the Rule List page (see [Configuring Storage Rules](#) on page 47), locate the storage rule you want to delete and click **Delete**. The confirmation page opens.
- 2 Click **OK**.

The storage rule is deleted from the Rosetta system.

Activating and Deactivating a Storage Rule

Administrators can activate or deactivate a storage rule. After a storage rule is deactivated, it is no longer available to the Rosetta system for matching.

On the Rule List page, the status of the storage rule is indicated by the check mark in the **Active** column:

- Yellow - The storage rule is active.
- Grey - The storage rule is inactive.

To activate or deactivate a storage rule:

- 1 On the Rule List page (see [Configuring Storage Rules](#) on page 47), locate the storage rule you want to activate or deactivate.
- 2 In the **Active** column, click the check mark. The check mark in the **Active** column indicates the new status.

The storage rule is changed from active to inactive, or from inactive to active.

Configuring External Metadata: SRU/SRW

External metadata about content objects in Rosetta is stored in external systems such as collection management systems. To enable Rosetta to communicate with external systems, Administrators can edit the configuration files from the Advanced Configuration page by clicking **Repository > External Interfaces > SRU/SRW sources and definitions**. The Configuration Files page opens to the SRU/SRW sub-group (below).

Filename	Description	Updated by	Update Date	
1 explain.properties	defines SRW/SRU explain operation properties	7/6/10	7/6/10	View Edit
2 external_resource_explorer_configuration.xml	define property file for voyager external database	7/20/10	7/20/10	View Edit

Figure 27: SRU/SRW Configuration Files

The fields are described in the following table:

Table 5. External Metadata Configuration Files

File	Defines...
explain.properties	Parameters of the Rosetta system server
external_resource_explorer_configuration.xml	Parameters of an external system server

The following tasks must be performed in the configuration of SRU:

- **Configure Rosetta Parameters – External Resource File** on page 54
- **Configure Rosetta Parameters – Explain File** on page 57
- **Adaptation Needed in the SRU Server** on page 57

Configure Rosetta Parameters – External Resource File

Use the following procedure to configure Rosetta parameters for external files:

- 1 With the **File Group** field set to **External Interface** and the **Sub-Group** set to **SRU/SRW**, click **Edit** for the file named `external_resource_explorer_configuration.xml`.

The file opens on the page. It contains all SRU configurations that are delivered with Rosetta.

- 2 Add your configuration to the file. The parameters that should be modified are:
 - `baseUrl` – the URL of the system that acts as an SRU server (CMS)
 - `version` – SRU version (currently 1.1 is supported by Rosetta)
 - `operation` – since Rosetta is the client, “searchRetrieve” is the only option
 - `recordSchema` – should be “dps”
 - `indexName` – should be “rec.id=”
 - `recordPacking` – should be “xml”
 - `updateUrl` – in cases where the server can get a response for updating the exists-in-Rosetta flag, the url should be added here (currently available only for Voyager and Aleph. For other systems – leave it empty).

The following is an example of an SRU configuration for Aleph and Voyager:

```
<?xml version="1.0" encoding="UTF-8"?>
<ExternalResourceExplorer xmlns="http://www.loc.gov/zing/srw/
configuration/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:xml="http://www.w3.org/XML/1998/namespace">
  <RepositoryName name="qa654mkdb">
    <protocol type="SRW">
      <parm name="baseUrl">http://10.100.2.36:17191/voyager</parm>
      <parm name="version">1.1</parm>
      <parm name="operation">searchRetrieve</parm>

      <parm name="recordSchema">dps</parm>
      <parm name="indexName">rec.id</parm>
      <parm name="recordPacking">xml</parm>
      <parm name="updateUrl">http://10.100.2.36:17114/vxws/
updateDPSFlag</parm>
    </protocol>
  </RepositoryName>

  <RepositoryName name="USM01" type="ALEPH">
    <protocol type="SRW">
      <parm name="baseUrl">http://il-aleph07:5667/usm01</parm>
      <parm name="version">1.1</parm>
      <parm name="operation">searchRetrieve</parm>
      <parm name="recordSchema">dps</parm>
      <parm name="indexName">rec.id</parm>
      <parm name="recordPacking">xml</parm>

      <parm name="updateUrl">http://il-aleph07:8997/
X?op=ros_doc&library=USM01</parm>
      <parm name="detachUrl">http://il-aleph07:8997/
X?op=ros_doc_del&library=USM01</parm>
    </protocol>
  </RepositoryName>
</ExternalResourceExplorer>
```

IMPORTANT:

It should be noted that `RepositoryName` must equal the name that is returned in the SRU/SRW response. For example, if

```
<identifier>DPS:10.100.2.36:ALEPH01:9611</identifier>
```

then the name attribute of `RepositoryName` must be `ALEPH01`, as in

```
<RepositoryName name="ALEPH01" type="ALEPH">
```

To get the parameters from the SRU server, call the `explain` URL.

The following is an example of the `explain` URL from the Aleph system:

```
<zs:explainResponse>
<zs:version>1.1</zs:version>
<zs:record>
<zs:recordSchema>http://explain.z3950.org/dtd/2.0/</
zs:recordSchema>
<zs:recordPacking>xml</zs:recordPacking>
<zs:recordData>
<explain>
<serverInfo>
<host>il-aleph07</host>
<port>9997</port>
<database>usm01</database>
</serverInfo>

<indexInfo>
<set identifier="info:srw/cql-context-set/1/cql-v1.1"
name="cql"/>
<set identifier="info:srw/cql-context-set/1/dc-v1.1"
name="dc"/>
<set identifier="http://zing.z3950.org/cql/bath/2.0/"
name="bath"/>
<index id="4">
<title>title</title>
<map>
<name set="dc">dc.title</name>
</map>
</index>
<index id="48">
<title>identifier</title>

<map>
<name set="dc">dc.identifier</name>
</map>
</index>
<index id="12">
<title>rec.id</title>
<map>
<name set="dc">rec.id</name>
</map>
</index>
</indexInfo>
</explain>
</zs:recordData>
</zs:record>
```

```
<zs:diagnostics>
<diagnostic>
<uri>info:srw/diagnostic/1/7</uri>
<message>Mandatory parameter not supplied</message>
<details>version</details>
</diagnostic>
<diagnostic>
<uri>info:srw/diagnostic/1/7</uri>
<message>Mandatory parameter not supplied</message>
<details>operation</details>
</diagnostic>
</zs:diagnostics>
</zs:explainResponse>
```

Configure Rosetta Parameters – Explain File

Use the following procedure to configure the `dbInfo` parameter in the `explain` file:

With the **File Group** field set to **External Interface** and the **Sub-Group** set to **SRU/SRW**, click **Edit** for the file named `explain.properties` and enter your SRU (Rosetta) details for `dbInfo={DPS SRU Database}`

For example,:

```
dbInfo=DPS SRU Database
```

Adaptation Needed in the SRU Server

Rosetta should store the CMS ID and metadata in the repository and link it to an IE. In order for Rosetta to perform the link, the following metadata should be returned as part of the response:

```
<identifier>{DPS}:{IP}:{NAME}:{CMS-ID}</identifier>
```

where:

- `{DPS}` = constant text
- `{IP}` = the IP of the server
- `{NAME}` = the name of the repository
- `{CMD-ID}` = the CMS ID

For example:

```
<identifier>DPS:10.100.2.36:qa654mkdb:9611</identifier>
```

Note: The response should be in DC format only.

The query:

```
http://10.100.2.36:17191/  
voyager?version=1.1&operation=searchRetrieve&query=title=boy&max  
imumRecords=10&recordSchema=dps
```

2

General Parameters

This section contains general parameters for the following modules:

- **Backoffice** on page 59
- **BIRT** on page 60
- **Delivery** on page 64
- **Deposit** on page 65
- **Format Library** on page 66
- **General** on page 67
- **Menu** on page 69
- **Network** on page 70
- **Authentication** on page 70
- **Repository** on page 71
- **SMTP Email** on page 72
- **Search** on page 73

General parameters define global settings that determine how the entire Rosetta system functions. The following sections describe general parameters by module.

NOTE:

Descriptions that are blank here will be added to the guide as they become available.

Backoffice

This module contains parameters that define settings for the Back Office, including default page size and password length.

Table 6. Backoffice Parameters

Parameter	Value	Description
file_data_in_producer_email	False	A boolean that indicates if the files' details will be sent as part of the email to the Producer Agent
menus_max_links	6	Maximum number of quick links that appear in the Advanced Configuration and Management menus.
email_updating_new_user	true/false	If true, the system sends a confirmation email to a newly registered user. If false, no confirmation email is sent.
email_sender	Management Department <do-not-reply@exlibris.co.il>	The string that shows in the sender field when emails are sent by the system to staff users.
email_contact	library@exlibris.co.il	The email address that appears as the contact address on system-generated email.
password_size	8	Minimal length of users' passwords
max_file_size_to_download	5	Maximum file size that can be downloaded (TA, Web editor). NOTE: If the file is bigger than the set value, it is copied to the server under: <code>{operational_shared}/staff_work_area/{userName}/export/streams/{pid}/{fileName}</code>
max_input_characters_allowed	2048	Maximum length for SIPs reject reason

BIRT

This module contains parameters that define BIRT reporting module settings.

Table 7. BIRT Parameters

Parameter	Value	Description
high_file_system	90	The threshold for activating the red traffic light on the file system report dashboard. The light turns red if there is less than 10% of file system space available.
high_IOWait	90	The threshold for activating the red traffic light on the CPU Status dashboard. The light turns red if 90% of the events generated in the past two hours indicate that the input-output (IO) wait is too long.
high_load	6	The threshold for activating the red traffic light on the CPU Status dashboard. The light turns red if six events generated in the past two hours indicate that the CPU is fully used.
high_memory	90	The threshold for activating the red traffic light on the memory report dashboard. The light turns red if there is less than 10% memory space available.
low_free_table_space	10%	The threshold for activating the red traffic light on the CPU Status dashboard. The light turns red if there is less than 10% of table space available.
com.exlibris.dps.adm.general.menu	adm_dashboard.rptdesign	The URL suffix of the reports from the Advanced Configuration home page
com.exlibris.dps.wrk.general.menu	most_viewed_objects.rptdesign	The URL suffix of the reports from the Management home page
com.exlibris.dps.wrk.general.menu.producer	total_producer_group_deposits.rptdesign	The URL suffix of the reports from the Producers menu page
com.exlibris.dps.wrk.general.menu.Submission	sip_processing_stats_pie.rptdesign	The URL suffix of the reports from the Submission menu page
com.exlibris.dps.wrk.general.menu.DataMng	file_type_groups_in_permanent.rptdesign	The URL suffix of the reports from the Data Management menu page
com.exlibris.dps.wrk.general.menu.Preservation	formats_at_risk_lobby.rptdesign	The URL suffix of the reports from the Preservation menu page
com.exlibris.dps.adm	adm_dashboard.rptdesign	Reports from the home page (Advanced Configuration)

Table 7. BIRT Parameters

Parameter	Value	Description
com.exlibris.dps.wrk	most_viewed_objects.rptdesign	Reports from the home page (Management)
com.exlibris.dps.wrk.general.menu.Preservation.AdvancedPreservationActivities.mngLibraryGlobal	file_type_groups_in_permanent.rptdesign'	The URL of the report that is shown in the global format library menu
com.exlibris.dps.wrk.general.menu.Preservation.AdvancedPreservationActivities.mngLibraryLocal	file_type_groups_in_permanent.rptdesign	The URL of the report that is shown in the local format library menu
com.exlibris.dps.wrk.general.menu.Submission.Advanced.Config	sip_processing_stats_pie.rptdesign	The URL of the report that is shown in the Submission - Configure Settings menu
com.exlibris.dps.adm.general.menu.advanced.general.generalHeader	adm_dashboard.rptdesign	The URL of the report that is shown in the Advanced Configuration sub-menus
com.exlibris.dps.adm.general.menu.advanced.SystemProcesses.SystemProcessesHeader	adm_dashboard.rptdesign'	The URL of the report that is shown in the System Process menu under Advanced Configuration
com.exlibris.dps.wrk.general.menu.Preservation.RiskAnalysis.reports	formats_at_risk_lobby.rptdesign'	The URL of the report that is shown in the Preservation - Risk Analysis menu
com.exlibris.dps.wrk.general.menu.producer.Advanced.ConfigureSettings	total_producer_group_deposits.rptdesign'	The URL of the report that is shown in the Producers - Configure Settings menu
com.exlibris.dps.adm.general.menu.advanced.deposit.deposit	adm_dashboard.rptdesign	The URL of the report that is shown in the Deposit Settings menu under Advanced Configuration
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader	adm_dashboard.rptdesign'	The URL of the report that is shown in the Repository menu under Advanced Configuration
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader.MD.MDHeader	adm_dashboard.rptdesign	The URL of the report that is shown in the Repository sub-menu under Advanced Configuration
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader.generalHeader	adm_dashboard.rptdesign	The URL of the report that is shown in the Repository sub-menu under Advanced Configuration

Table 7. BIRT Parameters

Parameter	Value	Description
com.exlibris.dps.wrk.general.menu.Preservation.AdvancedPreservationActivities.config	formats_at_risk_lo bby.rptdesign	The URL of the report that is shown in the Preservation - Configure Setting menu
com.exlibris.dps.adm.general.menu.advanced.SIPProcessing.SIPProcessingHeader	adm_dashboard.r ptdesign	The URL of the report that is shown in the SIP processing menu under Advanced Configuration
com.exlibris.dps.adm.general.menu.advanced.SIPProcessing.SIPProcessingHeader.conf.confHeader	adm_dashboard.r ptdesign	The URL of the report that is shown in the SIP processing sub-menu under Advanced Configuration
com.exlibris.dps.adm.general.menu.advanced.delivery.delivery	adm_dashboard.r ptdesign	The URL of the report that is shown in the Delivery menu under Advanced Configuration
com.exlibris.dps.wrk.general.menu.DataMng.Advanced.config	file_type_groups_i n_permanent.rptdesign	The URL of the report that is shown in the Data Management - Configure Setting menu
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader.ExternalInt.ExternalIntHeader	adm_dashboard.r ptdesign	The URL of the report that is shown in the External Interfaces menu under Advanced Configuration > Repository
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader.StorageRulesAndDefinitions.StorageRulesAndDefinitionsHeader	adm_dashboard.r ptdesign'	The URL of the report that is shown in the Storage definitions menu under Advanced Configuration > Repository
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader.utilities.utilitiesHeader	adm_dashboard.r ptdesign	The URL of the report that is shown in the Utilities and Files menu under Advanced Configuration > Repository
com.exlibris.dps.adm.general.menu.advanced.repository.RepositoryHeader.PersistentIdentifiersHeader	adm_dashboard.r ptdesign	The URL of the report that is shown in the Persistent Identifiers menu under Advanced Configuration > Repository
com.exlibris.dps.adm.general.menu.advanced.usersAndRoles.usersAndRolesHeader	adm_dashboard.r ptdesign	The URL of the report that is shown in the User Management menu under Advanced Configuration

Table 7. BIRT Parameters

Parameter	Value	Description
com.exlibris.dps.adm.general.menu.advanced.NonFunctional.NonFunctionalHeader	adm_dashboard.rptdesign	The URL of the report that is shown in the Multi-Language Support menu under Advanced Configuration
com.exlibris.dps.wrk.general.menu.producer.Advanced.ConfigureSettings.MaterialFlows	file_type_groups_in_permanent.rptdesign	The URL of the report that is shown in the Producers - Configure Setting > Material Flow Building Blocks menu

Delivery

This module contains parameters that define the behavior of the Delivery module.

Table 8. Delivery Parameters

Parameter	Value	Description
allow_access_by_default	True	Defines whether users can view an object when no specific access rights policy is defined for this object
anonymize_event	True	Replace the last octet of the user's IP address with a 0 in the delivery event.
anonymize_mets	True	The METS delivery does not include the producer information section
check_thumbnail_access_rights	True	Determines if access rights should be checked for thumbnails
client_ip_header	Rosetta-User-IP	Header location of the user IP address validated by Access Rights
collection_allow_navigation_strict	True	Collections with the Allow Navigation check box cleared are not displayed in the collection viewer
collection_page_size	10	The number of collections to be displayed in a collection viewer page. When set to 0 the collection display is optimized and fills the available space.
delivery_show_login_logout	False	Displays a Logout link during a registered user's delivery session and a Login link during a guest user's delivery session.

Table 8. Delivery Parameters

Parameter	Value	Description
display_j2k_plugin	False	Defines whether the J2K ActiveX must be embedded in the J2K Viewer (an alternative for the default J2K viewing application).
ie_thumbnail_limit	5	Max number of IE thumbnails to be displayed.
max_delivery_file_size_mb	0	Maximum file size for delivery
use_sso	True	If a user logs into Rosetta via Delivery, the session remains active for subsequent delivery request.

Deposit

This module contains parameters that define Deposit process settings, such as default screen size and event logging.

Table 9. Deposit Parameters

Parameter	Value	Description
fastTrack	False	Determines whether the fast track mechanism is enabled for Casual Producers
email_sender	Management Department <do-not-reply@exlibris.co.il>	The email address that appears as the source of the emails that are sent by the system to Producer Agents
email_contact	library@exlibris.co.il	The email address listed as the contact address on system-generated e-mails
applet_license_key	none	License key for the Aurigma uploader applet (Aurigma licensed users only)

Table 9. Deposit Parameters

Parameter	Value	Description
logic_deposit_area	@@dps.nfs.storage.dep@@	Path to directory used as the storage of the Deposit. Changing this value is possible only if there are no SIPs in the TA Deposit, Loading, or Validation sections nor deposit jobs with the status of Rejected or Draft.
deposit_ws_max_response	100	Maximum number of deposit activities to be returned in SDK Web services
validate_with_rosetta_mets	True	Validate submitted METS according to the Rosetta METS xsd. If true, invalid METS will be rejected
validate_with_mets	True	Validate submitted METS according to the METS xsd. If true, invalid METS will be rejected
use_captcha	True	Determines whether CAPTCHA validation will be used in self registration

Format Library

This module contains parameters that define settings for the format library.

Table 10. Format Library Parameters

Parameter	Value	Description
ftp_server	ftp.exlibris-usa.com	The FTP server that holds the Format Library versions that are ready for import
ftp_user	localFL	The FTP user name
ftp_pw		The FTP password

Table 10. Format Library Parameters

Parameter	Value	Description
fl_staff_email		The email address of the staff user who will be alerted when a new version is ready for import!
current_version	3.02	The Current version of the Format Library
format_library_is_global	false	Displays the global format library configuration menu.
use_sftp	False	Determines whether the format library update job uses sftp (SSH File Transfer Protocol).

General

This module contains parameters that define general settings, such as system time format.

Table 11. General Parameters

Parameter	Value	Description
apply_same_origin_policy	True	The HTTP header parameter X-FRAME-OPTIONS is configured to be SAMEORIGIN to prevent click jacking attacks.
collection_show_id	False	Displays the Collection ID alongside the collection name in the collection tree (Collection management and Search)
default_language	en	Default language for PDS
digital_certificate	digital_certificate.pfx	Name of the digital certificate file used for email signing. Stored in the operational shared directory
digital_certificate_password	Password	Password of the digital certificate used for email signing

Table 11. General Parameters

Parameter	Value	Description
digital_signature	True	Use of digital certificate to sign emails is permitted
dublincore_additional_namespaces		Additional Dublin Core namespaces for the Web editor. For example, add the following if you use namespaces in your descriptive metadata that are not standard dc or dcterms: <code>xmlns:mods="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-0.xsd"</code>
email_regex	<code>(^[0-9a-zA-Z]([-.\w]*[0-9a-zA-Z])*\@[0-9a-zA-Z]([-.\w]*[0-9a-zA-Z])\.)+[a-zA-Z]{2,9})\$) \s*</code>	Regular expression validation rules for email addresses
embedded_mode_session_timeout	number	Session time-out in seconds
export_flat	False	Boolean for indicating if the exported files will be tarred before they are exported
installation_code	DPS	
num_of_pages_per_pagination	3	The number of previous/next pages that will be available in pagination mode. For example " <code><<< 1 2 3 >>></code> ", " <code><<< 5 6 7 >>></code> "
url_regex	<code>^http\:\/\/[a-zA-Z0-9\-\.\+\.][a-zA-Z]{2,3}(\S*)?\$</code>	Regular expression validation rules for URL
system_time_format	HH:mm:ss	The format in which time is presented in the system - Hours:minutes:seconds
phone_regex	<code>((\s)*?\d(\s)*?-)*(\d)?(\s)*</code>	Regular expression validation rules for phone numbers in DPS
help_url	<code>http://il-dtldev08b.corp.exlibrisgroup.com:1801/</code>	Help page URL

Table 11. General Parameters

Parameter	Value	Description
isbn_regex	ISBN\x20(?=.{13}\$)\d{1,5}([-])\d{1,7}\1\d{1,6}\1(\d X)\$	Regular expression validation rules for ISBN strings in DPS
page_size	10	Number of list items displayed on the page.
operational_shared	@@dps.nfs.shared.all@@	Location of the operational shared directory
pronom_server_name	http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=detailReport&id=	URL for the format page in the Format library
enforce_ar_export	False	Enforces Access Rights during IE export (Search and Web Editor UIs only)
max_sys_load_avg	Default=0=no throttling	System load average level above which SIPs will wait in queue. NOTE: For more information, see SIP Processing Throttling on page 30.

Menu

This module contains parameters that define behavior of the menu.

Table 12. Menu Parameters

Parameter	Value	Description
graph	/mng/images	Directory in which graphs used in reports are stored

Network

This module contains parameters that define network settings, such as inactivity connection time-out.

Table 13. Network Parameters

Parameter	Value	Description
external_read_timeout	10	Http read timeout in milliseconds (when reading from an input stream)
external_connection_timeout	10	Http open connection timeout in milliseconds (when establishing a connection to the host)

Authentication

This module contains parameters that define the authentication settings.

Table 14. Authentication Parameters

Parameter	Value	Description
login_default_institution	<institution code> – for example, INS00	The institution that appears as the default on the login page. NOTE: The selected institution can be overridden by an institution querystring parameter., for example: <code>https://rosetta.myinstitution.edu/mng/login?auth=local&institution=INS01</code>
authentication_max_failure	By default, 3	If a user attempts to log in with an incorrect password more than the number of times specified here, the user is locked out of the system for the duration specified by the <code>authentication_lock_duration</code> parameter.

Table 14. Authentication Parameters

Parameter	Value	Description
authentication_lock_duration	By default, 300000	The duration for which a user attempting to log in to the system with an incorrect password is locked out of the system.
default_authentication_mode	<ul style="list-style-type: none"> ■ PDS ■ <integration profile name> 	Determines the default authentication method. NOTE: The default method can be overridden by appending the desired method to the URL as an auth querystring parameter.
deposit_self_registration	true/false	When false, the patron self-registration is not displayed in the local authentication login form.
sru_authentication	By default, false	Indicates whether SRU requests require authentication. If true, Rosetta authenticates a user based on a Basic Authentication Authorization header.

Repository

This module contains parameters that define Permanent Repository settings, such as the file storage system.

Table 15. Repository Parameters

Parameter	Value	Description
cach_mode		
run_virus_check_on_bitstream	True	Determines whether a virus check will be performed at the bitstream level
system_date_format	dd/MM/yyyy	Date format used for displaying dates in the Rosetta system user interface
http_port	1801	The port that Rosetta is listening on for http requests

Table 15. Repository Parameters

Parameter	Value	Description
tomcat_port	58821	The port that the Tomcat server is listening on
dc_date_format	dd/MM/yyyy	Date format used for displaying dates in the <code>dc:date</code> fields
keep_jhove	False	Defines whether the jhove output is written into the METS or only some of the fields are mapped to the DNX format
build_missing_indexes	True	Build missing object indexes on server restart
urn_fixed_length	5	Length for the URN number generated by the DOI plug-in
metadata_extraction_time_out_seconds	0	Number of seconds before metadata extraction times out

SMTP Email

Table 16. smtp_email Parameters

Parameter	Value	Description
use_email_smtp	false	
email_smtp_host	smtp.gmail.com	
email_smtp_port	587	
email_smtp_from	Management Department <rosetta.exlibris@gmail.com>	
email_smtp_pass	*****	

Search

This module contains parameters that define Search settings.

Table 17. Search

Parameter	Value	Description
index_supported_date_formats	yyyy-MM-dd HH:mm:ss,yyyy-MM-dd,dd/MM/yyyy,dd-MMM-yyyy,yyyyMMdd HHmm,'c'yyyy'.', 'c'yyyy'-'', 'p'yyyy'.', yyyy'.', yyyy']	The date format supported by the SOLR indexing
idx_load_balancer		Load balancer definition for the SOLR servers
search_max_number_of_columns	100	Number of columns displayed in the search UI
solr_server_capacity	30000000	The capacity of the SOLR server (number of files)
solr_commit_within	2000	The time interval between each commit of the SOLR indexing (in milliseconds)
search_results_splitter	10000	The bulk size of the export to Excel functionality
search_top_facets	5	max number of facets to display
max_facets_criteria	4	max allowed number of facet criteria
solr_commit_within	2000 ms	Controls the commit time of indexing. It can be changed before large re-index jobs.

3

Updating the Rosetta System

This section includes:

- **Hot Fix Installation** on page 76
- **Service Pack Installation** on page 76
- **Hot Fix Installation** on page 76
- **Copying Rosetta System Configurations** on page 78

Verifying your Rosetta Server Version

Basic information about the version and role(s) of Rosetta running on your server can be retrieved from the shell using the `ver` command:

```
il-dps01-d4(1) >>ver

Hostname : il-dps01
Rosetta Server Status : UP
Application role : DEP,REP,DEL,PER,IDX

Rosetta Ver  Server Name      Server Role  Build Date  Build Number  Rank
-----
4.0.0       il-dps01.corp.ex  DEP,REP,DEL,  SATURDAY   33045 (115)   1
              librisgroup.com  PER,IDX      NOVEMBER 23,
              2013

pds version : 2.0.1.00001 , (ACTIVE)
```

For extended information, run `ver_full`.

Service Pack Installation

For information on installing service packs, see the *Rosetta Service Pack Installation Guide* in the Installation and Implementation Information directory of the Rosetta customer portal.

Hot Fix Installation

This section explains how to install a hot fix for Rosetta.

Rosetta hot fixes are cumulative rather than incremental. For example, 3.1.2.2 includes 3.1.2.1. Please note that rolling back two or more consecutive hot fixes cannot be done. If there is a need to reverse changes applied by two or more consecutive hot fixes, please contact Rosetta Support.

NOTE:

Information about post-installation sanity tests is not included here.

The following terms are referred to in this segment:

- Distributed topology: two or more “all-in-on” application servers
- Distributed roles topology: two or more applications servers, each having a different role (for example, deposit, staging repository, or permanent repository)

Installing a Rosetta Hot Fix

Before installing a Rosetta hot fix, back up your Rosetta database using your tools/policies.

To install a Rosetta hot fix:

- 1 Create a folder to which the `.tar.gz` files will be downloaded:

```
cd $HOME  
mkdir hf_for_v<Rosetta_version> [for example, hf_for_v2.2.1]
```

In a distributed topology, repeat this step for each of the servers.

2 Obtain the .tar.gz file from the Ex Libris FTP server:

```
cd hf_for_v<Rosetta_version>
ftp -i ftp.exlibris.co.il [user name= dps_hf; password=4hfUpLoad]

cd hf_for_v<hot-fix-version> [for example, hf_for_v3.1]
bin
get Rosetta_<hot-fix version>.tar.gz [for example,
Rosetta_3.1_0_1.tar.gz]
quit
tar -zxvf <hot-fix-number>.tar.gz [for example, tar -zxvf
Rosetta_3.1_0_1.tar.gz]
```

In a distributed topology, copy or FTP and untar the .tar.gz files to each of the servers. -

NOTE:

The latest hot fix version in the version directory always includes the content of the previous hot fixes. Install only the higher version. (For example, if 2.tar.gz exists, do not download and install the file 1.tar.gz.)

3 Run the hot fix installation script:

```
csh -f install_hf.csh
```

In a distributed roles topology, run the script on each of the servers. Note that the script can be run consecutively or concurrently on all servers.

NOTE:

While the hot fix is being installed, a log file is written. The full log file can be found at /exlibris/dps/d4_1/local_sp_install/5.3.0.1/log/install_sp.log.

Rolling Back the Hot Fix Installation

If you must roll back to the previous version, use the following commands:

```
cd hf_for_v<Rosetta_version>
csh -f rollback_hf.csh
```

In a distributed topology, repeat these commands for each of the servers .

Restarting Rosetta

Once you have successfully installed the hot fix, restart Rosetta using the dps_start command. In a distributed topology, run this command on each of the servers.

Copying Rosetta System Configurations

The Copy Configuration feature reduces manual work when configuring the same settings in a newly installed environment as in an existing one.

Not all configuration items can be automatically copied. For example, specific environment variables such as IP addresses or NFS paths cannot be copied without causing errors.

Copy configuration is performed at the consortial level. All items from all institutions including the consortium hierarchy are copied.

The following sections include a description of the items that can be copied as well as instructions for performing the copy and transfer of configuration data.

Items Copied in Copy Configuration

- Repository
 - Task Chains
 - SIP Processing Configuration
 - Storage Definitions
 - Plug-In Instances
- Users And Deposit
 - Users
 - Producers
 - Producer Profiles
 - Metadata Profiles
 - CSV Templates
 - Content Structures
 - Submission Formats
 - Metadata Forms
 - Material Flows
- Administrative Structure
 - Administrative Structure
 - UI Customization
- Delivery
 - Viewers
 - Representation Profiles

- Delivery Rules
- Generic Thumbnails
- Metadata
 - Shared Metadata
 - History
- General
 - General Parameters
 - Authentication Profiles
 - Transformation Profiles
 - Configuration Files
 - Quartz Jobs
- Code and Mapping Tables
 - Code and Mapping Tables
- Rules
 - Rules

Implementing the Copy Configuration

The Copy Configuration feature is available by clicking **General** in the Administration module (consortium level). Two options are available from this page, Export Configuration and Import Configuration. Typically, a workflow begins with exporting a configuration, then importing it to a new environment.

To export a configuration:

- 1 From the Administration module > General page, click **Export Configuration**. The Export Configuration page opens (see [Figure 28](#) below).

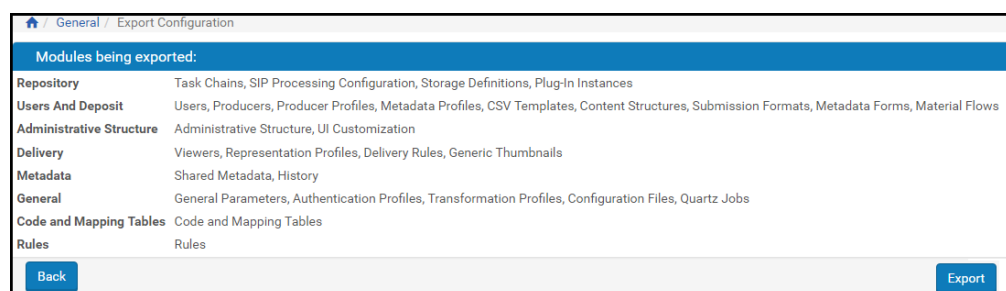


Figure 28: Export Configuration Page

2 Click the **Export** button to continue.

A progress bar is displayed as the export runs. The system creates a file containing the entire configuration. This step may take several minutes, depending on the size of the configuration. The file name containing the exported configuration is `<YYYY-MM-DD hh-mm>.zip` and is saved locally on the server under `.../operational_shared/cc_export_dir/`.

3 If necessary, copy the `.zip` file to a PC that can access the importing environment.

IMPORTANT:

Do not open or edit the `.zip` file. Any changes to the file will hinder your success when importing.

To import a configuration:

1 Confirm that:

- the importing system and the exporting system have the same Rosetta version and the same format library. To check which Rosetta version is installed, click the **Help** link in the upper right corner of the Web page and then **About Rosetta**. To check the format version, select **Preservation > Manage Format Library Version**.
- the importing environment is clean, without any data.
- all NFS storage locations are available and empty.

2 Log in to the administration module and follow the path **Advanced Configuration > General**.

3 Click the **Import Configuration** text. The Import Configuration Page opens (see [Figure 29](#) below).

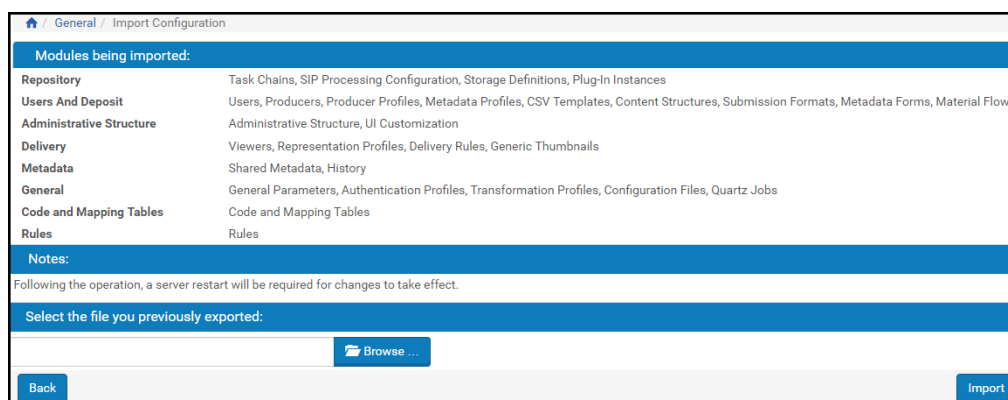


Figure 29: Import Configuration Page

- 4 Click the **Browse** button and navigate to the exported configuration .zip file you want to import. Select the file and click **Open**.

The file path and name populate the **Select the file you previously exported** field.

- 5 Click the **Import** button.

The system imports the configuration. This may take a few minutes. When finished, a confirmation message appears.

- 6 Restart Rosetta and perform a sanity test to verify a successful import.

NOTES:

The following activities must be performed manually:

- Copy the plug-ins jars (before the restart).
 - Align NFS path definitions (for example, operational shared paths, submission formats, and quartz jobs parameters).
 - Fix S3 storage definitions.
 - Add (S)FTP passwords.
 - Add logos to UI customizations.
-

4

Working with the Rosetta System

This section includes:

- [Starting and Stopping the Rosetta Application Servers](#) on page 83
- [Verifying the Operation of the Rosetta System](#) on page 85
- [Logging On to the Rosetta System](#) on page 85
- [System Checks](#) on page 86
- [Description of Aliases](#) on page 87
- [PDS Directory Tree](#) on page 88
- [Rosetta System Log Files](#) on page 88
- [System, Background, and Operational Jobs](#) on page 89
- [Shared Directories](#) on page 96

Starting and Stopping the Rosetta Application Servers

This section explains how to start and stop the various Rosetta system servers.

To start/stop the Patron Directory Service (PDS):

- 1 Log on to the PDS server.
- 2 Run the following commands as user dps:

```
apcb
apachectl stop
apachectl start
```

When running PDS on a port that requires root permissions (such as 80 or 443), the following command should be run (as root) to start/stop the service:

```
/exlibris/dps/d4_1/product/local/apache/bin/apachectl_auto  
/exlibris/dps/d4_1/product/local/apache/bin/apachectl stop
```

To start/stop an application server:

- 1 Log on to the server machine.
- 2 Run the following commands:

```
dps_start  
dps_stop
```

If there is more than one application server—either separate servers for Deposit, Staging, and Delivery, or in the case of a clustered environment (array of application servers managed by a load balancer)—these commands should be performed for each of the servers separately in the following order:

To shut down:

- 1 Deposit
- 2 Delivery
- 3 Staging
- 4 Permanent

To start up:

- 1 Permanent
- 2 Staging
- 3 Delivery
- 4 Deposit

NOTE:

If Rosetta is installed on a distributed environment (where each server can perform all roles—Deposit, Delivery, Staging, and so forth), there is no need to adhere to any shutdown/startup sequence.

To activate safe mode on the server:

NOTE:

While in safe mode no background jobs or processes (such as quartz jobs or queues) are running.

```
dps_start -s
```

or

```
dps_start --safe
```

To deactivate safe mode:

```
dps_start
```

Verifying the Operation of the Rosetta System

After you have installed and configured your Rosetta system, verify that the system is working properly.

To verify that the system is working correctly:

- 1 Log on to the Deposit Server at the following URL:
<http://<Deposit Server machine>:1801/deposit>
- 2 Log on to the Staging Server at the following URL:
<http://<Staging Server machine>:1801/mng/wrk/>
- 3 Using the following URL, verify that the Permanent Server is up and running:
<http://<Permanent Server machine>:1801/permanent/>

Logging On to the Rosetta System

To enable Administrators to log on to the Rosetta system, one user with all administrative privileges is installed with the software. Administrators can use the following information to log on to the system:

- User name: **admin1**
- Password: For the default password, contact the Rosetta implementation/support teams.
- Institution code: **INS01**

After logging on, Administrators can begin initial configuration, which includes

- changing their user name and password and
- adding new users.

IMPORTANT:

You must change the password after the initial logon and setup.

System Checks

System checks are used to check essential components needed for running Rosetta properly.

Table 18. System Checks

Name	Description (what is checked)	Action on warning/initial failure	Action on persisting failure
DBChecker	Application access to Oracle database	Rerun	Confirm Oracle is available
PDSCheckPlugin	Apache is running	Rerun	Restart Apache
PluginsCheckerPlugin	Plug-ins in plug-in directory are deployed/executable	Rerun	Check the plug-in directory is available and plug-ins have proper permission
SharedFoldersAccessChecker	Operational_shared is accessible (read and write)	Rerun	Fix permissions on operational_shared
SharedFoldersSpaceChecker	Operational_shared has at least 500MB free space	Rerun	Add space to operational_shared
StorageAccessChecker	All configured storage is accessible	Rerun	Mount storage to application server(s)
DepositWSChecker	Availability of web services	Rerun	Restart relevant server(s)
RepositoryWSChecker			
PermanentWSChecker			
DeliveryWSChecker			

Table 18. System Checks

Name	Description (what is checked)	Action on warning/initial failure	Action on persisting failure
Storage Space Checker	Checks that the free space of all storage is not below the configured threshold.	Rerun	Add storage to the specified storage and configure storage definitions accordingly or free up space by deleting objects using Rosetta UI/APIs.

Please contact Rosetta Support if you have difficulty resolving a failed system check.

Description of Aliases

The following table provides a list of aliases that are used by the system and in this guide.

Table 19. Aliases

Alias	Description
dps_bin	Shortcut to the Rosetta system's bin directory
dps_conf	Shortcut to the Rosetta system's configuration directory
dps_deploy	Shortcut to the Tomcat server's web applications deploy directory
dps_log	Shortcut to the Rosetta server's log directory
dps_util	Shortcut to the Oracle utilities
dps_start	Starts Rosetta
dps_stop	Stops Rosetta
pdsroot	Shortcut to the PDS directory
s+ [schema name]	Opens an SQL Plus session to the specified schema in Oracle
ver	Display basic information about the Rosetta environment
ver_full	Display extended information about the Rosetta environment

PDS Directory Tree

Administrators can access the root directory of the Patron Directory Service (PDS) by entering the following command:

```
>>pdsroot
```

In this directory, the following PDS subdirectories can be found:

- **conf_table** – Contains the main PDS `tab_service.institute` configuration files. Most of the PDS configurations are performed in this directory. These include LDAP configurations, attribute mapping file customizations, and so forth.
- **program** – Contains the PDS program files and modules
- **html_form** – Contains all of the HTML files used by the PDS, as well as customized versions of the various PDS screens
- **service_proc** – Contains internal PDS Perl routines used by the PDS for communication with various information sources
- **pds_files** – Stores all active session cookies
- **pds_proc** – Contains internal PDS utilities

Rosetta System Log Files

The Rosetta system automatically records process and event information in log files. All log files are stored in the `log` directory.

Table 20 describes the contents and locations of the log files in which the Rosetta system records this information:

Table 20. Rosetta System Log Files

Log File	Contains	Location
<code>server.log</code>	<ul style="list-style-type: none">■ Actions performed by users and recorded by the Rosetta system in real time■ Events initiated by the Rosetta system■ Execution and progress of internal system tasks■ Warning and error messages	<code>dps_log</code>

Table 20. Rosetta System Log Files

Log File	Contains	Location
pds_server.log	Information about internal interactions between the Rosetta system and Patron Directory Service (PDS)	/exlibris/dps/d4_1/log/

Rosetta server logs are automatically rotated and compressed. The following table describes Rosetta server.log rotation and compression naming convention policy:

Table 21. Rosetta server.log Naming Convention Policy

Event	Naming Convention	Comments
Midnight	server.log.[datestamp].gz	
Server restart	server.log.[datestamp].[timestamp].gz	
Log file reaches 1 GB	server.log.1.gz, server.log.2.gz, . . . server.log.[n].gz;	At midnight a datestamp is added (server.log.n.[datestamp].gz)

System, Background, and Operational Jobs

Rosetta runs the following types of jobs in the background:

- **System** – Jobs that are used for system maintenance and are not associated with a specific institution or workflow.
Only System Administrators can monitor and manage jobs of this type.
- **Background** – Jobs of this type are configured to run frequently in order to maintain the smooth operation of Rosetta.
Only System Administrators can monitor and re-execute jobs of this type, but they cannot modify the job's schedule.
- **Operational** – Jobs that are associated with a specific institution or workflow and are managed by staff users in the Rosetta Management interface (such as submission jobs, the Risk Analysis Report job, and the Producer Report job).

System Administrators can only monitor jobs of this type.

The following table lists all the jobs that System Administrators can monitor with the Rosetta Administration interface. For more information, see [Monitoring All Jobs](#) on page 92.

Table 22. Rosetta Jobs

Job Name	Description	Configuration
Submissions jobs	Bulk submission jobs that are created and managed by Preservation Managers and Analysts in the Rosetta Management interface. Type: Operational	
Risk Report	Generates the Risk Analysis report at an institutional level. Type: Operational	
Delete Deposits	Deletes deposit directories (parent defined under general parameter <code>logic_deposit_area</code>) for deposits that were declined or deleted. Type: Background Frequency: Every Sunday at 1:00	
Producer Report	Generates and emails the report of producer deposits at an institutional level. Type: Operational	
DepIndex Optimize	Indexes the deposit tables so that they can be searched by the system and users. Type: Background Frequency: Every day at 2:00	
Analyze Retention Policy Deleted IEs Process	Evaluates IEs with retention policies to determine if they should be deleted. This job generates a table of IEs that is used for a report called IEs to be Deleted. Type: System Frequency: Every day at 2:00	
Delete IE By Retention Policy	Deletes IEs whose retention periods have expired. Type: System Frequency: Every day at 2:00	
Index Optimize	IDX00 Oracle text index optimize Type: System Frequency: Every day at 2:00	

Table 22. Rosetta Jobs

Job Name	Description	Configuration
Metadata Loader	Aleph/Voyager updates Type: System Frequency: Every two hours	General Parameter: md_load_interval_in_s econd
Metadata Orphan Handler	Deletes unused CMS MD records. Type: System Frequency: Every day at 2:00	General Parameter: md_orphan_handler_in terval_in_second
Publishing Sync Job	Finds records to be published and publishes them. Type: System Frequency: Every day at 00:30	General Parameter: comparison_start_time comparison_interval
Startup Checks Job	Startup checks are used to check essential components needed for running Rosetta properly. Type: System	
Statistics Analyzer	Gather event statistics Type: System Frequency: Every 15 minutes	General Parameter: statistic_interval
Index Synch	Makes sure that each new object (IE) is indexed so that it can be searched in the system. Type: System Frequency: Every one minute	
Plan Execution Monitor	Preservation Processing Type: System Frequency: Every five seconds	
Process Automation Monitor	Process Execution Type: System Frequency: Every five seconds	
Format Library Updates	Checks for (and installs) new available Format Library versions. Type: System Frequency: Every Sunday at 2:00	Install Automatically (false)

Table 22. Rosetta Jobs

Job Name	Description	Configuration
Clean Up	<p>Removes old data from the file system and the database.</p> <p>Type: System</p> <p>Frequency: Every Sunday at 1:00</p>	<p>Data older than x days. Can be configured separately for different areas.</p> <hr/> <p>NOTE: When selecting Clean Deleted OAI-PMH records, it is recommended that you set the OAI deletedRecord policy to transient in the oaiproviderconfig.xml configuration file.</p> <hr/> <p>Clean Old Events can be configured to include/exclude specific event types.</p>
Find Duplicates	<p>Finds duplicates across the repository.</p> <p>Type: System</p> <p>Frequency: Do not run</p>	<p>Email distribution</p> <p>Maximum number of duplicates to be reported can be configured.</p>

Monitoring All Jobs

The Manage Scheduled Jobs page enables System Administrators to monitor the status of all jobs. To access this page, click **Manage Scheduled Jobs** on the Home > Advanced Configuration page.

	Name	Type	Institution	Previous F	Next FireTime	Frequency	State			
1	DepIndex Optimize	Background	CRS00	-	30/11/2012 02:00:00	Every day at 2:00	Normal	View		
2	Plan Execution Monitor	Background	CRS00	29/11/2012 20:00:25	29/11/2012 20:00:30	Every five seconds	Normal	View		
3	Process Automation Monitor	Background	CRS00	29/11/2012 20:00:25	29/11/2012 20:00:30	Every five seconds	Normal	View		
4	Delete IE By System Retention Policy	System	CRS00	29/11/2012 18:38:44	30/11/2012 02:00:00	Every day at 2:00	Normal	View	Edit	Run Now
5	Producer Report	Operational	CRS00.INS	29/11/2012 12:19:24	-	-	Not Running	View		
6	Risk Report	Operational	CRS00.INS	29/11/2012 15:36:09	-	-	Not Running	View		

Figure 30: Manage Scheduled Jobs

Depending on the type of job, System Administrators can perform the following tasks from this page:

- **View a job's details** – Click the **View** link next to the job you want to view.
- **Schedule a job** – For more information, see [Scheduling a System Job](#) on page 93.
- **Execute a job** – Click the **Run Now** link to run a job manually.

In addition, System Administrators can cancel a system job. For more information, see [Canceling a System Job](#) on page 95.

Scheduling a System Job

This task allows System Administrators to schedule a system job.

To schedule a system job:

- 1 On the Manage Scheduled Jobs page, click the **Edit** link next to the job that you want to modify.

The Job Details page opens.



Figure 31: Job Details Page

- 2 Select the interval at which to execute the job: **Hourly**, **Daily**, **Weekly**, **Monthly**, or **Advanced**.
- 3 To configure hourly, daily, weekly, and monthly intervals:

- a Use the following table to configure the common interval fields:

Table 23. Common Interval Fields

Section	Field	Description
Start At	Time	Select the hour and minutes from the drop-down fields to specify the time at which to run the job.
	Date	Use the calendar tool or select the month, day, and year from the drop-down fields to select the date at which to start running the job.
Repeat	Indefinitely	Select this option to run the job indefinitely.
	Until	Select the month, day, and year from the drop-down fields to specify the date at which to stop running the job. NOTE: After this date, the state of the job will change from Normal to Not Running .

- b Use the following table to configure the interval-specific fields:

Table 24. Interval-Specific Fields

Type of Interval	Perform this task:	Description
Hourly	Every	Select the hourly interval from the Hours drop-down field.
Weekly	Week days	Select which days of the week to run this job.
Monthly	Monthly At	Select the day of the month to run this job.

For information on how to configure advanced intervals, see section Configuring Advanced Schedules in the *Rosetta Staff User's Guide*.

- 4 Click the **Apply** button to add the job to the list of submission jobs.

Canceling a System Job

Canceling a system job allows the System Administrator to postpone the execution of the job indefinitely without deleting it from the system.

NOTE:

The **Unschedule** option will not be available when the state of the job is **Not Running**.

To cancel a system job:

- 1 On the Manage Scheduled Jobs page, click the **Edit** link next to the job that you want to modify.
The Job Details page opens.
- 2 Click the **Unschedule** button.
The state of the job should change to **Not Running**.

Shared Directories

Rosetta needs access to several mount points and shared directories in order to operate. The mount points are defined in the Operational Repository and the Permanent Repository. They can be found in the Storage Rules and Definitions area under **Advanced Configuration > Repository**.

For the deposit area (where the deposited files are initially stored) and for the operational shared objects (such as plug-ins and temporary folders), the mount point is defined in the General Parameters table (found in **Advanced Configuration > General > General Parameters**).

Operational and Permanent Repository Definitions

Both the Operational Repository and Permanent Repository are defined in Storage Definitions. For new installations, a mount point is assigned for each repository, and the out-of-the-box values are replaced with the values provided during installation through RosiKit.

Deposit Area

The deposit area is defined by the `logic_deposit_area` general parameter in the Deposit module.

Operational Shared

There is a single operational shared directory which must be shared among all servers. It is defined by the `operational_shared` general parameter in the General module. Under that folder, the following directories are defined and used by the system.

Table 25. Operational Shared Directories

Name	Module	Location	Usage
operational_delivery_shared	Delivery	operational_shared/ operational_delivery_shared/	Hold Delivery file cache, Delivery UI skins

Table 25. Operational Shared Directories

Name	Module	Location	Usage
plugin_directory	General	operational_shared/ plugins	Root folder for plug-ins
plugin_deploy_jar_dir	General	operational_shared/ plugins/deploy	Holds plug-in .jar files
plugin_deploy_script_dir	General	operational_shared/ plugins/bin	Holds plug-in scripts
preservation_export_directory	Backoffice	operational_shared/ operational_export_directory	Holds import and export representations during preservation test/action
convert_temp_directory	Repository	operational_shared/ operational_delivery_shared/ convert_temp/	Used by converter script (thumbnail, jpeg2000)
sip_temp_directory	Repository	operational_shared/ sipTmpDir/	Used by converter script (thumbnail, jpeg2000)
bytestream_working_dir	Repository	operational_shared/ bytestream_work	Used for creation of bytestreams
NamedFolders (Mapping Table)	Delivery	operational_shared/ operational_delivery_shared/work/ jmagick/	JMajick work directory; Only required when Operational and Delivery servers are separate
digital_certificate	General	operational_shared/ DigitalPreservationSystem.pfx	Location for email-signing certificate
staff_work_area_directory	Backoffice	operational_shared/ staff_work_area/ [user-name]	Location used by staff users for export/import
preserve_pp_docs	Backoffice	operational_shared/ preserve_pp_docs	Location that signed-off preservation plans are exported to. A submission job can be configured to ingest exports from this location.

Part IV

Managing the Rosetta Database

This part contains the following:

- **Chapter 6: Working with the Rosetta Database** on page 101

6

Working with the Rosetta Database

This section contains:

- **Connecting to the Database from UNIX** on page 102
- **Connecting to the Database from UNIX** on page 102
- **Changing the Database Schema's Password** on page 102
- **Starting and Stopping the Oracle Server** on page 102
- **Starting and Stopping the Oracle Listener** on page 103
- **Viewing and Managing Tablespaces** on page 106
- **Viewing and Managing Oracle Data File (DBF) Information** on page 108
- **Archiving Data** on page 110
- **Viewing and Managing Oracle Database Memory (SGA)** on page 112
- **Viewing Alert Log (Oracle Database Log) File Content** on page 114
- **Viewing and Managing Information on Oracle Database Users** on page 116
- **Creating Database Users with Privileges** on page 118

NOTES:

- For detailed information on the Oracle EM, refer to the following Oracle documentation:
http://download.oracle.com/docs/cd/B19306_01/server.102/b14196/em_manage003.htm
 - For detailed information on Oracle Administration, refer to the following Oracle documentation:
http://download.oracle.com/docs/cd/B19306_01/server.102/b14231/toc.htm
-

Connecting to the Database from UNIX

Administrators can connect to the Rosetta database installed on a UNIX machine.

To connect to the database:

- 1 Log on to the UNIX machine on which the Rosetta database is installed.
- 2 In the command line, type `s+` followed by the name of the schema. For example: `s+ shr00`

NOTE:

`s+` is an alias for `sqlplus ${ORA_USER_PREFIX}! :1/`get_ora_passwd ${ORA_USER_PREFIX}! :1``. For more information on aliases, see [Description of Aliases](#) on page 87.

Changing the Database Schema's Password

To change an existing schema's password:

- 1 Change the database schema's password on the Oracle side by logging in to Oracle with the user `exl_dba`.
- 2 For all application servers, perform the following:
 - Run the `put_ora_passwd` script: `put_ora_passwd <username> <password>`.
 - Stop the server using the `dps_stop` command.
 - Run the `$dps_dev/system.dir/bin/set_globals.sh`.
 - Restart the server using the `dps_start` command.

Starting and Stopping the Oracle Server

For Rosetta to interact with the Oracle database, the Oracle server must be running. During installation, you can configure the Oracle server to start automatically at boot time. Otherwise, you can start the Oracle server using the Oracle Enterprise Manager DB Controller (EM).

NOTES:

- Starting/stopping the Oracle server requires a database administrator user name and password.
- In most cases, you must run the Oracle Listener in order to use Rosetta in conjunction with the Oracle database. For details, see **Starting and Stopping the Oracle Listener** on page 103.

To start or stop the Oracle server:

- 1 Log on to the EM using the user **DPS_DBA**. This user is created during installation and has database administrator privileges to start or stop the Oracle database.
- 2 In the **Home** tab, click **Startup/Shutdown**.

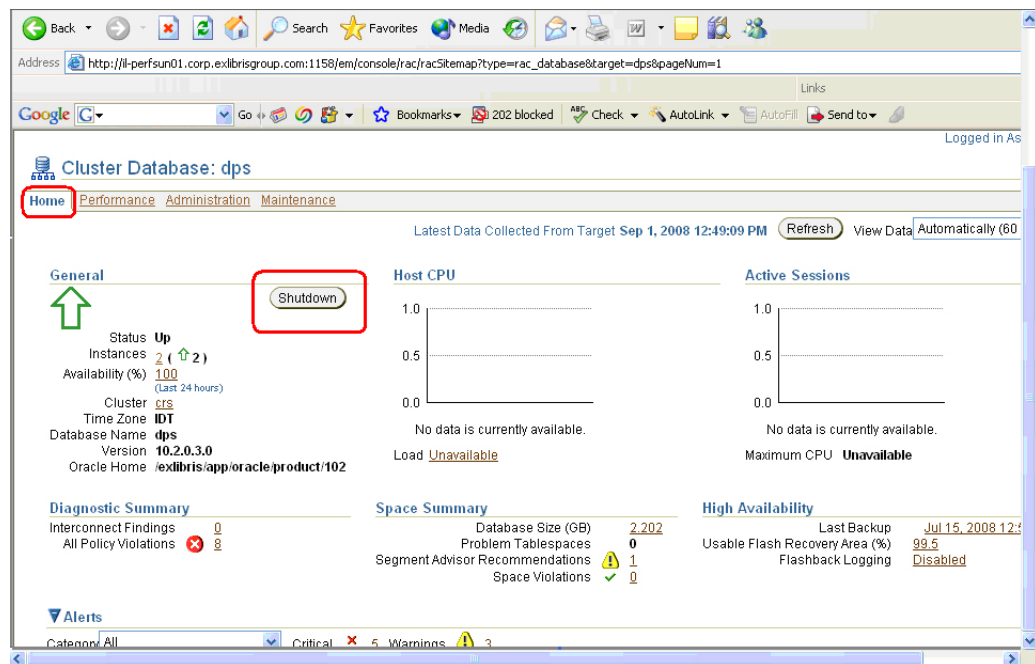


Figure 32: Oracle Server Startup/Shutdown

Starting and Stopping the Oracle Listener

When a user process initiates a connection request using a connect string, the Oracle Listener process examines the request and connects the user process to a server process.

If the Oracle database and Rosetta are installed on the same server, no third-party products are used to connect to the database, and no connections are made from a remote server, Rosetta can work without the Listener. Otherwise, the Listener must be running in order to work with Rosetta.

You can start and stop the Listener using the Oracle Enterprise Manager DB Controller (EM).

To start or stop the Oracle Listener:

- 1 Log on to the EM using the user **DPS_DBA**.
- 2 Select the **Administration** tab.

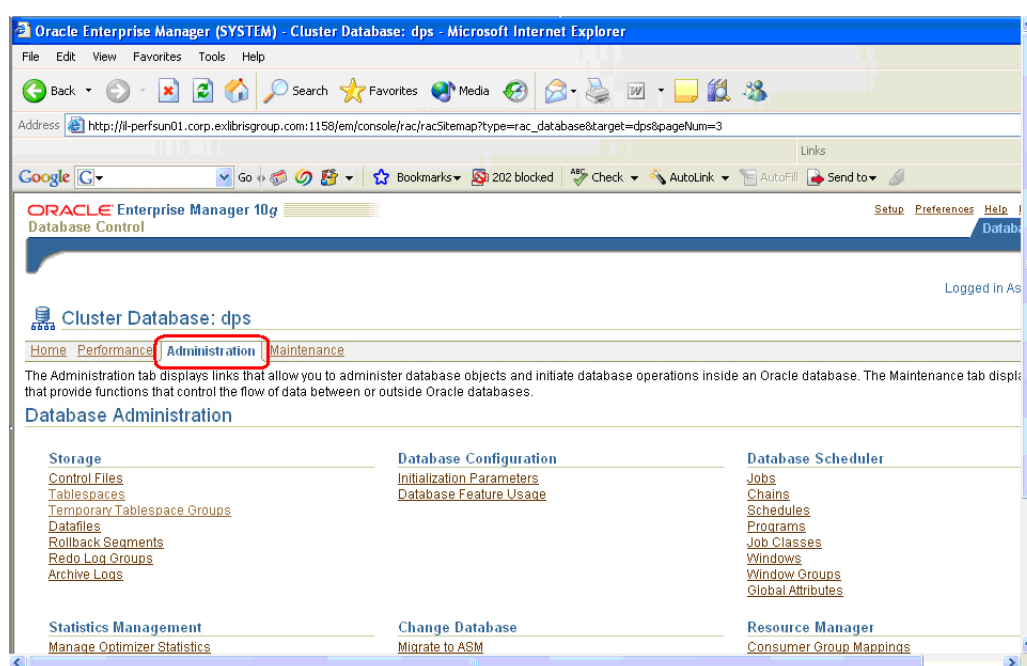


Figure 33: Administration Tab

- 3 Navigate to the required node (server).

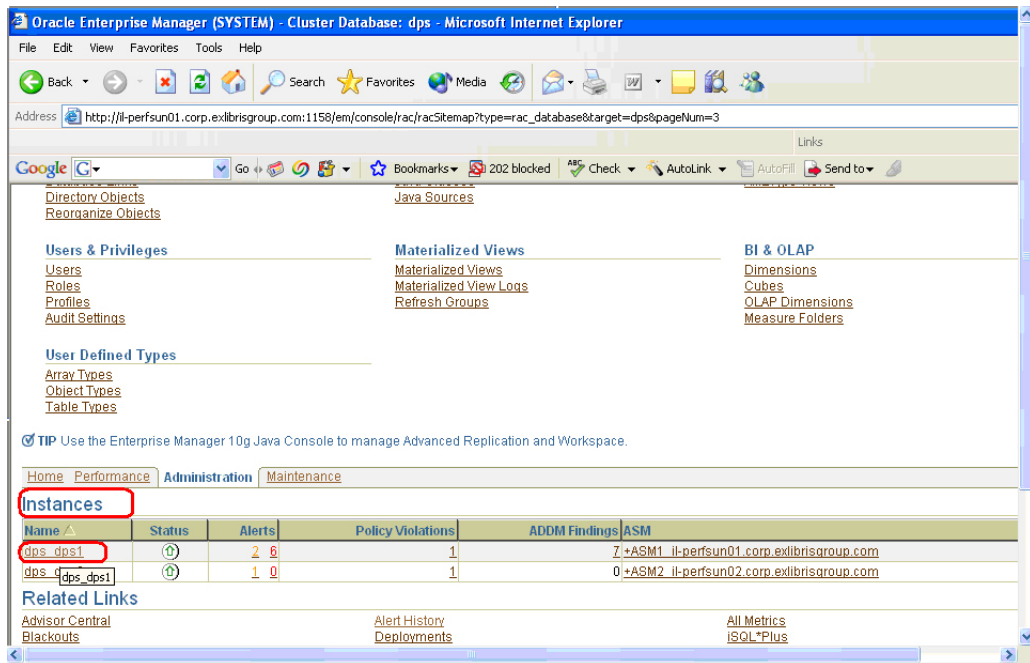


Figure 34: Node Navigation

- 4 Click the node to view the Listener details.

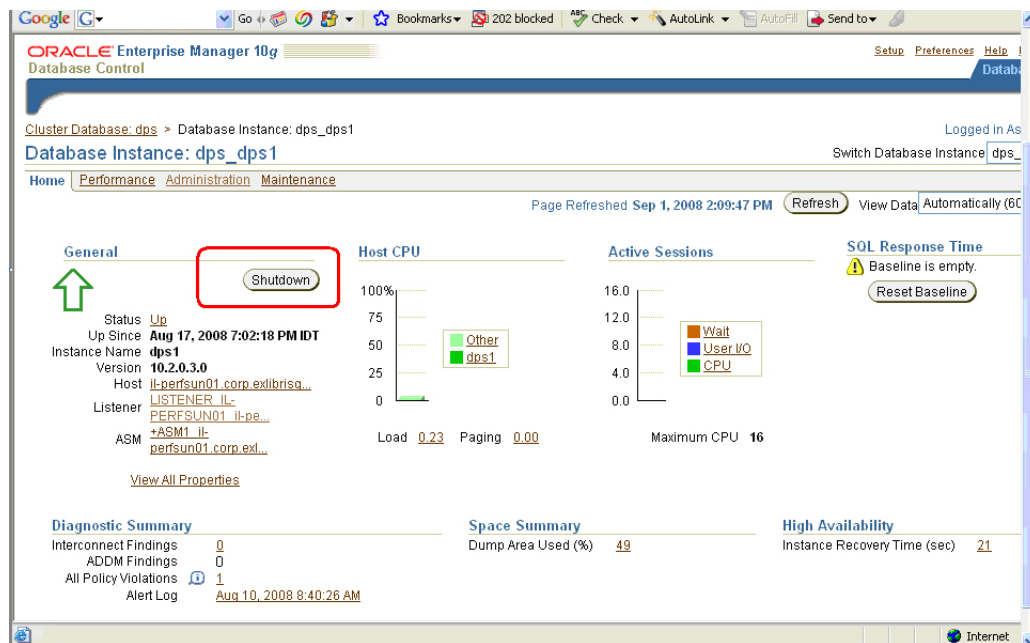


Figure 35: Listener Node Details

- 5 Click **Startup/Shutdown** to start or stop the Oracle Listener.

Viewing and Managing Tablespaces

You can view and manage Oracle tablespaces using the Oracle Enterprise Manager DB Controller (EM). For example, you can view the total space that each repository occupies in each tablespace and the amount of free space in each table.

To view and manage tablespaces:

- 1 Log on to the EM using the user **DPS_DBA**.
- 2 Select the **Administration** tab and click the **Tablespaces** link.

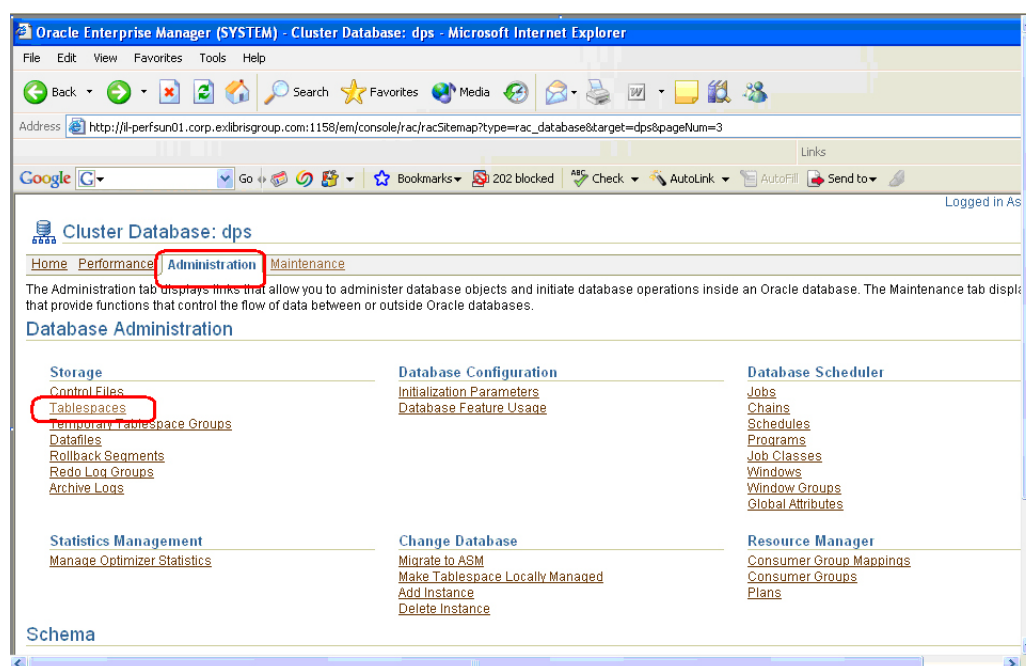


Figure 36: Administration Tab

The Tablespaces page is displayed, enabling you to view or manage each tablespace separately.

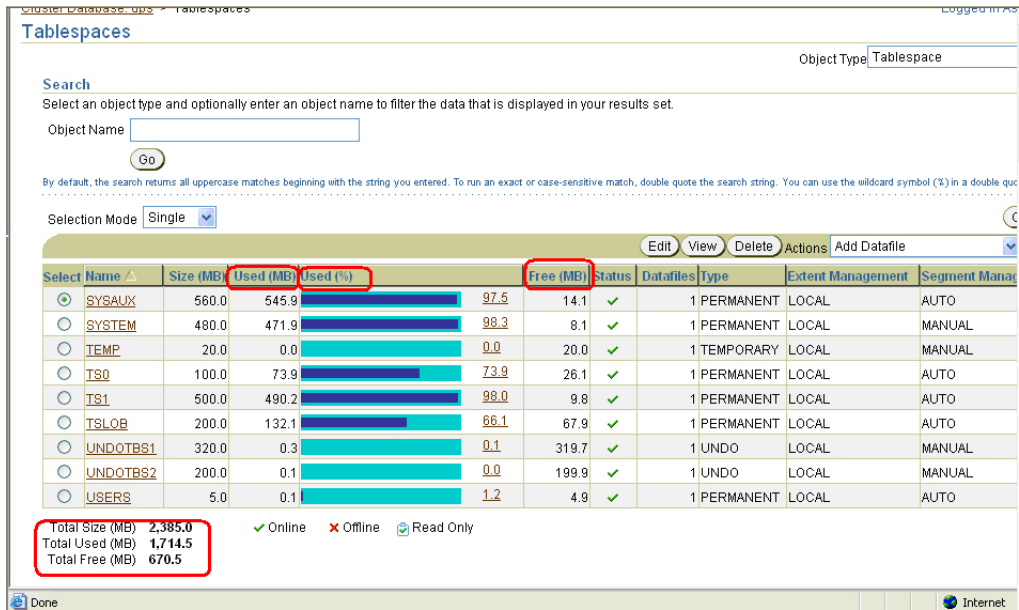


Figure 37: List of Tablespaces

- Click the required link to display the parameters of the tablespace you selected. For example, if you click the TS1 link, the following screen is displayed.

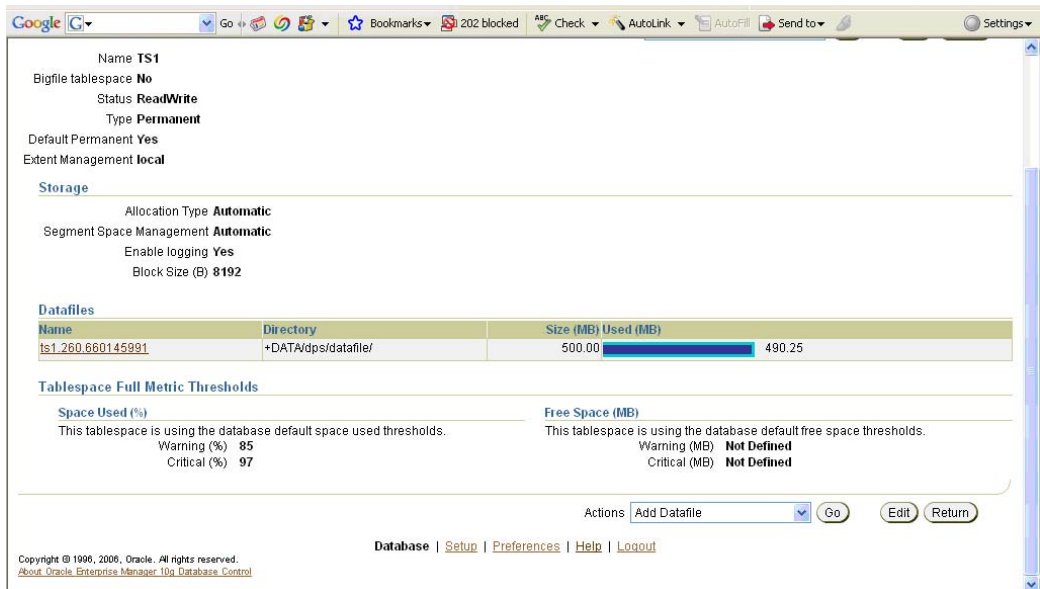


Figure 38: TS1 Tablespace Details

Viewing and Managing Oracle Data File (DBF) Information

You can view and manage Oracle data files using the Oracle Enterprise Manager DB Controller (EM).

To view and manage data files:

- 1 Log on to the EM using the user **DPS_DBA**.
- 2 Select the **Administration** tab and click the **Datafiles** link.

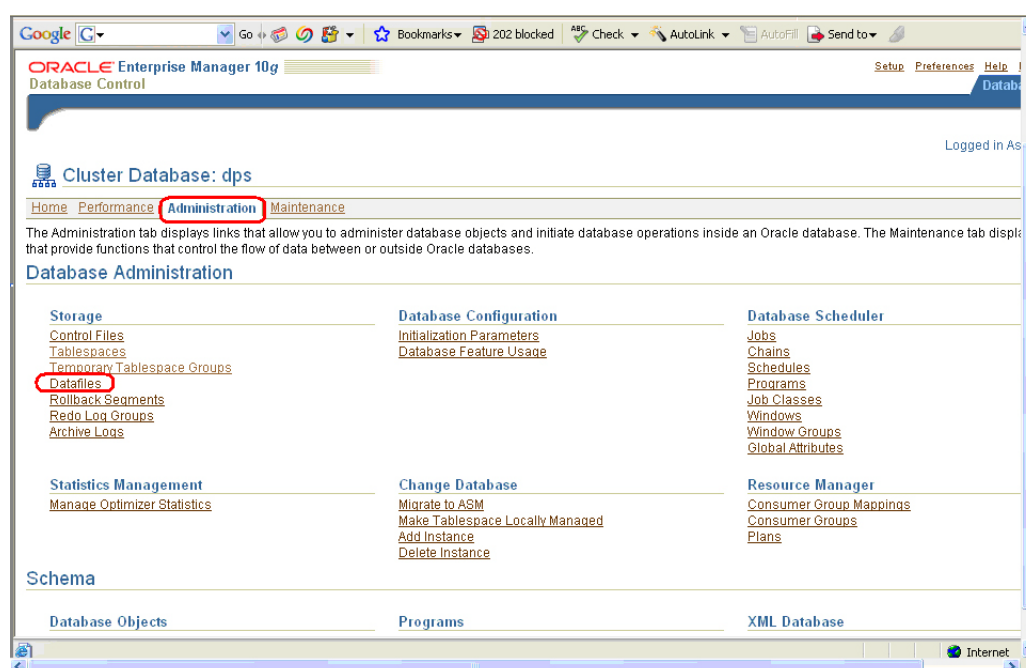


Figure 39: Administration Tab

The Datafiles page is displayed, enabling you to view or manage each file separately.

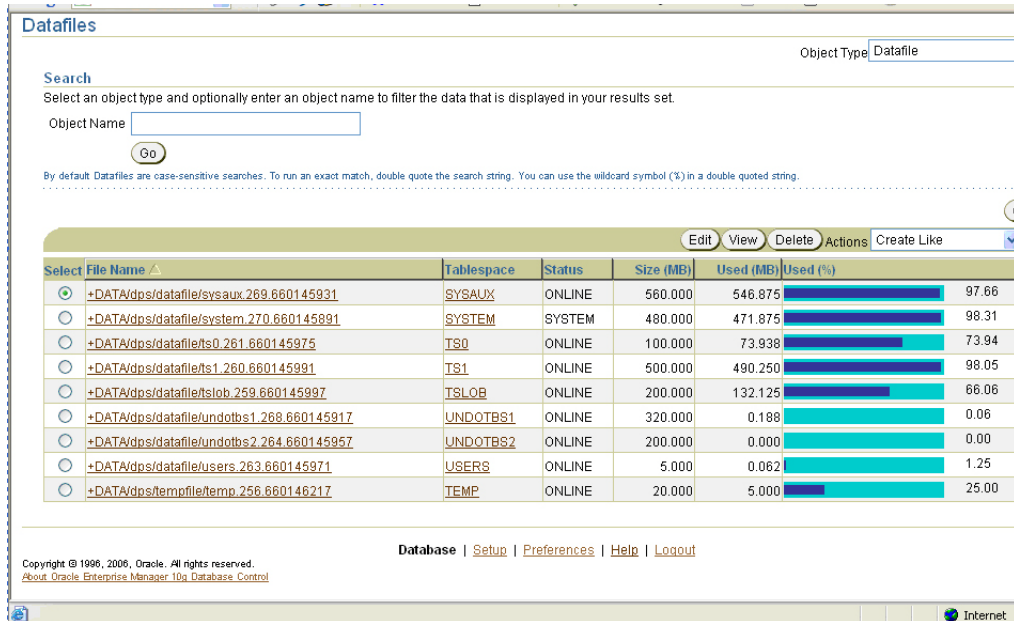


Figure 40: List of Data Files

- 3 Click the required link to display the parameters of the file you selected. For example, if you click the [+DATA/dps/datafile/ts1.260.660145991](#) link, the following screen is displayed.

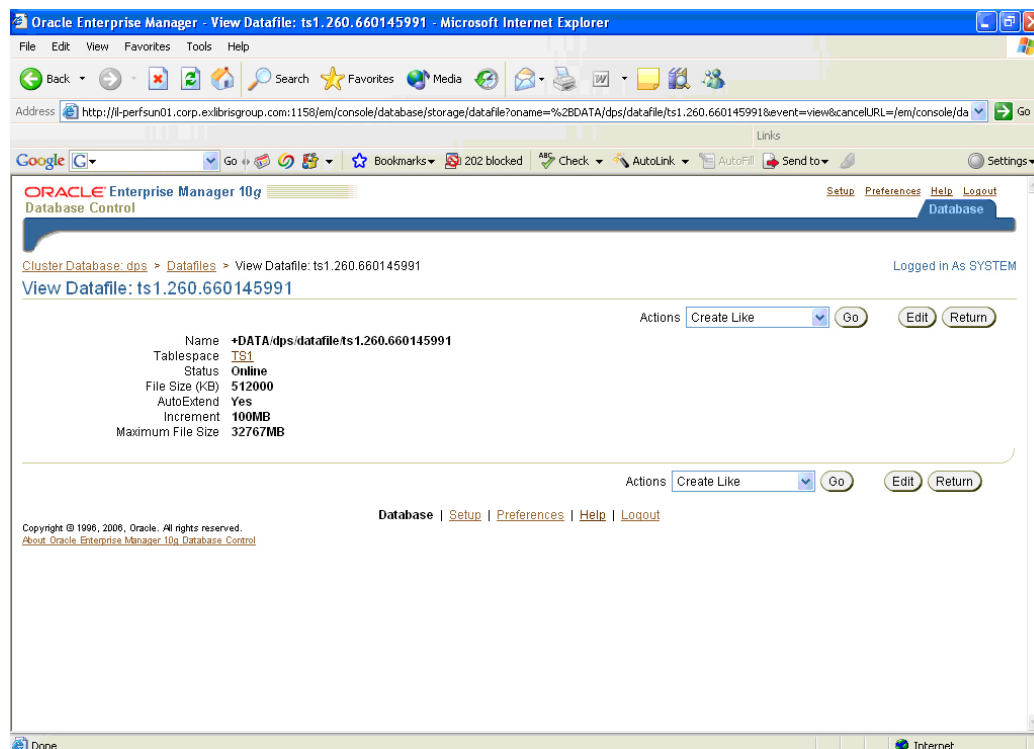


Figure 41: Individual Data File Example

Archiving Data

Rosetta backup and recovery procedures are based on Oracle. To provide complete recovery of data up to the time of failure, the Oracle database mode must be set to **archivelog** mode. Hot backup cannot be performed without first enabling the archivelog mode.

When the database is in **noarchivelog** mode, only cold backups can be performed. A cold backup restores data up to the time at which the last backup was performed. This means that all changes from the time of failure are lost.

Ex Libris highly recommends working in **archivelog** mode, which enables the use of both cold and hot backups to recover data up to the time of failure. To ensure a complete recovery, all the archive files that were generated from the time of the backup (hot or cold) until the time of failure must be available.

Three types of backup recovery can be configured from the Recovery Settings in the Oracle Enterprise Manager DB Controller (EM): Instance, Media, and Flash. Each category can be accessed by clicking the **Recovery Settings** link in the **Maintenance** tab.

Refer to the Oracle backup manual for more information on the preliminary actions that are required before using EM.

NOTES:

- To change the archiving mode, you must restart the Oracle database.
- You must log on as a SYS user (or DPS_DBA) to perform backup and recovery actions.
- For information on preliminary actions required before using the EM to configure backup recovery, refer to Oracle backup documentation.

To view and manage Instance, Media, and Flash backup and recovery:

- 1 Log on to the EM.
- 2 Select the **Maintenance** tab and click the **Recovery Settings** link.

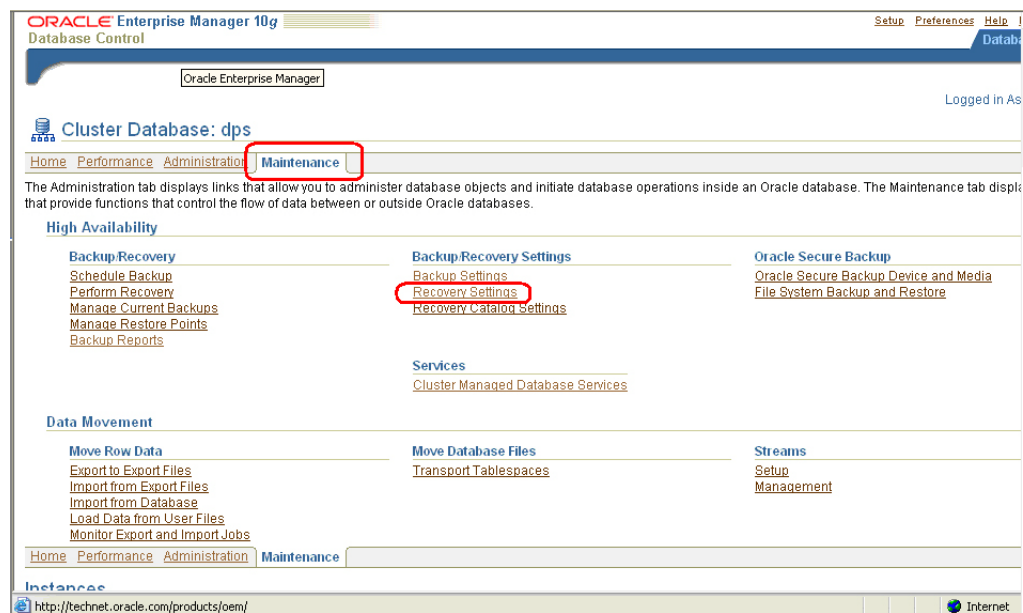


Figure 42: Maintenance Tab

The Recovery Settings page is divided into three sections.

- 3 Navigate to the section you require. For example, to manage Media recovery, navigate to the Media Recovery section.

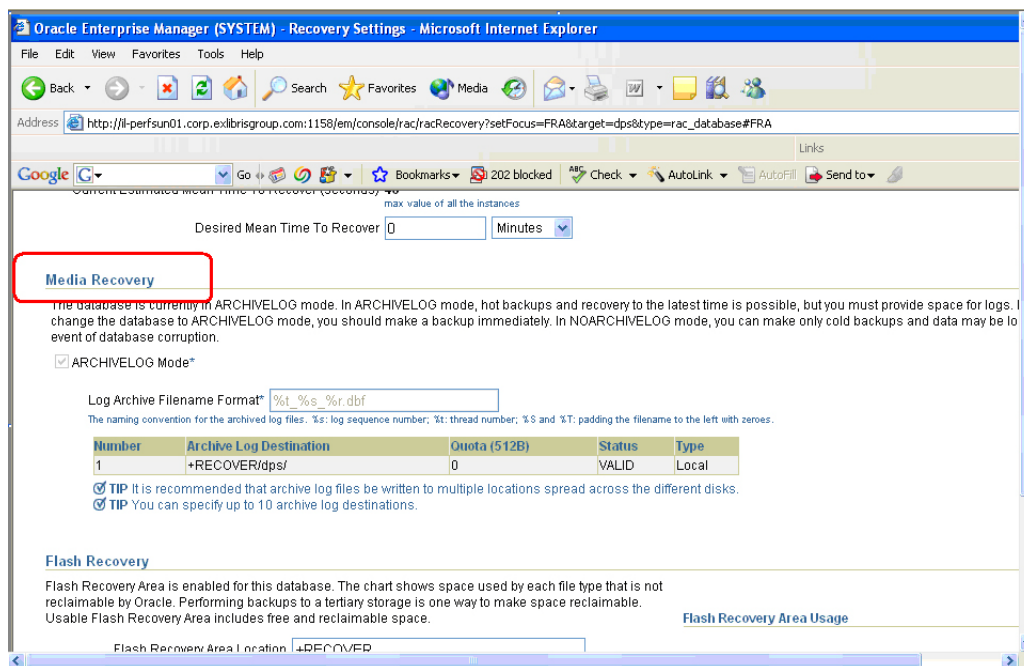


Figure 43: Media Recovery

Viewing and Managing Oracle Database Memory (SGA)

You can view and manage Oracle database memory (SGA) using the Oracle Enterprise Manager DB Controller (EM).

To view and manage memory (SGA):

- 1 Log on to the EM using the user **DPS_DBA**.
- 2 Select the **Administration** tab.
- 3 Navigate to the required node (server) and click it.

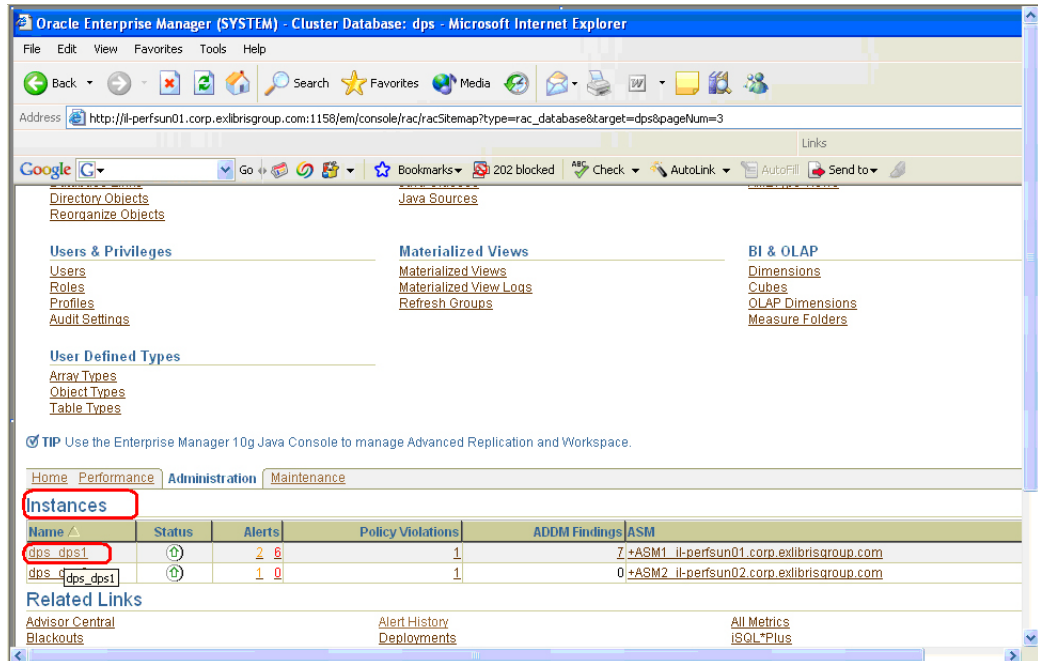


Figure 44: Node Selection

4 Click the **Memory Parameters** link.

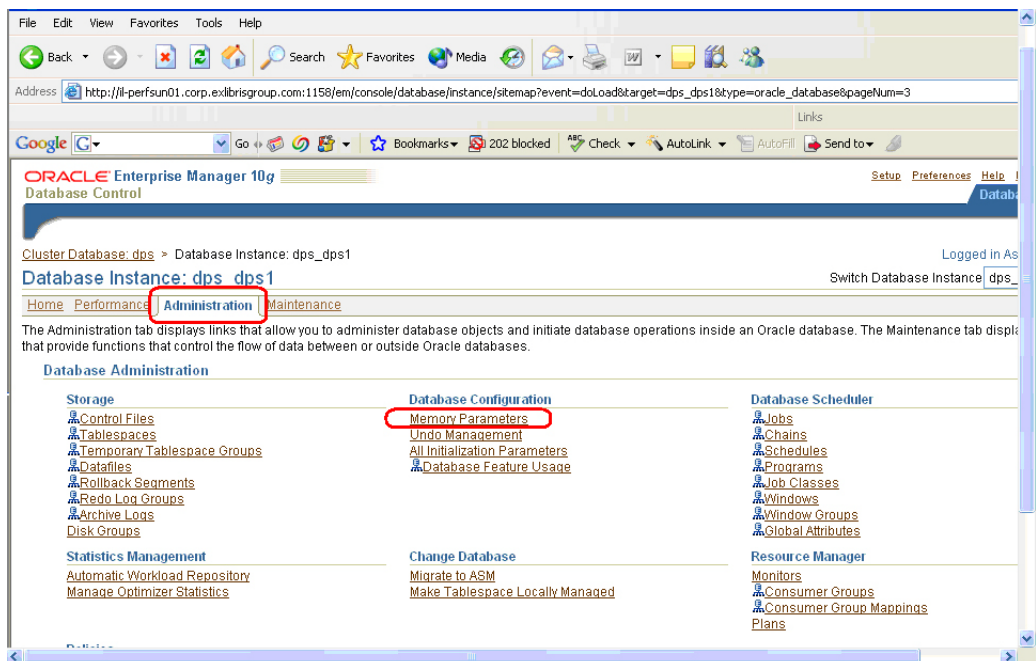


Figure 45: Memory Parameters Link

The selected database's memory parameters are displayed.

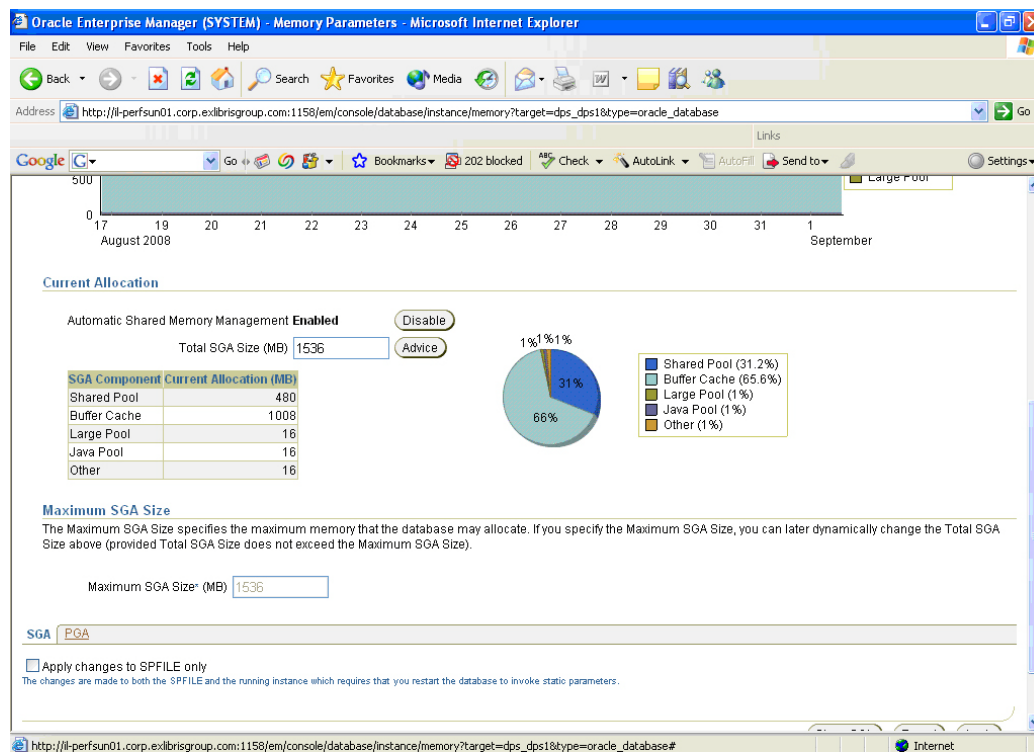


Figure 46: Memory Parameters

Viewing Alert Log (Oracle Database Log) File Content

You can view the content of the alert log (or Oracle database log) files using the Oracle Enterprise Manager DB Controller (EM).

To view alert log file content:

- 1 Log on to the EM using the user **DPS_DBA**.
- 2 Select the **Administration** tab.
- 3 Navigate to the required node (server) and click it.

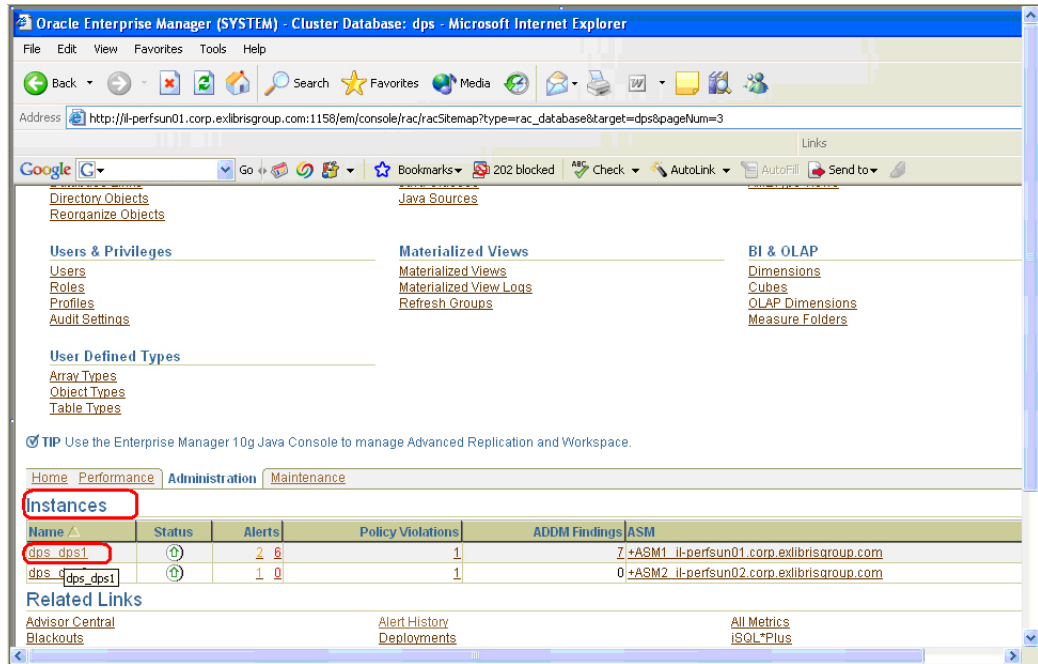


Figure 47: Node Selection

4 Click the **Alert Log Content** link.

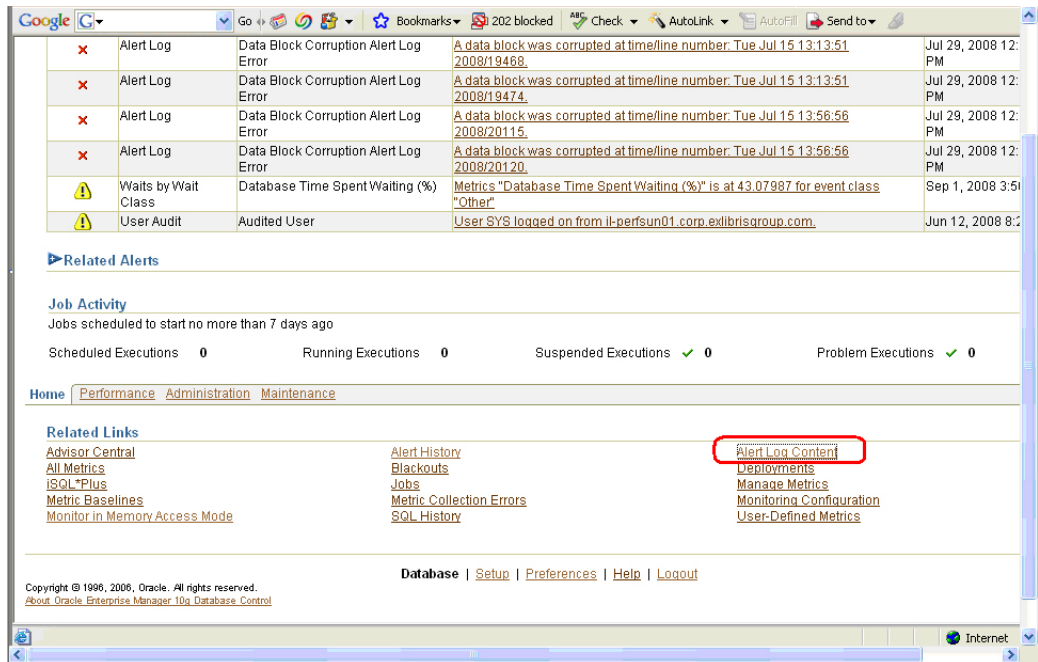


Figure 48: Alert Log Content Link

Viewing and Managing Information on Oracle Database Users

You can view and manage information on existing Oracle database users using the Oracle Enterprise Manager DB Controller (EM).

NOTE:

To create a new database user for the Rosetta system, use the command line script described in **Creating Database Users with Privileges** on page 118.

To view and manage information on existing Oracle database users:

- 1 Log on to the EM using the user **DPS_DBA**.
- 2 Select the **Administration** tab and in the **Users & Privileges** section, click the **Users** link.

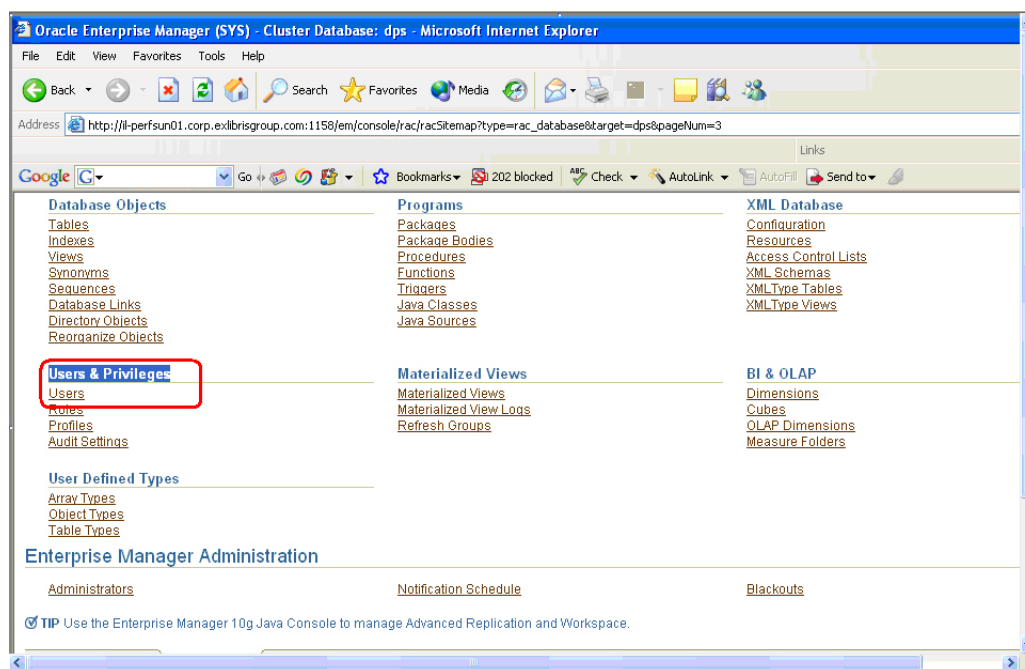


Figure 49: Users & Privileges List

- 3 Select a user to view and manage its status.

The screenshot shows the Oracle Enterprise Manager interface in a Microsoft Internet Explorer browser window. The browser address bar shows the URL: `http://li-perfsun01.corp.exlibrisgroup.com:1158/em/console/database/databaseObjectsSearch?event=search&otype=USER&target=dps&type=rac_database`. The page title is "Oracle Enterprise Manager (SYS) - Users". Below the browser window, there is a search results page with a table of database users. The table has columns for "Select", "UserName", "Account Status", "Expiration Date", "Default Tablespace", "Temporary Tablespace", "Profile", and "Created". The "Select" column contains radio buttons, with the first one for "ANONYMOUS" selected. The "Expiration Date" column shows "Jun 10, 2008 12:38:59 PM IDT" for several users. The "Created" column shows various dates in 2008.

Select	UserName	Account Status	Expiration Date	Default Tablespace	Temporary Tablespace	Profile	Created
<input checked="" type="radio"/>	ANONYMOUS	EXPIRED & LOCKED	Jun 10, 2008 12:38:59 PM IDT	SYSAUX	TEMP	DEFAULT	Jun 10, 2008 12:17:1
<input type="radio"/>	CTXSYS	OPEN		SYSAUX	TEMP	DEFAULT	Jun 10, 2008 12:16:4
<input type="radio"/>	D410_DEP00	OPEN		TS1	TEMP	EXL_PROFILE	Jul 29, 2008 8:16:55.
<input type="radio"/>	D410_JBPM00	OPEN		TS1	TEMP	EXL_PROFILE	Jul 29, 2008 3:00:47
<input type="radio"/>	D410_PER00	OPEN		TS1	TEMP	EXL_PROFILE	Jul 29, 2008 8:17:16.
<input type="radio"/>	D410_REP00	OPEN		TS1	TEMP	EXL_PROFILE	Jul 29, 2008 8:16:34.
<input type="radio"/>	D410_RPT00	OPEN		TS1	TEMP	EXL_PROFILE	Jul 29, 2008 8:17:33.
<input type="radio"/>	D410_SHR00	OPEN		TS1	TEMP	EXL_PROFILE	Jul 30, 2008 1:52:33
<input type="radio"/>	DBSNMP	OPEN		SYSAUX	TEMP	DEFAULT	Jun 10, 2008 12:11:2
<input type="radio"/>	DIP	EXPIRED & LOCKED		TS1	TEMP	DEFAULT	Jun 10, 2008 12:06:5
<input type="radio"/>	DMSYS	EXPIRED & LOCKED	Jun 10, 2008 12:38:59 PM IDT	SYSAUX	TEMP	DEFAULT	Jun 10, 2008 12:16:2
<input type="radio"/>	EXFSYS	EXPIRED & LOCKED	Jun 10, 2008 12:38:59 PM IDT	SYSAUX	TEMP	DEFAULT	Jun 10, 2008 12:16:1

Figure 50: List of Database Users

The following figure displays the configuration of the CTXSYS user:

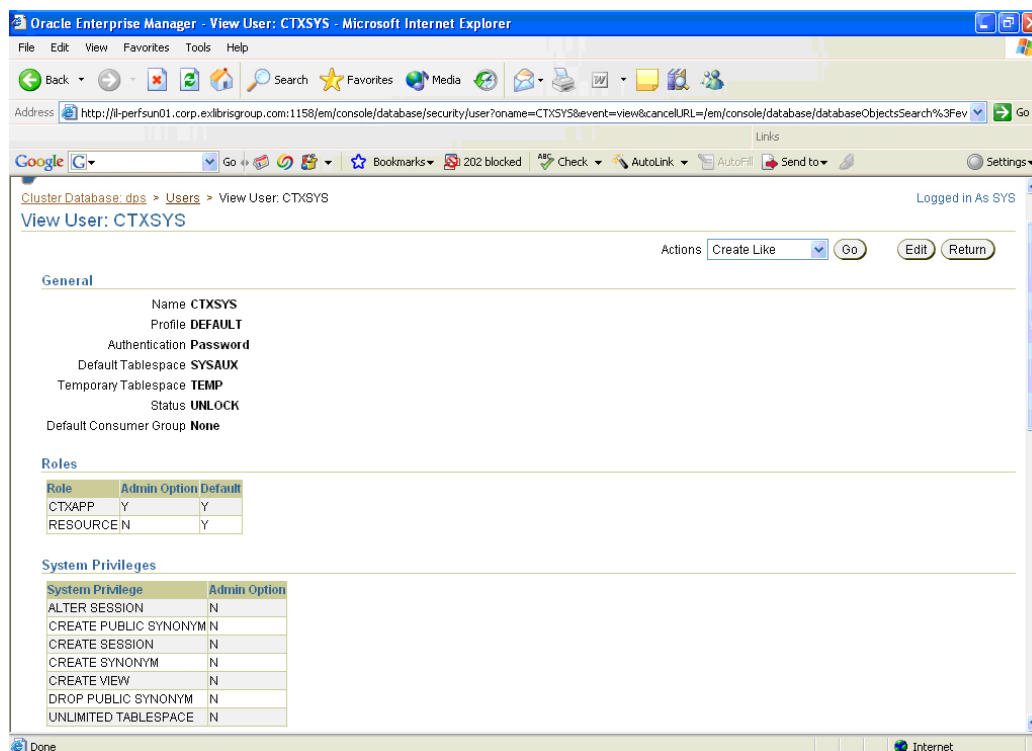


Figure 51: Example of Database User

Creating Database Users with Privileges

You create Oracle database users for Rosetta using the following command line script.

To create Oracle database users:

- 1 At the UNIX prompt, enter %dps_util.
- 2 To activate the script, enter the following command:

```
csh -f create_profile_role_user.csh
```

The following response is displayed:

```
Create Profile/Role/User Utility
-----
This script will create a profile and role, and/or admin users
and/or oracle users

Would you like to continue Y/[N]?
```

- 3 Choose **Y**. The following response is displayed:

```
-----
0. Quit
1. Create Oracle profile and roles
2. Create Oracle admin users
3. Create oracle user

Choose one of the following options:
```

- 4 Choose option **3** to create a new user. The new user name prompt is displayed:

```
Enter user name [D4I0_]:
```

- 5 Enter the user name. The following response is displayed:

```
-----
-----
Creating user: D4I0_TEST                               Application:
EXL
User password: D4I0_TEST
Oracle SID: dps
-----
-----
Warning: If Oracle schema D4I0_TEST exists all its data will
be erased!!!
Enter y to create oracle user D4I0_TEST [n]:
```

- 6 When you enter **y**, the following summary is displayed:

```
Creating D4I0_TEST user
.....OK
Granting EXL_SCHEMA role to D4I0_TEST user
..... OK
Granting encryption to D4I0_TEST user
..... OK
```


Part V

Data Backup and Recovery

This part contains the following:

- **Chapter 8: Backing Up and Recovering Data** on page 123

8

Backing Up and Recovering Data

This section contains:

- **About Backing Up Data** on page 123
- **Rosetta Scripts** on page 124
- **Backup Strategies** on page 125
- **Recovering Data** on page 127

About Backing Up Data

To minimize possible data loss, System Administrators must back up the following data:

- Rosetta software
- Producer Agent content
- Oracle database

Each of these data components must be backed up separately on the Deposit Server, Staging Server, and Permanent Repository. **Table 26** describes the location of each component:

Table 26. Data Location

Sub-folder	Contents
bytestream_work	Bytestream temporary working folder
cc_export_dir	Exported copy configuration files
digital_certificate	email-signing certificate
format_library_downloads	Downloaded Format Library files
logs	Application server logs
operational_delivery_shared	Delivery temporary working folder

Table 26. Data Location

Sub-folder	Contents
operational_export_directory	Import and export representations during preservation test/action
plugins	Plugins packaged with Rosetta are under <code>bundled</code> and custom plugins under <code>custom</code>
preserve_pp_docs	Signed-off preservation plans (see page 163 the Test Sign-Off section of the <i>Rosetta Preservation Guide</i>)
sipTmpDir	SIP temporary working folder
software	(legacy - can be removed)
sp_bck	Service Pack backup files
staff_work_area	Staff users' export/import files
ui	Customized logo files

Rosetta Scripts

System Administrators can use scripts to extract the names of the directories in which Producer Agent content is stored, as described in the following sections:

- **Deposit Application Role** on page 124
- **Repository Application Role** on page 125
- **Permanent Repository** on page 125

Deposit Application Role

The following script extracts the name of the directory in which Producer Agent content is stored by the Deposit application role:

```
#!/bin/csh -f
rm -f /tmp/deposit_directories
sqlplus -s > /tmp/deposit_directories << !
${ORA_USER_PREFIX}shr00/${ORA_USER_PREFIX}shr00
set pagesize 0 linesize 2048 heading off feedback off
trimspool on
column tablespace_name format a20
select value from hfrgeneralparameter where
name='logic_deposit_area';
exit
!
```

Repository Application Role

The following script extracts the name of the directory in which Producer Agent content is stored by the Repository application role:

```
rm -f /tmp/staging_directories
sqlplus -s > /tmp/staging_directories << !
${ORA_USER_PREFIX}shr00/${ORA_USER_PREFIX}shr00
set pagesize 0 linesize 2048 heading off feedback off
trimsPOOL on
column tablespace_name format a20
select value from storage_parameter,storage,storage_group
where key='DIR_ROOT' and module_type='REPOSITORY' and
storage_id = storage.id and group_id=storage_group.id;
exit
!
```

Permanent Repository

The following script extracts the name of the directory in which Producer Agent content is stored in the Permanent Repository:

```
rm -f /tmp/permanent_directories
sqlplus -s > /tmp/permanent_directories << !
${ORA_USER_PREFIX}shr00/${ORA_USER_PREFIX}shr00
set pagesize 0 linesize 2048 heading off feedback off
trimsPOOL on
column tablespace_name format a20
select value from storage_parameter,storage,storage_group
where key='DIR_ROOT' and module_type='PERMANENT' and
storage_id = storage.id and group_id=storage_group.id;
exit
!
```

Backup Strategies

System Administrators can perform backups, as described in the following sections:

- **About Backup Methods** on page 126
- **Implementing Backup Strategies** on page 126
- **Backup Scenarios** on page 127

About Backup Methods

System Administrators can back up database data using the methods described in [Table 27](#):

Table 27. Backup Methods

Method	Description	When to use
Cold backup	Performs a full backup on a closed database. Recovery of data that was backed up using this method does not require any additional files. The database can be restored to the same state it was in at the time the backup copy was created.	Each time the database is closed
Hot backup	Performs a full backup on an open database that runs in the archive log mode. Data that was backed up using this method must be recovered together with the archived redo logs, in order to synchronize the database.	Any time the cold backup method cannot be used due to down time
Archive redo logs	Contains only those changes that occurred since the last full backup (either cold or hot). Using archive redo logs allows System Administrators to minimize loss of new data that is stored between the creation of full backup copies.	Regularly, especially when the hot backup method is used

Implementing Backup Strategies

The best backup strategy is to perform, nightly, a cold backup, followed by the archiving of redo logs. When a cold backup cannot be performed each night due to down time, it is recommended to perform a hot backup each night instead.

In this case, a System Administrator can always recover data using the backup copies created on a previous night.

When a System Administrator cannot run a full backup every night, it is recommended to minimize the time interval between full backups. In this case, the role of archived redo logs becomes critical because they enable the System Administrator to restore the latest full backup, reapply transactions, and bring the database up to date.

In addition, backup tapes must be read regularly to check their validity. It is recommended to run a full backup tape listing at least once a week. Aside from verifying that the tape can be read properly, System Administrators must check the listing and make sure all expected directories and files were backed up.

The ability to restore an up-to-date site configuration file depends on the frequency at which backups are performed.

Backup Scenarios

System Administrators can use the following combinations of backup methods:

- C.A.S.E. - Cold + Archived redo logs + Site configuration + Export
- H.A.S.E. - Hot + Archived redo logs + Site configuration + Export
- A.S.E. - Archived redo logs + Site configuration + Export
- T.V. - Tape validity check
- B.I. - Backup integrity check

Table 28 lists possible scenarios for creating backup copies:

Table 28. Backup Scenarios

Day	Scenario 1	Scenario 2	Scenario 3
Monday	C.A.S.E.	C.A.S.E.	A.S.E.
Tuesday	C.A.S.E.	H.A.S.E.	H.A.S.E.
Wednesday	C.A.S.E.	C.A.S.E.	A.S.E.
Thursday	C.A.S.E.	H.A.S.E.	A.S.E.
Friday	C.A.S.E.	C.A.S.E.	C.A.S.E.
Saturday	no backup	no backup	no backup
Sunday	no backup	no backup	no backup
Weekly	T.V. + B.I.	T.V. + B.I.	T.V. + B.I.

Recovering Data

After multiple backup copies have been created, System Administrators must recover data using the following rules:

- A database backup must always be followed by file backup.
- If, after creating a backup copy of the database, multiple backup copies of files were created, a System Administrator must recover data using the

latest database backup, and the file backup closest to the latest database backup, as shown in **Figure 52**:

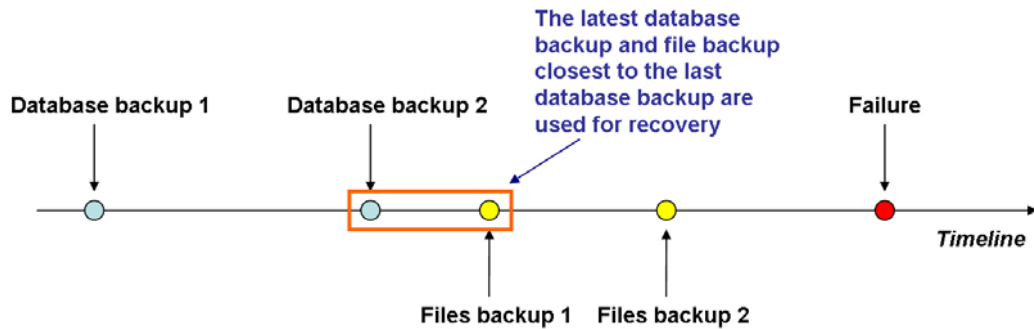


Figure 52: Recovering Data From Multiple Backup Copies

- If, after creating a backup copy of the database — but before creating a backup copy of files — new files were stored in the system, these files are lost because there are no appropriate records for them in the database, as shown in **Figure 53**:

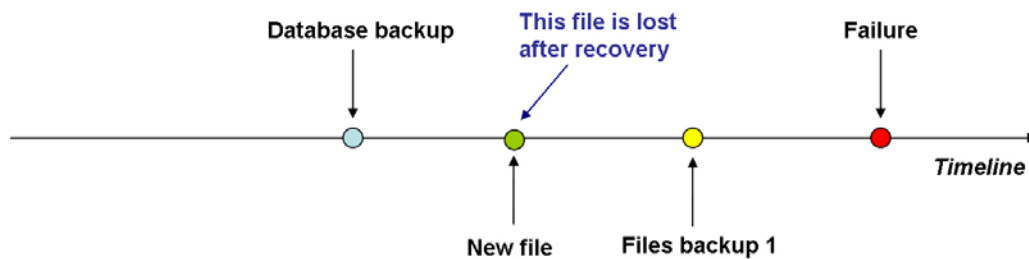


Figure 53: Losing Files Added Between Backups

Index

Numerics

2-Tier Topology, 24
2-tier topology, 24, 26

A

alert log, 114
aliases, 87
All-in-One Topology, 23
all-in-one topology, 23
application role server
 permanent, 18
archivelog mode, 110
archiving data, 110

B

background jobs, 89
backing up
 data, 123
 methods, 126
 strategies, 125, 127
 implementing, 126
 scenarios, 127
Backoffice parameters, 60
BIRT parameters, 60

C

common interval fields, 95
configuration
 external metadata, 53
connecting to database, 102
copy configuration, 78
 content, 78
 procedures, 79

D

data files, 108
database
 schema, 14
 password, 102
 users, 116
 privileges, 118
Delivery parameters, 64
Deposit
 parameters, 65

 role, 13
 storage, 15
disk space, 34

G

general parameters, 67
global.properties, 26

I

index
 rank, 20
 size, 20
 status, 21
interval fields, 95

J

jobs, 89
 monitoring, 92
 table, 90

L

log directory, 88
logging on, 85

M

Manage Scheduled Jobs link, 92
menu parameters, 69
metadata
 external, 53

N

network parameters, 70

O

operational jobs, 89
Operational Repository, 96
Oracle, 101–119
 data files, 108
 database log, 114
 database memory (SGA), 112
 Enterprise Manager DB Controller (EM), 108
 starting/stopping
 Listener, 103

- server, 102
- tablespaces, 106
- users, 116
 - creating, 118

P

- Patron Directory Service (PDS), 14
 - subdirectories, 88
- PDS, 24
- Permanent Repository, 96
- permanent repository settings, 71

R

- recovering data, 127
- repository
 - definitions, 96
 - parameters, 71
- repository application role
 - staging, 15
- Rosetta scripts, 124
 - Deposit application role, 124
 - Permanent Repository, 125
 - Repository application role, 125
- Rule List page, 47
- rules
 - match, 49
 - operators, 48
- Rules Details page, 47

S

- scalability, 27
- server
 - Solr, 20
- servers
 - starting/stopping, 83
 - Patron Directory Service (PDS), 83
 - Storage, 34
- shared directories, 96
- soft link, 43
- Solr server, 20
- storage
 - architecture, 35
 - definition, 37
 - groups, 36
 - adding, 37
 - deleting, 46
 - re-ordering, 45
 - updating, 45
 - handling method, 43
 - NFS plug-in, 40
 - rules, 47
 - activating, 53
 - adding, 50

- configuring default, 52
 - deactivating, 53
 - deleting, 52
 - re-ordering, 52
 - updating, 51
- Storage Groups List page, 36
- system
- aliases, 87
 - architecture, 11
 - jobs, 89
 - canceling, 95
 - scheduling, 93
 - layers, 11
 - verify operation, 85
- system configuration
- copying, 78

T

- tablespaces, 106
- topologies, 23

U

- Users & Privileges list, 116