HPSS Storage Broker Administrator’s Guide
# Table of Contents

1. System overview .................................................................................................................................................. ix

1.1. HSB Application Components ......................................................................................................................... 1

1.1.1. Copier ................................................................................................................................................................. 1

1.1.2. Storage endpoint ................................................................................................................................................. 1

1.1.2.1. Storage endpoint credentials ......................................................................................................................... 2

1.1.3. Storage policy ..................................................................................................................................................... 7

1.1.4. Project ............................................................................................................................................................... 8

1.1.5. Managed data set (MDS) ............................................................................................................................... 11

1.1.5.1. MDS catalog ...................................................................................................................................................... 12

1.1.5.2. MDS fragments ................................................................................................................................................. 12

1.1.5.3. MDS manifest .................................................................................................................................................... 13

1.1.5.4. MDS user metadata ......................................................................................................................................... 13

1.1.6. User .................................................................................................................................................................. 13

1.1.7. Security Domain ............................................................................................................................................... 14

1.1.8. Events .............................................................................................................................................................. 14

1.1.9. HSB database ................................................................................................................................................... 14

1.1.10. HSB logging ................................................................................................................................................... 15

1.1.10.1. HSB log location ........................................................................................................................................... 15

1.1.10.2. HSB log rotation ........................................................................................................................................... 15

1.2. HSB services ...................................................................................................................................................... 16

1.2.1. HSB request processor .................................................................................................................................... 16

1.2.2. HSB web services daemon ............................................................................................................................. 16

1.2.3. HSB copier .......................................................................................................................................................... 16

1.2.4. HSB credential expirer ..................................................................................................................................... 16

1.2.5. HSB client interface ......................................................................................................................................... 16

1.2.6. HSB administrative interface ....................................................................................................................... 17

1.2.7. HSB database ................................................................................................................................................... 17

1.2.8. HSB ZooKeeper ............................................................................................................................................... 17

2. Configuring the HSB instance overview ............................................................................................................. 18

2.1. Configure a HSB instance .................................................................................................................................. 18

2.2. Logging into the HSB administrative GUI ........................................................................................................ 19

3. Configuring HSB instance components ............................................................................................................. 21

3.1. Security menu .................................................................................................................................................. 21

3.1.1. User configuration ........................................................................................................................................... 21

3.1.1.1. Add user ........................................................................................................................................................... 21

3.1.1.2. User fields ...................................................................................................................................................... 23

3.1.1.3. Display user ................................................................................................................................................... 24

3.1.1.4. Edit user ........................................................................................................................................................ 25

3.1.1.5. Delete user ................................................................................................................................................... 27

3.1.2. Project role configuration ................................................................................................................................ 28

3.1.2.1. Add project role ............................................................................................................................................. 28

3.1.2.2. Project role fields ......................................................................................................................................... 29

3.1.2.3. Display project role ...................................................................................................................................... 29

3.1.2.4. Edit project role ............................................................................................................................................ 30

3.1.2.5. Delete project role ....................................................................................................................................... 30

3.1.3. Data privacy notice configuration .................................................................................................................. 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.3.1. Add data privacy notice</td>
<td>31</td>
</tr>
<tr>
<td>3.1.3.2. Data privacy fields</td>
<td>33</td>
</tr>
<tr>
<td>3.1.3.3. Display data privacy notice</td>
<td>33</td>
</tr>
<tr>
<td>3.1.3.4. Delete data privacy notice</td>
<td>34</td>
</tr>
<tr>
<td>3.1.4. Security domains</td>
<td>35</td>
</tr>
<tr>
<td>3.1.4.1. Add security domain</td>
<td>35</td>
</tr>
<tr>
<td>3.1.4.2. Security domain fields</td>
<td>36</td>
</tr>
<tr>
<td>3.1.4.3. Display security domain</td>
<td>37</td>
</tr>
<tr>
<td>3.1.4.4. Edit security domain</td>
<td>38</td>
</tr>
<tr>
<td>3.1.4.5. Delete security domain</td>
<td>38</td>
</tr>
<tr>
<td>3.2. Configure menu</td>
<td>39</td>
</tr>
<tr>
<td>3.2.1. Endpoint configuration</td>
<td>39</td>
</tr>
<tr>
<td>3.2.1.1. Add endpoint</td>
<td>39</td>
</tr>
<tr>
<td>3.2.1.2. Display endpoint</td>
<td>41</td>
</tr>
<tr>
<td>3.2.1.3. Display endpoint project credentials</td>
<td>43</td>
</tr>
<tr>
<td>3.2.1.4. Display endpoint user credentials</td>
<td>44</td>
</tr>
<tr>
<td>3.2.1.5. Edit endpoint</td>
<td>45</td>
</tr>
<tr>
<td>3.2.1.6. Edit endpoint project credentials</td>
<td>47</td>
</tr>
<tr>
<td>3.2.1.7. Edit endpoint user credentials</td>
<td>47</td>
</tr>
<tr>
<td>3.2.1.8. Delete endpoint</td>
<td>48</td>
</tr>
<tr>
<td>3.2.1.9. Delete endpoint project or user credentials</td>
<td>48</td>
</tr>
<tr>
<td>3.2.1.10. Download endpoint project or user credentials</td>
<td>48</td>
</tr>
<tr>
<td>3.2.1.11. Upload endpoint project or user credentials</td>
<td>49</td>
</tr>
<tr>
<td>3.2.1.12. Endpoint fields</td>
<td>50</td>
</tr>
<tr>
<td>3.2.1.13. Endpoint default credential fields</td>
<td>52</td>
</tr>
<tr>
<td>3.2.1.14. Endpoint project credential fields</td>
<td>53</td>
</tr>
<tr>
<td>3.2.1.15. Endpoint user credential fields</td>
<td>54</td>
</tr>
<tr>
<td>3.2.1.16. Endpoint specific configuration</td>
<td>55</td>
</tr>
<tr>
<td>3.2.2. Copier configuration</td>
<td>61</td>
</tr>
<tr>
<td>3.2.2.1. Add copier</td>
<td>61</td>
</tr>
<tr>
<td>3.2.2.2. Copier fields</td>
<td>62</td>
</tr>
<tr>
<td>3.2.2.3. Display copier</td>
<td>63</td>
</tr>
<tr>
<td>3.2.2.4. Edit copier</td>
<td>65</td>
</tr>
<tr>
<td>3.2.2.5. Delete copier</td>
<td>66</td>
</tr>
<tr>
<td>3.2.2.6. Copier logging</td>
<td>66</td>
</tr>
<tr>
<td>3.2.2.7. Controlling copier java heap size allocation</td>
<td>67</td>
</tr>
<tr>
<td>3.2.2.8. Controlling the copier thread pool size</td>
<td>67</td>
</tr>
<tr>
<td>3.2.2.9. Controlling copier HPSS PFTP data movement settings</td>
<td>67</td>
</tr>
<tr>
<td>3.2.3. Storage policy configuration</td>
<td>69</td>
</tr>
<tr>
<td>3.2.3.1. Create storage policy</td>
<td>69</td>
</tr>
<tr>
<td>3.2.3.2. Storage policy fields</td>
<td>70</td>
</tr>
<tr>
<td>3.2.3.3. Storage policy copy fields</td>
<td>71</td>
</tr>
<tr>
<td>3.2.3.4. Storage policy import source fields</td>
<td>72</td>
</tr>
<tr>
<td>3.2.3.5. Display storage policy</td>
<td>73</td>
</tr>
<tr>
<td>3.2.3.6. Edit storage policy</td>
<td>75</td>
</tr>
<tr>
<td>3.2.3.7. Edit storage policy copy</td>
<td>76</td>
</tr>
<tr>
<td>3.2.3.8. Edit import source</td>
<td>77</td>
</tr>
<tr>
<td>3.2.3.9. Delete storage policy</td>
<td>78</td>
</tr>
<tr>
<td>3.2.4. Project configuration</td>
<td>79</td>
</tr>
</tbody>
</table>
3.2.4.1. Create project ................................................................. 79
3.2.4.2. Project project ............................................................ 80
3.2.4.3. Display project ............................................................ 80
3.2.4.4. Project operation fields ............................................... 83
3.2.4.5. Project role fields ........................................................ 83
3.2.4.6. Edit project ................................................................. 84
3.2.4.7. Delete project ............................................................. 84
3.2.5. System configuration ...................................................... 86
3.2.5.1. List system configuration ........................................... 86
3.2.5.2. Display system configuration ..................................... 86
3.2.5.3. Edit system configuration .......................................... 87
3.2.5.4. System configuration fields ........................................ 88
3.3. Monitor menu ................................................................. 90
3.3.1. Events ........................................................................... 90
3.3.1.1. List event messages .................................................. 90
3.3.1.2. Display event message ............................................. 90
3.3.2. Requests ....................................................................... 92
3.3.2.1. List requests ........................................................... 92
3.3.3. Copiers ......................................................................... 96
3.3.3.1. List copiers ............................................................ 96
3.3.3.2. Monitor copier controls ........................................... 96
3.3.3.3. Monitor copier fields ............................................... 96
3.3.4. Endpoints .................................................................... 99
3.3.4.1. List endpoints .......................................................... 99
3.3.4.2. Monitor endpoint controls ....................................... 99
3.3.4.3. Monitor endpoint fields .......................................... 100
3.3.4.4. Monitor endpoint state at copiers fields ....................... 101
3.3.5. Servers ...................................................................... 102
3.3.5.1. List servers ............................................................ 102
3.3.5.2. Monitor servers fields ............................................. 102
3.4. Tools menu ..................................................................... 104
3.4.1. Datasets ..................................................................... 104
3.4.1.1. List datasets .......................................................... 104
3.4.1.2. Display dataset ..................................................... 104
3.4.1.3. Dataset fields ........................................................ 106
3.4.2. Fragment chunks ....................................................... 108
3.4.2.1. List fragment chunks ............................................. 108
3.4.2.2. Display fragment chunk ......................................... 108
3.4.2.3. Edit fragment chunk .............................................. 109
3.4.2.4. Fragment chunk fields ........................................... 109
3.4.3. Download Linux client installer .................................... 111
3.4.4. Download windows client installer ............................ 112
3.5. Help ............................................................................... 113
3.5.1. About .......................................................................... 113
3.6. HSB instance management ............................................ 114
3.6.1. Starting ...................................................................... 114
3.6.2. Stopping ..................................................................... 114
3.6.3. Database ................................................................. 114
3.6.4. Monitoring ................................................................. 115
List of Figures

1.1. Copier storage endpoint assignment example. ................................................................. 1
1.2. Storage policy storage endpoint assignment example. .................................................. 2
1.3. Storage endpoint credential assignment example. ......................................................... 3
1.4. Administrative GUI endpoint location entry ................................................................. 4
1.5. Administrative GUI default credential entry ................................................................. 4
1.6. User GUI endpoint explorer failure ................................................................................ 5
1.7. User GUI endpoint status screen .................................................................................... 5
1.8. User GUI endpoint set credentials ................................................................................ 6
1.9. User GUI endpoint explorer .......................................................................................... 7
1.10. Storage policy selection for managed data set creation ............................................... 8
1.11. Example of project and managed data set grouping. .................................................... 9
1.12. Project role hierarchy. ................................................................................................... 10
1.13. Example of project role assignments. .......................................................................... 10
1.14. Example of data flowing from static storage endpoint to a managed data set. ............ 11
1.15. Example of copier writing managed data set fragments to storage endpoints defined in the selected storage policy. ................................................................. 12
1.16. Example of user metadata (XML file) being added to a managed data set. ................. 13
2.1. HSB administrative GUI login screen ............................................................................ 20
Copyright notification. Copyright © 2022 International Business Machines Corporation, The Regents of the University of California, Triad National Security, LLC, Lawrence Livermore National Security, LLC, National Technology & Engineering Solutions of Sandia, LLC, and UT-Battelle. All rights reserved.

Portions of this work were produced by Lawrence Livermore National Security, LLC, Lawrence Livermore National Laboratory (LLNL) under Contract No. DE-AC52-07NA27344 with the U.S. Department of Energy (DOE); by the University of California, Lawrence Berkeley National Laboratory (LBNL) under Contract No. DE-AC02-05CH11231 with DOE; by Triad National Security, LLC, Los Alamos National Laboratory (LANL) under Contract No. 89233218CNA000001 with DOE; by National Technology & Engineering Solutions of Sandia, LLC (NTESS), Sandia National Laboratories (SNL) under Contract No. DE-NA0003525 with DOE; and by UT-Battelle, Oak Ridge National Laboratory (ORNL) under Contract No. DE-AC05-00OR22725 with DOE. The U.S. Government has certain reserved rights under its prime contracts with the Laboratories.

DISCLAIMER. Portions of this software were sponsored by an agency of the United States Government. Neither the United States, DOE, The Regents of the University of California, Triad National Security, LLC, Lawrence Livermore National Security, LLC, National Technology & Engineering Solutions of Sandia, LLC, UT-Battelle, nor any of their employees, makes any warranty, express or implied, or assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

Trademark usage. High Performance Storage System is a trademark of International Business Machines Corporation.

IBM is a registered trademark of International Business Machines Corporation.

IBM, Db2, Db2 Universal Database, and AIX are trademarks or registered trademarks of International Business Machines Corporation.

UNIX is a registered trademark of the Open Group.

Linux is a registered trademark of Linus Torvalds in the United States and other countries.

Java is a registered trademark of Oracle and/or its affiliates.

Microsoft Windows or Windows is a registered trademark of Microsoft Corporation.

Amazon Web Services or AWS is a registered trademark of Amazon Technologies, Inc.

Swift is a registered trademark of Apple Inc.

Optical Disc Archive is a registered trademark of Sony Group Corporation.

Other brands and product names appearing herein may be trademarks or registered trademarks of third parties.

Who should read this book. The HPSS Storage Broker Administrator’s Guide is intended as a resource for HPSS Storage Broker (HSB) administrators. Chapter 1 provides a quick overview
of the system. Chapter 2 assists in the configuration of an HSB instance. Chapter 3 assists in the configuration of specific HSB components. Chapter 4 provides instructions on managing an HSB instance.

**Conventions used in this book.** Example commands that should be typed at a command line are proceeded by a percent sign ("%") and be presented in a Courier font:

```plaintext
% sample command
```

Any text preceded by a pound sign ("#") should be considered comment lines:

```plaintext
# This is a comment
```

Angle brackets ("<>") denote a required argument for a command:

```plaintext
% sample command <argument>
```

Square brackets ("[]") denote an optional argument for a command:

```plaintext
% sample command [optional argument]
```

Vertical bars ("|") denote different choices within an argument:

```plaintext
% sample command <argument1 | argument2>
```
Chapter 1. System overview

1.1. HSB Application Components

1.1.1. Copier

The copiers are responsible for reading and writing data. Each copier is assigned one or more storage endpoints. More than one copier can be assigned to read and write from a storage endpoint.

Figure 1.1. Copier storage endpoint assignment example.

Each HSB system requires at least one copier and can support multiple copiers.

Each copier provides the following functions:

- Creates, deletes, and recalls managed data sets.
- Copy data between storage endpoints.
- List files, objects, directories and containers on storage endpoints.

1.1.2. Storage endpoint

Storage endpoints define the locations the copiers can read from and write to. There are two types of storage endpoints:

1. Static - These are storage locations that contain user files, objects, directories, and containers, usually visible to the HSB servers. They can also used as targets for managed data set recalls.

2. Repository - These endpoints are used to store the managed data set fragments and catalog files.

The following protocols are supported when configuring storage endpoints:

- filesystem  Filesystem local to the copier.
- hpss  HPSS High Performance Storage System.
before storage endpoints can be used, they must be assigned to copiers (see copier storage endpoint assignment example). once assigned, they can be included in storage policy definitions. each storage policy copy must be configured with one or more repository storage endpoints before it can be used to write managed data set fragments.

**figure 1.2. storage policy storage endpoint assignment example.**

**1.1.2.1. Storage endpoint credentials**

Credentials are required to access storage endpoints. Endpoint credentials are used by HSB to access the storage endpoint, they are not used for authenticating with HSB. Default credentials are required for the application to report endpoint availability. Endpoint availability is verified by reading the endpoint folder/container. The default credentials should provide enough privilege to accomplish this task. You can also configure credentials for projects and for individual users.
Here is an example on how the administrator can use credentials for an endpoint. We will be setting credentials for a static filesystem endpoint on a Linux server.

1. For the endpoint default credential, HSB will need a user on the Linux system. On the test system, we used "hsbep" as the endpoint default credential account name and "tfuser1" and "tfuser2" as user accounts. They are configured like this:

   ```bash
   $ sudo -E groupadd hsbep
   $ sudo -E /usr/sbin/useradd hsbep -r -g hsbep -p $(openssl passwd -1 mypassword)
   $ sudo -E /usr/sbin/useradd tfuser1 -r -g tfuser1 -p $(openssl passwd -1 mypassword)
   $ sudo -E /usr/sbin/useradd tfuser2 -r -g tfuser2 -p $(openssl passwd -1 mypassword)
   $ id hsbep
   uid=985(hsbep) gid=1004(hsbep) groups=1004(hsbep)
   ```

2. The new account was then tested by running the `su` command with the non-root user to make sure we could log in with that password.

3. The HSB endpoint will also need storage space and a path that points to that space. On our Linux system we created a path `/hsb_endpoints/home`.

   ```bash
   ls -ld /hsb_endpoints/home
   drwxr-xr-x 5 hsbep hsbep 4096 May 21 15:20 /hsb_endpoints/home
   ```

4. The administrator then created individual user directories under the endpoint path on the Linux system.

   ```bash
   [hsbep@testsystem:/hsb_endpoints] ls -l /hsb_endpoints/home
   total 12
   drwxr-x---  4 tfuser1 tfuser1 4096 May 21 15:27 tfuser1
   drwxr-x---  4 tfuser2 tfuser2 4096 May 21 15:29 tfuser2
   ```
5. On our HSB test system we used the administrative GUI to configure a static endpoint named "HSB User Home" and set the path to /hsb_endpoints/home.

**Figure 1.4. Administrative GUI endpoint location entry**

![Administrative GUI endpoint location entry](image)

6. We use the *hsbe* account and password (*mypassword*) to configure the default endpoint credential. Once the endpoint was configured we assigned it to the copier on the test system.

When adding user credentials via the HSB administrative GUI, the administrator may not know the users username and password for an endpoint. When this occurs, the administrator can add a user credential for the HSB user and specify "none" for the user credential client name and password. This will allow the HSB user to see and edit the credential username and password via the client GUI *Endpoint Status and Credential Configuration* screen.

**Figure 1.5. Administrative GUI default credential entry**

![Administrative GUI default credential entry](image)

The screen below indicates the error a user would receive if attempting to view an endpoint in endpoint explorer when the administrator has specified "none" for the user credential client name and password.
If the user looks at the Endpoint Status and Credential Configuration screen (see below), they can see the *none* user credential created by the administrator.

From this screen the user can update the endpoint user credential, to their Linux user and secret.
Once the credential is updated, the user can see the contents of their Linux directory from the explorer. They will not be able to access directories their Linux user does not have permissions for.
Figure 1.9. User GUI endpoint explorer

The process for setting project credentials is similar to the process for setting the user credentials. However only the HSB project owners will be able to modify the project credentials with the user GUI.

1.1.3. Storage policy

A storage policy defines how managed data sets get written to the storage endpoints.

Each storage policy defines:

- The number of copies to create.
- The storage endpoints each copy is written to.
- Each copy’s fragmentation and parity characteristics.

One or more storage policies can be assigned to a project. A user can select a storage policy to use when creating a managed data set under the project.
Figure 1.10. Storage policy selection for managed data set creation

1.1.4. Project

Projects are used to group managed data sets and provide access control.
Access to projects, and the managed data sets they contain, is controlled by project roles. Users are granted a specific role on a project, which provides access to its managed data sets. Users can be granted one of the following project roles (ordered from least to most permissive) per project:

- **Read** list managed data sets and recall data content.
- **Update** create new managed data sets. Includes all read privileges.
- **Delete** delete managed data sets. Includes all update privileges.
- **Owner** add, update and delete other non-owner users and modify their project roles. Includes all delete privileges.
Figure 1.12. Project role hierarchy.

Each project role includes the permissions of the less powerful roles.

Figure 1.13. Example of project role assignments.

Each project has the following attributes:
• Unique name.
• Event record retention period.
• Default managed data set manifest retention period.
• Operation Limits - represents the number of requests of a specific type that can execute simultaneously:
  • Create
  • Recall
  • Verify
  • Delete

1.1.5. Managed data set (MDS)

A managed data set is an immutable collection of files or objects managed by the HSB service. Managed data sets are intended as a mechanism to physically group associated data on storage endpoints and to optimize the desirable characteristics of high-latency, high-capacity storage. User-defined metadata can be associated with each managed data set to facilitate efficient location and retrieval.

Figure 1.14. Example of data flowing from static storage endpoint to a managed data set.
The HSB copier logically aggregates all of the files/objects and directories/containers from the selected storage endpoint and then writes the MDS fragments (generated from the logical aggregate) to the repository storage endpoints as defined by the selected storage policy. An MDS catalog file is also written to each endpoint with each fragment.

**Figure 1.15. Example of copier writing managed data set fragments to storage endpoints defined in the selected storage policy.**

![Diagram of managed data set and copier](image)

### 1.1.5.1. MDS catalog

An MDS catalog file is generated describing the contents of the managed data set (MDS) when the MDS is created. The MDS catalog file uses the SNIA-SIRF self-contained information format. A separate MDS catalog file is generated for each copy defined in the selected storage policy. A copy of the MDS catalog file is written with each MDS fragment. A checksum value is generated for each MDS catalog file associated with a storage policy copy and stored in the database.

### 1.1.5.2. MDS fragments

An MDS fragment represents a contiguous subset of the data within a managed data set. Managed data sets are fragmented as equally as possible into the number of fragments specified by storage policy copy. Fragmentation facilitates increased transfer performance via concurrent, parallel transfers and provides redundancy via the generation of parity fragments.
1.1.5.3. MDS manifest

An MDS manifest contains details about each file or object in the managed data set and is stored in the HSB database.

1.1.5.4. MDS user metadata

Users can optionally associate metadata with a managed data set to facilitate quick and efficient searches. The user metadata interface allows HSB users to upload one or more XML files that further describe the contents of the managed data set. Custom indexes can be created over the XML data to increase the efficiency of the searches.

The XML files can optionally be validated against XML schemas that have been registered in the HSB database and associated with the project containing the managed data set.

Figure 1.16. Example of user metadata (XML file) being added to a managed data set.

1.1.6. User

HSB users must be registered by a HSB administrator using the HSB administrative GUI. The HSB installation tool creates the default HSB administrative account "admin".

Users can have one of the following roles:

- Admin
• Has privileges to perform all administrative functions including:
  • Managing the HSB server configuration.
  • Registering, updating, and deleting HSB users.
  • Can also be given project privileges via project roles.

User

• Given project privileges via project roles.

• Will have limited privileges in the HSB administrative GUI.

1.1.7. Security Domain

A security domain is used to define the configuration of a Lightweight Directory Access Protocol (LDAP) server that HSB can use to authenticate a HSB user with a LDAP directory service. Defining a security domain is only necessary if you would like to use LDAP for authentication on one or more HSB user accounts. Multiple security domains can be defined if there are LDAP user accounts stored across different LDAP directory services.

1.1.8. Events

The HSB system records various events in the database event log. These events include major events such as:

• Successful creation of a managed data set
• Successful deletion of a managed data set
• Successful recall of a managed data set
• Failed request
• Suspended request
• Successful user login
• Failed user login
• User password expired
• User added to project

1.1.9. HSB database

The HSB system uses an IBM Db2 database to manage the application metadata. This includes:

• Configuration data
• Copiers
• Storage endpoint assignments
• Storage endpoints
• Storage policies
• Storage policy copy definitions
• Users
• Project data
• Managed data sets
• Managed data set catalogs
• Managed data set fragments
• Managed data set manifests
• Managed data set user metadata
• Projects
• Project roles

1.1.10. HSB logging

1.1.10.1. HSB log location

By default, HSB logs are stored in the following locations:

• /var/log/treefrog/treefrog.log - Contains HSB request processor logs.

• /var/log/treefrog/infra.log - Contains HSB infrastructure logs.

The above locations can be modified from the /etc/rsyslog.d/treefrog.conf file. Please consult with HPSS support if changes to the HSB logging configuration are necessary.

• /var/log/treefrog/copier.log - Contains HSB copier logs.

The location above can be modified from the /etc/rsyslog.d/treefrog-copier.conf file. Please consult with HPSS support if changes to the HSB logging configuration are necessary.

1.1.10.2. HSB log rotation

Rotating a log file renames and compresses it. The format of the rotated log file name is <filename>-YYYYMMDD.gz. The rotated log file remains in the default logging directory.

For example: if /var/log/treefrog/treefrog.log is rotated on February 3rd 2020 it is renamed and compressed, the name becomes /var/log/treefrog/treefrog.log-20200203.gz and a new /var/log/treefrog/treefrog.log is created.
By default logs are:

- Rotated weekly if larger than 100 KiB.
- Compressed after rotation.
- Rotated automatically when the log file size reaches 100 MiB.

Log rotation is facilitated by the system administration tool **logrotate**.

Log rotation settings for `treefrog.log` and `infra.log` are stored in `/etc/logrotate.d/treefrog` and can be modified.

Log rotation settings for `copier.log` are stored in `/etc/logrotate.d/treefrog-copier` and can be modified.

See the man page for **logrotate** for details on the parameters you can set.

### 1.2. HSB services

#### 1.2.1. HSB request processor

The HSB request processor is responsible for taking a request from start to completion. It handles updating the current request state and status throughout the process. The list of current requests is maintained in ZooKeeper to protect against server failures.

Requests come from the HSB web services daemon and are processed by the HSB copiers.

Completed dataset information is stored in Db2.

#### 1.2.2. HSB web services daemon

The HSB web services daemon is a web application that uses Apache httpd. It handles processing requests to the HSB user interfaces. Users and administrators interact with the HSB web services daemon, not the HSB request processor.

#### 1.2.3. HSB copier

The HSB copier handles transferring data between storage endpoints. It uses Apache JClouds to transfer the data. The HSB request processor schedules requests on the HSB copier.

#### 1.2.4. HSB credential expirer

The HSB credential expirer looks for expired user credentials and deletes them.

#### 1.2.5. HSB client interface

The HSB client interface allows users to list, create, or recall managed data sets for projects they have been granted a project role. Users can also view their running jobs or query for the results of previous jobs.
1.2.6. HSB administrative interface

The HSB administrative interface is a web-based UI. It allows the administrators to configure, manage, and monitor an HSB instance.

1.2.7. HSB database

The HSB database contains all of the persistent data for the HSB instance. This includes:

- copiers
- storage endpoints
- storage policies
- users
- roles
- projects
- managed datasets
- dataset manifest
- request events

The information stored in the database is accessed by the HSB request processor and the HSB web services daemon.

1.2.8. HSB ZooKeeper

Apache ZooKeeper is used to store semi-persistent data about the running HSB system. For example, the request queue, containing all of the active HSB requests along with their status, is maintained in ZooKeeper.

ZooKeeper data is stored in the dataDir location defined in the zookeeper.cfg. The default location is /var/treefrog/zk/data. This can be changed before the system is running.
Chapter 2. Configuring the HSB instance overview

2.1. Configure a HSB instance

This section provides a brief overview on how to configure a HSB system. For details about installing HSB, see the HSB Installation Guide.

1. Install HSB server and copiers.

   Install the HSB server RPMs on the designated system. Once complete, install the HSB copier RPMs on the designated copier systems. Set up the instance using the `treefrog_setup` setup command and make note of the primary administrator password.

2. Log in as the primary administrator.

   Open a browser and login to the HSB administrative GUI as the HSB administrator that was created during the installation.

3. Configure the endpoints.

   Use the steps for adding an endpoint, see Endpoint configuration section, to create the planned endpoints for the system. Be sure to add default credentials. Don’t worry about setting up user credentials during this step. That takes place in step 7 below.
   Check to make sure all of the endpoints have been created on the endpoint monitor view, see Monitor endpoints section.

4. Configure the copiers.

   Use the steps for adding a copier, see Copier configuration section, to create the planned copiers for the system. Be sure to download and install the copier bundles.
   Check to make sure all of the copiers have been created and are in the **ONLINE** state in the copier monitor view, see Monitor copiers section.

5. Configure the policies.

   Use the steps for adding a storage policy, see Storage policy configuration section, to create the planned storage policies for the system.
   After creating the policy, be sure to configure the storage policy copies. Once the storage policy copies have been created, then edit the storage policy and set the **Preferred Recall Copy** field.

6. Configure the projects.

   Use the steps for adding a project, see Project configuration section, to create the planned projects for the system.

7. Optionally configure the security domain(s).
Configuring the HSB
instance overview

If HSB users will be using LDAP for user authentication a security domain should be added before adding the users. Use the steps for adding a security domain, see Security domains section, to create the security domain(s).

Once the security domain(s) are created, a security domain can be associated with a HSB user account during the steps for adding a user, see User configuration section

8. Configure the users.

Use the steps for adding a user, see User configuration section, to create the planned users for the system.

Once the users are created, then give the users roles for the existing projects using the project role list view, see the Project role configuration section.

2.2. Logging into the HSB administrative GUI

1. Open a browser and enter the URL https://<hostname> where "<hostname>" is replaced with the hostname of the system the HSB server was installed on.

2. If this is the first time you’ve visited this URL with this browser you will be presented with a browser warning message. This warning message is displayed because the HSB server uses a self-signed certificate. The warning message and the steps to trust the self-signed certificate vary depending on the particular browser being used. For some of the more commonly used browsers follow the steps below to accept the self-signed certificate and get access to the HSB administrative GUI.

   • Firefox

   1. Firefox will display the warning message "Your connection is not secure". Click the <Advanced> button.

   2. Click the <Add Exception> button.

   3. Click the <Confirm Security Exception> button.

   • Chrome

   1. Chrome will display the warning message "Your connection is not private". Click the <Advanced> button.

   2. Click the link that says "Proceed to <hostname> (unsafe)".

   • Safari

   1. Safari will display the warning message "This Connection Is Not Private". Click the <Show Details> button.

   2. Click the link that says "visit this website".

   • Edge
1. Edge will display the warning message "This site is not secure". Click the "Details" text.

2. Click the link that says "Go on to the webpage (Not recommended)".

3. On the HSB administrative GUI landing page select Login. If you’re initially setting up the HSB system for the first time you will need to log in with the username "admin" and enter the password you specified when you ran the `treefrog_setup setup` command (see the *HSB Installation Guide*). Otherwise, enter a valid HSB username and password.

**Figure 2.1. HSB administrative GUI login screen**
Chapter 3. Configuring HSB instance components

3.1. Security menu

3.1.1. User configuration

3.1.1.1. Add user

Select Security > Users in the HSB administrative GUI:

The List Users window is displayed:

This window shows all of the users configured to use the HSB system.

From the List Users window, click the plus button to open the Add User window.
After entering the user information, click on **Save**.
3.1.1.2. User fields

User Name
Required. User name for the user. The maximum size is 64 alpha-numeric characters, including underscores. The user name must be unique within the HSB instance.

User Roles
Roles the user holds in the system. The defined system roles are:

- **Admin:** Indicates that this is an administrative account. Users holding this role are able to add, update, and delete system configuration items. These users can also monitor and modify ongoing request activities.

  The admin role confers no project roles. Any required project roles must be assigned separately.

- **User:** Indicates that this is a non-administrative account. Users holding this role are able to be members of projects and be assigned project roles. See Project roles section for the project roles available for users.

Password
Password for the user’s account. A password is required if the user account is not associated with a security domain. If a user account is associated with a security domain the password field must be left blank. The maximum size is 256 characters.

Confirm Password
Confirmation of the password for the user’s account. The maximum size is 256 characters.

User State
Set whether the HSB user account is enabled or disabled.

Security Domain
Optional. The security domain this user account will authenticate with. If an existing HSB user account does not have the security domain set, setting the security domain will clear out any existing password associated with the user account as authentication will now be handled through the security domain. If an existing HSB user account has the security domain set, removing the security domain will require the HSB user to reset their password as authentication will now be handled locally through HSB.

Security Domain Account
Optional. Allows you to identify what user name on the external security domain will be used for authentication of this account, if it differs from the HSB user name choosen for this user.

First Name
Required. First name of the user. The maximum size is 256 characters.

Last Name
Required. Last name of the user. The maximum size is 256 characters.

Email Address
Required. Email address for the user. The maximum size is 256 characters.
Configuring HSB
instance components

**Phone (Mobile)**
Mobile phone number of the user. The maximum size is 32 characters.

**Phone (Office)**
Office phone number for the user. The maximum size is 32 characters.

**Office Address**
Office address for the user. The maximum size is 256 characters.

---

### 3.1.1.3. Display user

From the *List Users* window, click the magnifying glass button to open the *Show User* window.

When finished select the "back" button.
Configuring HSB
instance components

**Login Count**
Number of times user has logged in.

**Last Logged In On**
Date when user last logged in.

**Failed Login Count**
Number of times user has failed to log in.

**Created On**
Date the user was created.

**Updated On**
Date the user was last updated.

**Password Last Changed On**
Date the user’s password was last changed.

See *User fields* section for other project field descriptions.

To reset a password click the **Reset Password** button.

![Reset Password Form](image)

Enter a new password, then enter it again to confirm it, and click **Save**.

### 3.1.1.4. Edit user

From the *List Users* window, click the notepad-and-pen button to open the *Edit User* window.
After you have entered the user information, click on Save:

See User fields section for user field descriptions.
3.1.1.5. Delete user

From the *List Users* window, click the eraser button next to the desired user record to delete the user.

A user cannot be deleted if they have been assigned any project roles. The project role assignments must be removed first.
3.1.2. Project role configuration

3.1.2.1. Add project role

Select Security > Project Roles in the HSB administrative GUI:

The List Project Roles window is displayed:

This window shows every user of a project and the role they have on the project.

From the List Project Roles window, click the plus button to open the Add Project Role window.

Before adding a project role you must have defined one or more projects and one or more users on the system.
After entering the user information, click on Save.

See Project role fields section for project role field descriptions.

### 3.1.2.2. Project role fields

**Project**
An immutable dropdown list of the projects currently defined on the system. Select a project from the list for which you want to add a new role.

**User**
An immutable dropdown list of the users currently defined on the system. Select a user from the list for which you want to assign a role on the project.

**Role**
The role assigned to the user for the project. The roles have the permission described below and the permissions for all of the roles below it. The roles are:

- **owner**: An owner of the project. Owners are considered the owner of the data stored in the project. They have permission to access and delete datasets stored in the project. They can modify the metadata associated datasets. They can also add users to the project and modify roles for members of the project (except for other owners).

- **delete**: The user can delete datasets from the project.

- **update**: The user can modify the metadata for datasets in the project.

- **create**: The user can create datasets in the project.

- **read**: The user can see the datasets in the project and recall them from the system.

### 3.1.2.3. Display project role

From the List Project Roles window, click the magnifying glass button next to the desired project role record to open the Show Project Role window.

See Project role fields section for project role field descriptions.
Click the "back" button when finished.

### 3.1.2.4. Edit project role

From the *List Project Roles* window, click the notepad-and-pen button next to the desired project role record to open the *Edit Project Role* window.

![Edit Project Roles](image.png)

See *Project role fields* section for project role field descriptions.

After editing the project role information, click on *Save*.

### 3.1.2.5. Delete project role

From the *List Project Roles* window, select the eraser button next to the desired project role to delete the project role.

Deleting a project role is prevented if the user is the last owner of the project and the project contains one or more managed data sets.
3.1.3. Data privacy notice configuration

Select **Security > Data Privacy** in the HSB administrative GUI:

The **List Data Privacy Notices** window is displayed:

This window shows a history of the data privacy notices that have been created, including the start date and the expiration date of each notice. Only one data privacy notice can be active at a time. When a new data privacy notice is created, the previous active notice is set to expired.

3.1.3.1. Add data privacy notice

To create a new data privacy notice from the **List Data Privacy Notices** window, click the plus button to open the **Upload a new Data Privacy Notice** window.
After entering the data privacy notice, click on Save.

When a new notice is created, this notice is displayed to each user after logging into the client GUI. The user must accept the notice to use the client GUI. Once accepted by the user, they will not be shown the Privacy Notice screen again unless a new notice is created.
3.1.3.2. Data privacy fields

**Name**  
Required. Enter a unique name to identify the data privacy notice. The maximum size is 100 ASCII characters.

**Message Text**  
Enter the message text you would like the user to acknowledge after logging into the client GUI. The maximum size is 32,000 ASCII characters.

3.1.3.3. Display data privacy notice

From the *List Data Privacy Notices* window, click the magnifying glass button next to the desired privacy notice to open the *Show Data Privacy Notice* window.
Configuring HSB instance components

Identity Numeric identity assigned to notice at creation time.
Name The name of the notice.
Message Text The text of notice.
Start Date The date the notice was created.
Expiration Date The date this notice was replaced by a new one.

Click the "back" button when finished.

3.1.3.4. Delete data privacy notice

From the List Data Privacy Notices window, click the eraser button next to the desired privacy notice to delete it.

Only notices that are expired can be deleted.
3.1.4. Security domains

3.1.4.1. Add security domain

Select Security > Security Domains in the HSB administrative GUI:

The List Security Domains window is displayed:

This window shows all of the security domains configured on the HSB system.

From the List Security Domains window, click the plus button to open the Add LDAP Security Domain Configuration window.
3.1.4.2. Security domain fields

**Domain Name**
Required. A unique name used to refer to this security domain. The maximum size is 128 characters.

**Host Name**
Required. The host name of the LDAP server.

**Host Port**
Required. The port used to connect to the LDAP server. Defaults to port 636 which is the port most commonly used for LDAP TLS/SSL communication.

**Bind User Name**
Required. The user name of an LDAP account to use to bind to the LDAP server for authentication.

After entering the user information, click on **Save**.
Bind User Password
Required. The password associated with the bind user account.

Confirm Password
Required. Confirmation of the password associated with the bind user account.

Base DN Name
Required. The base DN from where the LDAP server will search for user accounts.

User Object Filter
Required. Filter used to query for user account information.

OpenLDAP
If checked indicates the security domain is using the open-source implementation of the LDAP software. This field must be set correctly for successful authentication to occur.

Allow Self-Signed Certificates
If checked indicates any errors related to the TLS/SSL certificate being self-signed are ignored.

CA Certificate Path
File path for the location of the client certificate being used for TLS/SSL communication with the LDAP server.

3.1.4.3. Display security domain

From the List Security Domains window, click the magnifying glass button to open the Show LDAP Security Domain Configuration window.

When finished select the "back" button.
See Security domain fields section for field descriptions.

3.1.4.4. Edit security domain

From the List Security Domains window, click the notepad-and-pen button to open the Edit LDAP Security Domain Configuration window.

After you have entered the security domain information, click on Save:

See Security domain fields section for field descriptions.

3.1.4.5. Delete security domain

From the List Security Domains window, click the eraser button next to the desired security domain record to delete the security domain.
3.2. Configure menu

3.2.1. Endpoint configuration

3.2.1.1. Add endpoint

Select **Configure > Endpoints** in the HSB administrative GUI:

In the **List Storage Endpoints** window, select the plus button to open the **Add Storage Endpoint** window.
### Configuring HSB

#### instance components

**Add Storage Endpoint**

<table>
<thead>
<tr>
<th><strong>General</strong></th>
<th><strong>Add Storage Endpoint</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endpoint Name</strong></td>
<td>Endpoint Name</td>
</tr>
<tr>
<td><strong>Endpoint Description</strong></td>
<td>Endpoint Description</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>repository</td>
</tr>
<tr>
<td><strong>Endpoint in Enabled</strong></td>
<td>Is the endpoint enabled?</td>
</tr>
<tr>
<td><strong>Passwords Stored</strong></td>
<td>User and project passwords are stored for endpoint credentials</td>
</tr>
<tr>
<td><strong>Endpoint Idle Connection Timeout</strong></td>
<td>Endpoint idle connection timeout in seconds. If left blank a default timeout is used based on the endpoint protocol</td>
</tr>
<tr>
<td><strong>Ping Limit Interval</strong></td>
<td>Ping Limit Interval</td>
</tr>
<tr>
<td><strong>Maximum Concurrent Requests</strong></td>
<td>Maximum Concurrent Requests</td>
</tr>
<tr>
<td><strong>Maximum object size in MiB</strong></td>
<td>Maximum object size in MiB to store on this endpoint</td>
</tr>
</tbody>
</table>

**Location**

<table>
<thead>
<tr>
<th><strong>Protocol</strong></th>
<th>NEsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Uri Host</strong></td>
<td>ioss1.deerlake.ibm.com</td>
</tr>
<tr>
<td><strong>The host name or IP address used by this endpoint</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Uri Port</strong></td>
<td>URI Port</td>
</tr>
<tr>
<td><strong>The port number used by this endpoint</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Uri Path</strong></td>
<td>URI Path</td>
</tr>
<tr>
<td><strong>The absolute path relative to this endpoint</strong></td>
<td></td>
</tr>
</tbody>
</table>
See *Endpoint fields* section for endpoint field descriptions.

After entering the storage endpoint information, click on **Save**.

### 3.2.1.2. Display endpoint

In the *List Storage Endpoints* window, click the magnifying glass button to open the *Show Storage Endpoint* window.
### Storage Endpoint Identifier
Unique numeric ID assigned to endpoint at creation time.

### Supporting Copiers
List of copiers configured to support this endpoint.

### Supported Endpoint Policies
List of policies that have this static endpoint defined as an import source in their policy.

See *Endpoint fields* section for other endpoint field descriptions.

From the *Show Storage Endpoint* screen the administrator can reset the password being used for the default credential by clicking *Reset Default Credentials*. 

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint Name</td>
<td>TreeFrogFS</td>
</tr>
<tr>
<td>Endpoint Description</td>
<td>Filesystem static endpoint</td>
</tr>
<tr>
<td>Storage Endpoint Type</td>
<td>static</td>
</tr>
<tr>
<td>Storage Endpoint Identifier</td>
<td>1</td>
</tr>
<tr>
<td>Endpoint is Enabled</td>
<td>True</td>
</tr>
<tr>
<td>Passwords Stored</td>
<td>True</td>
</tr>
<tr>
<td>Endpoint idle Connection Timeout</td>
<td>3600</td>
</tr>
<tr>
<td>Ping Limit Interval</td>
<td>None</td>
</tr>
<tr>
<td>Maximum Concurrent Requests</td>
<td>100</td>
</tr>
<tr>
<td>Maximum object size in MB</td>
<td>None</td>
</tr>
<tr>
<td>Supporting Copiers</td>
<td>[Treefrog_Copier]</td>
</tr>
<tr>
<td>Supported Import Policies</td>
<td>[Filesystem to Filesystem Policy, Filesystem_1+0_1tep, AWS-LS Policy, HPSS_1+0_1tep, policy1]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint URI scheme</td>
<td>filesystem</td>
</tr>
<tr>
<td>Host or IP</td>
<td>sara.dearlake.bm.com</td>
</tr>
<tr>
<td>Network Port</td>
<td>None</td>
</tr>
<tr>
<td>Path</td>
<td>/TreeFrogEndpoints/TreeFrogFS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Name</td>
<td>treedfrog</td>
</tr>
<tr>
<td>Expiration Date</td>
<td>2022-09-13</td>
</tr>
</tbody>
</table>
The administrator can validate that the combination of the Client Name and password being used for the default credential are valid by clicking Test Credential.

The administrator can also add endpoint project and user credentials by clicking Upload Credentials. See Upload endpoint project or user credentials section for more information.

### 3.2.1.3. Display endpoint project credentials

Select List Project Credentials to display a list of project credentials associated with the endpoint:

Click the magnifying glass button next to the credential you would like to display.
From the *Show Project Credential* screen the administrator can validate that the combination of the **Client Name** and password being used for the project credential are valid by clicking **Test Project Credential**. See *Endpoint project credential fields* section for field descriptions.

The administrator can reset the password being used for the project credential by clicking **Reset Project Credential Secret**.

### 3.2.1.4. Display endpoint user credentials

Select **List User Credentials** to display a list of user credentials associated with the endpoint:

Click the magnifying glass button next to the credential you would like to display.
From the Show User Credential screen the administrator can validate that the combination of the Client Name and password being used for the user credential are valid by clicking Test User Credential. See Endpoint user credential fields section for field descriptions.

The administrator can reset the password being used for the user credential by clicking Reset User Credential Secret.

3.2.1.5. Edit endpoint

In the List Storage Endpoints window, click the notepad-and-pen button to open the Edit Storage Endpoint window.
## Configuring HSB instance components

### Edit Storage Endpoint

**Detail** List Project Credentials  List User Credentials

**General**

<table>
<thead>
<tr>
<th>Endpoint Name</th>
<th>F8_repo10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endpoint Description</td>
<td>Filesystem repo10 endpoint</td>
</tr>
<tr>
<td>Type</td>
<td>repository</td>
</tr>
<tr>
<td>Endpoint is Enabled</td>
<td>On</td>
</tr>
<tr>
<td>Passwords Stored</td>
<td>On</td>
</tr>
</tbody>
</table>

**Endpoint Idle Connection Timeout**

<table>
<thead>
<tr>
<th>Endpoint Idle Connection Timeout</th>
<th>Endpoint Idle Connection Timeout</th>
</tr>
</thead>
</table>

**Ping Limit Interval**

<table>
<thead>
<tr>
<th>Ping Limit Interval</th>
<th>Minimum seconds between endpoint ping. Use -1 to disable pings and use 0 for the system default ping interval (30 seconds)</th>
</tr>
</thead>
</table>

**Maximum Concurrent Requests**

<table>
<thead>
<tr>
<th>Maximum Concurrent Requests</th>
<th>100</th>
</tr>
</thead>
</table>

**Maximum object size in MB**

<table>
<thead>
<tr>
<th>Maximum object size in MB</th>
<th>Maximum object size in MB to store on this endpoint</th>
</tr>
</thead>
</table>

### Location

**Protocol**

<table>
<thead>
<tr>
<th>Protocol</th>
<th>filesystem</th>
</tr>
</thead>
</table>

**Host or IP**

<table>
<thead>
<tr>
<th>Host or IP</th>
<th>sara.clearlake.brn.com</th>
</tr>
</thead>
</table>

**Network Port**

<table>
<thead>
<tr>
<th>Network Port</th>
<th></th>
</tr>
</thead>
</table>

**Path**

<table>
<thead>
<tr>
<th>Path</th>
<th>/hsbepingEndpoints/repodir10/</th>
</tr>
</thead>
</table>

**Default Credential**

<table>
<thead>
<tr>
<th>Client Name</th>
<th>freetng</th>
</tr>
</thead>
</table>

**Expiration Date**

<table>
<thead>
<tr>
<th>Expiration Date</th>
<th></th>
</tr>
</thead>
</table>
Changes cannot be made to the following fields while the endpoint contains dataset data: type, protocol, host, network port, path. In order to change these fields the datasets on the endpoint would have to be deleted.

Changes will take effect on the system when you click Save.

See Endpoint fields section for endpoint field descriptions.

### 3.2.1.6. Edit endpoint project credentials

In the Show Storage Endpoint Project Credential window, click the notepad-and-pen button to open the Edit Project Credential window.

Changes cannot be made to the following fields while the endpoint contains dataset data: type, protocol, host, network port, path. In order to change these fields the datasets on the endpoint would have to be deleted.

See Endpoint project credential fields section for field descriptions.

### 3.2.1.7. Edit endpoint user credentials

In the Show Storage Endpoint User Credential window, click the notepad-and-pen button to open the Edit User Credential window.
See *Endpoint user credential fields* section for field descriptions.

### 3.2.1.8. Delete endpoint

In the *List Storage Endpoints* window, select the eraser button next to the desired storage endpoint to delete it.

- All project and user credentials associated with the endpoint being deleted are also deleted.
- If the endpoint is referenced by a *storage policy copy* or a *copier*, it cannot be deleted.

### 3.2.1.9. Delete endpoint project or user credentials

In the *List Storage Endpoints* window, click the magnifying glass button next to the desired storage endpoint whose credentials you would like to delete. Next select either *List Project Credentials* or *List User Credentials* to bring up a list of credentials associated with endpoint. Use the checkboxes to select all or multiple credentials and from the *Actions* dropdown menu select *Delete Selected*. Select *OK* to the user confirmation message to delete the credentials.

### 3.2.1.10. Download endpoint project or user credentials

In the *List Storage Endpoints* window, click the magnifying glass button next to the desired storage endpoint whose credentials you would like to download. Next select either *List Project Credentials* or *List User Credentials* to bring up a list of credentials associated with endpoint. Use the checkboxes to select all or multiple credentials and from the *Actions* dropdown menu select *Download Selected*. Select *OK* to the browser message to save the credentials file to your browsers download directory.
3.2.1.11. Upload endpoint project or user credentials

In the *List Storage Endpoints* window, click the magnifying glass button next to the desired storage endpoint you would like to add credentials to. On the *Show Storage Endpoint* detail window click the *Upload Credentials* button to open the *Upload Endpoint Credentials* window.

Select **Browse** and locate and select the file containing the credentials you would like to upload. Select **Open**.
Configuring HSB instance components

Next select **Save** on the the *Upload Endpoint Credentials* window to complete the credential upload.

### 3.2.1.12. Endpoint fields

**Endpoint Name**
   Required. Administrator-defined name of the endpoint. The maximum size is 128 characters.

**Endpoint Description**
   Administrator-defined description of endpoint. The maximum size is 256 characters.

**Type**
   Required. The type of the endpoint; choices are:

   - **Static**: A persistent storage endpoint that is used to connect to and access a storage system (which is defined by the protocol; see the Protocol field description, below). It does not contain MDS objects. It is the source for MDS creation requests and the destination for MDS recall requests.

   - **Repository**: A storage endpoint that is used by HSB to store managed data sets.

**Enabled**
   When checked the endpoint is marked as enabled. When unchecked the endpoint is considered disabled and appears as such on the monitor screen.
Passwords Stored
When checked project and user credential passwords for this endpoint will be stored in the HSB database. This should be unchecked when the project and user credentials being used to access the endpoint are authenticated with a security domain.

This field is only applicable for filesystem endpoints. Non-filesystem endpoints should always have this field checked. In order to switch off or on passwords stored, no project or user credentials must exist on the endpoint. When changing the passwords stored value the endpoints project and user credentials can be downloaded (see Download endpoint project or user credentials section), the credentials can then be deleted (see Delete endpoint project or user credentials section), the passwords stored value can be changed, and then the project and user credentials can be added back by uploading the credential files (see Upload endpoint project or user credentials section).

Endpoint Idle Connection Timeout
Endpoint idle connection timeout in seconds. If blank, a default timeout is used based on the protocol. If the endpoint idle connection timeout is exceeded during the execution of a request, that request fails and is retried. Create and delete requests retry continuously. Recall requests are retried a set number of times based on the storage policy copy retry count for each storage policy copy. The timeout default is 600 seconds.

For HPSS, administrative functions like migrations or repacks can result in data not being accessible until the action completes. Therefore, for HPSS endpoints the timeout value should be long enough to allow those actions to complete. This value should be adjusted based on the number of tape drives available, the frequency of migrations/repacks, if staged data is read directly from tape, and if the same tape data is accessed frequently. Contact HPSS support if help is needed in determining the best timeout value.

Ping Limit Interval
The number of seconds the system will wait between attempts to determine if the endpoint is connected and available. This value can be set to "-1" to disable any ping activity, or to "0" to use the default ping interval (30 seconds). This value must be supplied for cloud endpoints (for example, AWS-S3). This value allows for control of any access costs that might be associated with contacting the storage system underlying the endpoint.

Max Requests
The maximum number of requests that can be running on this endpoint simultaneously. This defaults to "1000".

Max Object Size
The maximum size of a fragment chunk that can be stored on this endpoint in mebibytes. HSB automatically breaks larger objects into smaller fragment chunks to respect this value. If the endpoint is used in a storage policy copy that has parity generation enabled, the maximum object size must be a multiple of that copy’s parity block size.

Setting the Max Object Size on HPSS endpoints is not allowed.
Protocol
The transfer protocol of the endpoint. Select one of the options below that best describes your endpoint.

- filesystem: Filesystem local to the copier.
- hpss: High Performance Storage System.
- swift-keystone: Swift Keystone.
- aws-s3: Amazon Web Services Simple Storage Service.
- s3: Amazon Simple Storage Service compatible.
- sony-oai: Sony Optical Archive Incorporated.

Host
Required. The hostname or the IP address of the system hosting the endpoint.

Network Port
The port used by the endpoint. The port is required for endpoints using the HPSS protocol; however, it must not be specified for endpoints using the filesystem protocol. For all other protocols besides those two, the port is optional.

Path
Required. The absolute path relative to this endpoint and it is the base path of the endpoint. The supplied Client Name must have permissions to access this path.

Filesystem protocol endpoints must use a unique combination of host, network port, and path. Sub paths of other endpoints using the same host and network port are not allowed.

Endpoints using non-filesystem protocols can have duplicates of host, network port, and path.

Endpoints using the HPSS protocol cannot have endpoints defined with paths that are subpaths of other endpoints using the same protocol, host, and network port.

3.2.1.13. Endpoint default credential fields

Client Name
Required. The default credential client name to use on the endpoint. The client name is the username that will be used to access and save data on the endpoint. At a minimum, this client name needs to be able to access the top level of the endpoint’s path. If it cannot, then the HSB system will not be able to status the endpoint.

Secret
Required for non-filesystem endpoints. Secret (password) verifying credential.

Confirm Secret
Required for non-filesystem endpoints. Secret (password) verifying credential that must match the secret above.
Configuring HSB instance components

**Cred Expiration**

The date the credential expires. If left blank, the credential does not expire.

⚠️ This credential is used as an identity for querying the existence, accessibility and capacity of the associated endpoint. An identity with minimal permissions (metadata read permissions only) on the default endpoint credential should be selected for this purpose. This identity must exist on the endpoint using the specific authentication mechanism of that endpoint.

The default endpoint credential is used when determining if the endpoint can be marked ONLINE.

There is a hierarchy associated with the endpoint credential selection. When a request requires access to an endpoint, the HSB server uses a defined user endpoint credential defined for the current HSB system’s user identity and the associated endpoint. If no credential is found, it uses a project endpoint credential defined for the associated project. At least one of these credentials must be provided for the user to create, read and delete dataset fragments on the endpoint. It is best practice to not use the same identity multiple times for the HSB credentials on an endpoint.

The **Client Name** is not an HSB user.

For example:

- When configuring an HPSS endpoint, specify the HPSS username and password HSB will use when accessing the HPSS endpoint.
- When configuring an S3 endpoint, specify the S3 account username and password HSB will use when accessing the S3 endpoint.
- When configuring a filesystem endpoint, use the OS username and password HSB will use when accessing the filesystem endpoint.

⚠️ Changing permissions on the endpoint can affect HSB access. For example: For an S3 endpoint, if the path is set to a shared bucket owned by S3 user A, and the HSB endpoint client name is set to S3 user B. If S3 user A revokes the shared access to the bucket then the HSB endpoint will not be accessible until the access is restored.

### 3.2.1.14. Endpoint project credential fields

**Project** Required. The name of the project associated with the credential.

**Client Name** Required. The credential client name is the username on the endpoint system that will be used to access and save data on the endpoint, it is *not* an HSB user name. The client name is *not* an HSB user credential.

HSB will only have access to the data on the endpoint for which the client name has permission. If there is data in the endpoint’s path that the client name does not have
Configuring HSB instance components

There is a hierarchy associated with the endpoint credential selection. When a request requires access to an endpoint, the HSB server uses a defined user endpoint credential defined for the current HSB system’s user identity and the associated endpoint. If no credential is found, it uses a project endpoint credential defined for the associated project. At least one of these credentials must be provided for HSB to create, read and delete dataset data on the endpoint. It is best practice to not use the same identity multiple times for the HSB credentials on an endpoint.

The client name is not an HSB user.

For example:

- When configuring an HPSS endpoint, specify the HPSS username and password HSB will use when accessing the HPSS endpoint.

- When configuring an S3 endpoint, specify the S3 account username and password HSB will use when accessing the S3 endpoint.

- When configuring a filesystem endpoint, use the OS username and password HSB will use when accessing the filesystem endpoint.

Changing permissions on the endpoint can affect HSB access. For example: For an S3 endpoint, if the path is set to a shared bucket owned by S3 user A, and the HSB endpoint client name is set to S3 user B. If S3 user A revokes the shared access to the bucket then the HSB endpoint will not be accessible until the access is restored.

Expiration Date

The date the credential expires. If left blank, the credential does not expire.

### 3.2.1.15. Endpoint user credential fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>Required. The HSB user associated with the credential.</td>
</tr>
<tr>
<td>Client Name</td>
<td>Required. The credential client name is the username on the endpoint system that will be used to access and save data on the endpoint, it is <em>not</em> an HSB user name.</td>
</tr>
</tbody>
</table>

HSB will only have access to the data on the endpoint for which the client name has permission. If there is data in the endpoint’s path that the client name does not have permissions to use (permissions as defined by the endpoint not HSB), then HSB will not be able to access the data.

Expiration Date

The date the credential expires. If left blank, the credential does not expire.

There is a hierarchy associated with the endpoint credential selection. When a request requires access to an endpoint, the HSB server uses a defined user endpoint credential defined for the current HSB system’s user identity and the associated endpoint. If no
Configuring HSB instance components

credential is found, it uses a project endpoint credential defined for the associated project. At least one of these credentials must be provided for the user to create, read and delete dataset fragments on the endpoint. It is best practice to not use the same identity multiple times for the HSB credentials on an endpoint.

The client name is not an HSB user.

For example:

• When configuring an HPSS endpoint, specify the HPSS username and password HSB will use when accessing the HPSS endpoint.

• When configuring an S3 endpoint, specify the S3 account username and password HSB will use when accessing the S3 endpoint.

• When configuring a filesystem endpoint, use the OS username and password HSB will use when accessing the filesystem endpoint.

Changing permissions on the endpoint can affect HSB access. For example: For an S3 endpoint, if the path is set to a shared bucket owned by S3 user A, and the HSB endpoint client name is set to S3 user B. If S3 user A revokes the shared access to the bucket then the HSB endpoint will not be accessible until the access is restored.

3.2.1.16. Endpoint specific configuration

Some of the supported endpoint types provide additional configuration attributes.

HPSS endpoint attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPSS Class of Service</td>
<td>The HPSS COS identifier to use for objects written to this endpoint. The COS must be defined in HPSS before you can use this endpoint.</td>
</tr>
<tr>
<td>HPSS File Family</td>
<td>The HPSS file family identifier to use for objects written to this endpoint. This allows objects stored on the same endpoint to be co-located on the same HPSS tape cartridges. See the HPSS tape-only class of service considerations section for more information about when this attribute is needed.</td>
</tr>
</tbody>
</table>
HPSS Media Section Bytes

This attribute applies to endpoints that are assigned to a HPSS COS that is associated with a HPSS tape-only storage hierarchy. It specifies the maximum number of bytes HPSS (and associated tape drives) buffers for objects written to the endpoint. This attribute should be set to the product of the tape block size and the number of tape blocks in an HPSS tape section for the tape storage class associated with the specified HPSS COS. If the endpoint is used for a copy level with parity generation enabled, this value must also be a multiple of the parity block size.

This attribute is required for all tape-only HPSS endpoints.

The following example shows an HPSS storage class configuration screen. To find the corresponding HPSS Media Section Bytes value, multiply together the values for the Media Block Size and Blocks Between Tape Marks settings (highlighted in the figure below).
The **HPSS Media Section Bytes** for the example storage class is:

\[262144 \times 28800 = 7549747200\]

**HPSS tape-only class of service considerations**

When configuring endpoints for HPSS tape-only classes of service, great care must be taken to specify the required HPSS endpoint attribute values correctly. Failure to do this can result in failures when the endpoint is used to store data.

The **HPSS Media Section Bytes** attribute must always be specified for tape-only classes of service. See the guidance above to find the correct value.

HPSS tape-only storage policies with a default fragmentation of greater than 1 require a number of HPSS tape drives greater than or equal to the default fragmentation. If parity fragments are to be generated, then there must be an additional tape drive for each parity fragment generated. Failure to ensure that adequate drives are available can cause create or recall requests to deadlock and eventually fail.

If the endpoint is to be used in a storage policy with a default fragmentation of greater than 1, then one of the following guidelines should be followed:

- Each endpoint associated with the storage policy must specify a unique value for the **HPSS Class of Service** attribute.

  **Or**

- Each endpoint associated with the storage policy with the same HPSS COS must specify a unique value for the **HPSS File Family** attribute.

Since tape drives are sequential I/O devices, requests utilizing HPSS tape-only storage policies could be delayed due to other activity competing for tape drive access. The **Endpoint Idle Connection Timeout** attribute should be set high enough to account for the delays that can occur when there is contention for the tape drives.

Each HPSS COS and HPSS file family combination will result in an additional pool of HPSS tape resources. Care should be taken when configuring the endpoints so that more pools aren’t allocated than there will be available tape cartridges.

Each HPSS Tape-Only COS supporting an endpoint should also have the **Hash Algorithm** set to use **MD-5** (see the image below). This allows for the generation of a hash for each fragment stored on the endpoint. This hash is validated against a hash value generated at the copier before declaring dataset creation successful.
HPSS PFTP configuration

To support HPSS endpoints, the HPSS PFTP server must be configured and able to accept FTP and parallel FTP requests from the copiers supporting the HPSS endpoints. The copiers maintain a limited size pool of active PFTP connections. In order for HPSS endpoint access to function properly, the `xinetd` supporting the PFTP server must allow for a copier to connect based on the maximum connection pool size (the default is 40 connections). The following `xinetd` configuration setting should be updated based on the HPSS endpoint configuration.

- **per_source** - The maximum number of connections per source. This value should be set to 40, to reflect that copiers will pool up to 40 connections.

- **cps** - Allows you to limit the rate of incoming connections to allow per second. It accepts two arguments. The first argument is the number of connections per second to handle. This value should be "40". The second argument is the number of seconds to wait before re-enabling the service after it has been disabled. This value should be "10".

- **instances** - The maximum number of PFTP instances system-wide. This number should be the number of copiers multiplied by 40.

The following is an example of `xinetd` configurations in support of an HPSS endpoint:

```bash
--- /etc/xinetd.conf.bak 2020-03-11 11:26:18.980424506 -0500
+++ /etc/xinetd.conf 2020-03-11 08:33:08.133634364 -0500
@@ -23,9 +23,9 @@
    no_access       =
    only_from       =
    max_load        = 0
-   cps             = 50 10
-   instances       = 50
-   per_source      = 10
+   cps             = 40 10
+   instances       = 200
+   per_source      = 40

# Address and networking defaults
#
```

See `xinetd.conf(5)` for more detail about the `xinetd` configuration settings.

If your usage profile requires more than 40 PFTP connections, this value can be increased by adjusting the value of the "hpss-pftp-max-connections" setting in the `copier.conf` file. This setting controls the maximum HPSS PFTP connection pool size for a HSB copier instance.

The following is an example of a configuration that supports up to 80 HPSS PFTP connections:

```
copier-id=1
copier-port=8181
copier-logging-level=INFO
server-uri=https://tallulah.clearlake.ibm.com:443
server-cn=tallulah.clearlake.ibm.com
ping-interval-milliseconds=5000
ssl-key-store=/etc/treefrog/ssl/copierkeystore.jks
```
An additional PFTP configuration consideration involves reuse of data transfer connections between the HPSS movers and the HSB copiers. The default behavior for PFTP is to not request that the HPSS movers reuse connections. If any of your storage policies are configured to support parity, then it is important for PFTP to remain configured to disallow the reuse of data transfer connections. To do this make sure that the PortRange option (PFTP daemon configuration stanza) in the `HPSS.conf` file remain commented out. The following is an example of this configuration:

```bash
$ cat /var/hpss/etc/HPSS.conf

PFTP Daemon = {
    # Allow the Daemon to take Core Dumps
    Allow Core Files
    ...
    # The Port Range to be used for connections to other machines
    # Useful if Client is used to cross Network Filters
    # (The older ncacn_ip_tcp[10100-12100] format is supported for now)
    # Default = 10100-65535
    ; PortRange = 10100-12100
    ...
}
```

See the *HPSS User's Guide* for more information on PFTP configuration.
3.2.2. Copier configuration

3.2.2.1. Add copier

Select Configure > Copiers in the HSB administrative GUI:

The *List Copiers* window is displayed:

In the *List Copiers* window, select the plus button to open the *Add Copiers* window.
Enter the copier information and then click **Save**. See the description of the **Copier fields** below, noting any specific rules that may apply when configuring the copier. Once the copier information has been saved, additional steps are needed to download and install the copier certificate bundle. See section Download copier certificate bundle.

### 3.2.2.2. Copier fields

**Copier Name**

The name of the copier must be unique and has a maximum size of 128 characters.

**Execute Host Name**

The hostname or the IP address where the copier runs. Note that the combination of "Execute Host Name" and "Port Number" must be unique for each copier entry.

**Port Number**

The network port number that the copier listens on.

**Copier is Enabled**

When checked the copier is enabled. When unchecked the copier is disabled and appears in the DISABLED state on the list and on monitor copier screen.
Configuring HSB
instance components

Logging Level
Controls which message types are sent to the copier log for this copier.

Maximum Concurrent Requests
The maximum number of concurrent requests that can be on this copier at one time. This defaults to 1000.

Retry Interval Seconds
The number of seconds to wait between retrying requests.

Endpoints
Select the endpoints that this copier is responsible for handling data to or from.

Key Store Password
This is the password used to protect the Java key store for the copier.

Confirm Password
This password should match the value entered in the "Key Store Password" field.

Certificate Valid Days
The number of days you wish the copier SSL certificate to remain valid.

3.2.2.3. Display copier

In the List Copiers window, click the magnifying glass button next to the copier record to be displayed to open the Show Copier window.
Configuring HSB instance components

<table>
<thead>
<tr>
<th>Copier Identifier</th>
<th>Unique numeric ID assigned to copier at creation time.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supported Endpoints</td>
<td>Endpoints this copier can access.</td>
</tr>
<tr>
<td>Certificate Serial Number</td>
<td>Unique ID assigned to SSL Certificate at creation time.</td>
</tr>
<tr>
<td>Certificate Invalid Before Date</td>
<td>Start date for valid certificate.</td>
</tr>
<tr>
<td>Certificate Invalid After Date</td>
<td>End date for valid certificate.</td>
</tr>
</tbody>
</table>

See *Copier fields* section for other copier field descriptions.

**Download copier certificate bundle**

In order to complete the copier configuration, download the copier certificate bundle and install it on the copier machine. Once downloaded the bundle is placed in the web browser’s download directory. This must be transferred to the copier machine, then executed as root user. This bundle generation imports the HSB server certificate into a copier truststore file `copiertruststore.jks` and creates a copier SSL certificate and imports it into a copier keystore file `/etc/treefrog/ssl/copierkeystore.jks`. Installing the bundle extracts the truststore and keystore files to the location indicated by the `TREEFROG_ETC_DIR` environment variable. If the environment variable doesn’t exist, it defaults to the directory `/etc/treefrog/ssl`.

To download and install the certificate bundle, select the "Download Copier Certificate Bundle" button on the *Show Copier* window. This downloads the "copier_cfg_installer" script using your web browser. Execute the script on the copier machine, simply hit return key to accept the default settings to install the truststore, keystore, and other configurations.

Note: The certificate bundle is unique for each copier added to the HSB. It cannot be shared among the different copiers.

```bash
> sh ./copier_cfg_installer
Self Extracting Installer
installer
copiersettings.inp
copierkeystore.jks
copiertruststore.jks
```

**Reset KeyStore Password**
Clicking the **Reset KeyStore Password** button from the *Show Copier* window allows you to change the current copier keystore password.

**Generate New Certificate**

Clicking the **Generate New Certificate** button from the *Show Copier* window generates a new copier certificate that needs to be downloaded and installed on the Copier. Click *OK* on the confirmation dialog box that appears. To download and install the certificate on the copier, follow the directions in the section Download Copier Certificate Bundle above.

**3.2.2.4. Edit copier**

In the *List Copiers* window, select the notepad-and-pen button next to the copier record to open the *Edit Copiers* window.
See Copier fields section for copier field descriptions.

### 3.2.2.5. Delete copier

In the List Copiers window, select the eraser button next to a copier record to delete it.

A copier cannot be deleted if the storage endpoints referenced are defined in a storage policy copy and no other copiers are configured to access the same storage endpoints.

### 3.2.2.6. Copier logging

The HSB copier stores its current log file, `copier.log`, as well as archived copier logs under the directory `/var/log/treefrog`. Copier log file rotation is handled by the `logrotate` utility and the logrotate configuration file for the copier located under `/etc/logrotate.d/treefrog-copier`. Below is the copier configuration file that is installed as part of the copier RPM installation:

```bash
compress

/var/log/treefrog/copier.log {  
  weekly
  missingok
}
Configuring HSB instance components

### 3.2.2.7. Controlling copier java heap size allocation

The Java heap size allocation is controlled via the `/usr/bin/treefrog_copier` script file. The default initial heap size setting (`-Xms`) is 10 GiB and the maximum heap size setting (`-Xmx`) is 26 GiB. These settings should be adjusted based on the amount of memory installed on the system running an instance of the copier. Contact HPSS support if you would like assistance with determining the appropriate values for these settings.

### 3.2.2.8. Controlling the copier thread pool size

The default maximum number of threads supported by the copier is 200. For copiers that experience heavy workloads, the maximum number can be increased. The maximum number of threads can be controlled by the `copier-max-threads` configuration parameter.

This parameter is specified in the `/etc/treefrog/copier.conf` file on the copier host machine. The valid range for this value is 200-2000. In the following example, the maximum number of threads is set to 400:

```
copier-id=1
copier-port=8181
copier-logging-level=INFO
server-uri=https://tallulah.clearlake.ibm.com:443
server-cn=tallulah.clearlake.ibm.com
ping-interval-milliseconds=5000
ssl-key-store=/etc/treefrog/ssl/copierkeystore.jks
ssl-key-store-password=treefrog
ssl-trust-store=/etc/treefrog/ssl/copiertruststore.jks
ssl-trust-store-password=treefrog
copier-max-threads=400
```

Server log messages found in `/var/log/treefrog/treefrog.log` similar to the following could indicate an insufficient copier maximum thread pool setting:

```
Aug 18 07:56:21 myhost treefrog[3569]: <ERROR> resource_thread._status_endpoint:346
Communications error: ConnectionPoolWrapper(host='myhost.mycompany.com', port=8181):
Read timed out. (read timeout=30), copier 61:https://myhost.mycompany.com:8181
```

Aug 18 05:21:04 myhost treefrog[21768]: <ERROR> resource_thread._status_copier:218
Communications error: ConnectionPoolWrapper(host='myhost.mycompany.com', port=8181):
Read timed out. (read timeout=2), copier 1:https://myhost.mycompany.com:8181
```

### 3.2.2.9. Controlling copier HPSS PFTP data movement settings

For data transfers to and from HPSS endpoints, the copier utilizes connections directly to the HPSS movers. To optimize data transfer performance, the socket buffer sizes used for these connections

```
compress
copytruncate
maxsize 100M
minsize 100K
postrotate
/bin/kill -HUP `cat /var/run/syslogd.pid 2> /dev/null` 2> /dev/null || true
endscript
```

See the man page for `logrotate` for details on the parameters you can set.
may need to be adjusted from the system defaults. The `hpss-pftp-sockbuf-size` parameter is used to set the socket buffer size used by the copier for data connections. The parameter argument is specified in bytes, with 4 KiB (4096) being the minimum allowed value. If this parameter is not specified the system default socket buffer sizes are used.

It has also been observed that HPSS transfer network performance can be optimized by adjusting the size of the socket writes used during data transmission. The size of the socket writes can be controlled with the use of the `hpss-pftp-iobuf-size` parameter. The parameter argument is specified in bytes, with the default being 32 KiB.

These parameters are specified in the `/etc/treefrog/copier.conf` file on the copier host machine. In the following example, the socket buffer size is set to 1 MiB and the I/O buffer size is set to 128 KiB:

```
copier-id=1
copier-port=8181
copier-logging-level=INFO
server-uri=https://tallulah.clearlake.ibm.com:443
server-cn=tallulah.clearlake.ibm.com
ping-interval-milliseconds=5000
ssl-key-store=/etc/treefrog/ssl/copierkeystore.jks
ssl-key-store-password=treefrog
ssl-trust-store=/etc/treefrog/ssl/copiertruststore.jks
ssl-trust-store-password=treefrog
hpss-pftp-sockbuf-size=1048576
hpss-pftp-iobuf-size=32768
```

Care should be taken when setting socket buffer sizes for the HPSS connections. The systems network memory parameters must allow for the specified socket buffer size. On Linux, the `net.core.rmem_max` and `net.core.wmem_max` kernel parameters limit user socket buffer allocation size. If the requested size cannot be allocated the copier will output warning messages.
3.2.3. Storage policy configuration

3.2.3.1. Create storage policy

Select Configure > Policies in the HSB administrative GUI:

The Storage Policies window is displayed:

From the Storage Policies window, select the plus button to open the Add Storage Policy window.
Configuring HSB instance components

After you have entered the storage policy information, click on **Save**.

3.2.3.2. Storage policy fields

**Policy Name**
Required. Administrator defined name for the storage policy. The maximum size is 256 characters.

**Policy Description**
Administrator defined description for the storage policy. The maximum size is 256 characters.

**Minimum Dataset Size**
The minimum size of the dataset in MiB that is allowed to be stored on this storage policy.

**Maximum Dataset Size**
The maximum size of the dataset in MiB that is allowed to be stored on this storage policy.

**Access Permissions**
The permissions to use for this storage policy. Choose from one the following:

- **Disabled**: Policy is disabled and can not be read from or written to.
- **Read/Write**: Policy can be read from and written to. Reading is used for recall requests. Writing is used for create and delete requests.
- **Read**: Policy is read only. Only recall requests can be run on this storage policy.

**Preferred Recall Copy**
The preferred copy level to use for recalls.

**Copies**
Storage policy copy levels. If the policy has multiple copies, then the data is copied to all copy levels on create. On delete, the data is deleted from all copy levels. On recall, the preferred copy
level is used to retrieve the data. If the preferred copy level cannot recall the data, then the other enabled copy levels are attempted starting with the first copy level in the list. Copy levels are ordered by creation date currently. If no copy levels can recall the data, the request is suspended.

3.2.3.3. Storage policy copy fields

**Copy Name**
Required. Administrator defined name for the storage policy copy. The maximum size is 256 characters.

**Copy Description**
Administrator defined description for the storage policy copy. The maximum size is 256 characters.

**Default Fragmentation**
This is the default number of fragments for datasets created for this copy. This value must be greater than the parity count value. If \((\text{dataset size})/\text{Minimum Fragment Size}\) (see below) is greater than or equal to this value, then the copy consists of default fragmentation number of fragments. If the \((\text{dataset size})/\text{minimum fragment size}\) is less than the default fragmentation and parity count is zero, then the dataset consists of \((\text{dataset size})/\text{minimum fragment size}\) number of fragments. If the parity count greater than zero, then that value plus 1 is the lower bound for the number of fragments created for this copy.

**Minimum Fragment Size**
This is the smallest fragment (in bytes) allowed when storing datasets for this copy.

A fragment can be smaller than the minimum fragment size if the dataset is smaller than the minimum or if the last fragment is smaller than the minimum size.

**Recall Request Priority**
Default is 0. The valid value range is 0-9, where 0 is the lowest priority and 9 is the highest. This value represents a system level recall priority for recall and verification requests for datasets stored with this storage policy.

**Parity Count**
This is the number of parity fragments to generate for datasets created for this copy. The default for this is zero. This value must be less than the default fragmentation value.

**Parity Block Size**
The block size used for generation of parity for copies with a parity count Greater than zero. Default is 4MiB. If the endpoints associated with the copy level has max object size specified, then the maximum object size must be a multiple of this value. If the endpoints associated with the copy level are HPSS endpoints and have HPSS media section bytes specified, then the section bytes must be a multiple of this value.

**Recall Retry Count**
This is the number of recall retries to be performed from this copy of the dataset before moving to the next copy. The default value for this is 3. Recall retries start at the preferred recall copy (storage policy), then retry from the next copy level. If all available copies are retried unsuccessfully, then the request fails.
Enable for Reads
Default is Enabled. Enables reading for this copy level of the storage policy.

Endpoints
Storage endpoints associated with this copy level of the storage policy.

All endpoints selected for a storage policy copy must be defined with the same protocol and endpoint URI scheme.

Object Hashing Enabled
Default is Enabled. Enables creation of a hash for each file or object included in a dataset.

Chunk Hashing Enabled
Default is Enabled. Enables creation of a hash for each fragment chunk created with the dataset.

3.2.3.4. Storage policy import source fields

Source Endpoint
Required. A static endpoint for which to add the ability to import a dataset from an archive file.

Import Types
Required. A list of archive formats that can be imported for the import source. Currently supported types are htar and tar.

Import Management Type
Required. Indicates how the imported files will be managed. Currently the only management type available is user. This type has the following characteristics:

1. The archive file owner is responsible for managing the data.
   a. If the file is deleted before the HSB dataset is deleted then the data will not be accessible using HSB.
   b. If the file permissions change, HSB may not be able to access the data.
   c. When the HSB dataset is deleted the archive file will not be deleted.

2. A directory for the dataset will be created under the HSB project. It will contain a HSB magic file and the dataset catalog.

3. The original archive file will remain with the original owner and group permissions.

Can Access Archive
Selecting this indicates that an import using this source may read the archive, as opposed to reading the index file, during the import. For tar import types selecting this is necessary. For htar import types the normal behavior is to read the archive contents from the HTAR index file, which is typically stored on fast random access storage.

Import suffix overrides fields
Import Type

Required. The archive type to associate with the Archive File Suffixes. Currently supported types are `htar` and `tar`. The default suffix for HTAR (index) archive files is `idx` and the default for TAR archive files is `tar`. If you require other or additional suffixes to be identified as an archive file, then you can add overrides.

Archive File Suffixes

Required. A comma delimited list of all the file suffixes associated with an Import Type.

3.2.3.5. Display storage policy

In the List Storage Policies window, click the magnifying glass button next to the desired policy to open the Show Storage Policy panel.

Display storage policy copy

In the Storage Policy Copy window, click the magnifying glass button next to the desired policy copy to open the Storage Policy Copy window.
See Storage policy copy fields section for storage policy copy field descriptions.

**Display import source**

In the Show Storage Policy window, click the magnifying glass button next to the desired import source to open the Import Source window.

**Policy Name**  The policy name the import source is associated with.

See Storage policy import source fields section for storage policy import source field descriptions.

**Display import suffix overrides**

From the Import Source window select List Import Suffix Overrides to display a list of the file suffixes associated with each import type.
Configuring HSB instance components

See Import suffix overrides fields section for field descriptions.

**Delete storage policy copy**

In the Show Storage Policy window, select the eraser button next to the desired policy copy to delete it.

If a project is referencing the storage policy, then the deletion of the storage policy copy is prevented.

**Delete import source**

In the Show Storage Policy window, select the eraser button next to the desired import source to delete it.

### 3.2.3.6. Edit storage policy

In the Storage Policies window, click the notepad-and-pen button next to the desired policy to open the Edit Storage Policy window.
See *Storage policy fields* section for storage policy field descriptions.

- See *Display storage policy copy* above on how to display storage policy copies.
- See *Edit storage policy copy* on how to edit storage policy copies.
- See *Edit import source* on how to edit an import source.

**3.2.3.7. Edit storage policy copy**

In the *Show Storage Policy* window, click the notepad-and-pen button next to the desired policy copy to open the *Edit Storage Policy Copy* window.
3.2.3.8. Edit import source

In the *Show Storage Policy* window, click the notepad-and-pen button next to the desired import source to open the *Edit Import Source* window.

---

See *Storage policy copy fields* section for storage policy copy field descriptions.
Configuring HSB instance components

See Storage policy import source fields section for storage policy import source field descriptions.

From the Import Source window select List Import Suffix Overrides to display a list of the file suffixes associated with each import type.

If a record already exists for the import type select the eraser button next to the import type and delete it.

Now select the plus button to open the Add Import Suffix Overrides window.

Select a Import Type from the dropdown list and in the Archive File Suffixes field identify the file suffixes you want to associate with that import type.

See Import suffix overrides fields section for field descriptions.

3.2.3.9. Delete storage policy

In the List Storage Policies window, select the eraser button next to the desired policy to delete it.

- See Delete storage policy copy on how to delete storage policy copies.
- See Delete import source on how to delete an import source.
- Deleting the storage policy is prevented if there are any projects referencing it or if any of the storage policy copies have storage endpoints defined.
3.2.4. Project configuration

3.2.4.1. Create project

Select **Configure > Projects** in the HSB administrative GUI:

The *List Projects* window is displayed:

In the *Storage Projects* window, select the plus button to open the *Add Project* window.
See Project fields section for project field descriptions.

After you have entered the project information, click on Save.

3.2.4.2. Project fields

**Project Name** - Required. Administrator defined name of the project. Maximum size is 128 characters.

**Project Description** - Required. Administrator defined description of the project. Maximum size is 128 characters.

**Creation Priority** - The relative scheduling priority for datasets created in this project. The valid range is 0 (lowest) to 9 (highest).

**Event Record Retention Period** - The number of days that the system retains event messages associated with this project. The default value is 60 days.

**Policies** - List of storage policies that can be used when creating datasets for this project.

3.2.4.3. Display project

In the List Projects window, click the magnifying glass button next to the desired project to open the Show Projects window:

![Show Projects window](image)

See Project fields section for project field descriptions.

**List project operation limits**

In the Show Projects window, click on the Op Limits tab to open the Op Limits view:
See *Project operation fields* section for project operation field descriptions.

**Display project operation limit**

From the "Show Projects" Op Limits tab, click the magnifying glass button next to the desired operation limit to open the *Op Limits* window.

See *Project operation fields* section for project operation field descriptions.

**Edit project operation limit**

*Operation Limits* - represents the number of requests of a specific type that can execute simultaneously.

In the *Show Project Op Limits* window, click the notepad-and-pen button next to the desired operation limit to open the *Edit Op Limits* window.
Configuring HSB instance components

See Project operation fields section for project operation field descriptions.

**List project roles**

In the Show Projects window, select the "Project Roles" tab to open the Project Roles window:

Display project role

In the Project Roles window, click the magnifying glass button next to the desired project role to open the Project Role window.

See Project role fields section for project role field descriptions.

**Edit project role**

In the Show Projects window, click the notepad-and-pen button next to the desired project role to open the Edit Project Roles window.
See *Project role fields* section for project role field descriptions.

### Delete project role

In the *Show Projects* window, select the eraser button next to the desired project role to delete it.

![Project Roles](image)

The last project owner cannot be deleted when the project contains managed data sets.

#### 3.2.4.4. Project operation fields

**Operation Limits** - represents the number of requests of a specific type that can execute simultaneously associated with the project.

**Operation** - The type of the operation. Types listed below:

- **Create** - Create a managed data set.
- **Recall** - Recall a managed data set.
- **Verify** - Verify a managed data set.
- **Delete** - Delete a managed data set.

**Limit** - The number of requests of specific type that can be running on this project simultaneously. If blank, the limit is infinite. Recommended limits are heavily dependent on your hardware, workload and workflow. Contact your HSB support representative if recommendations are required for your specific configuration.

#### 3.2.4.5. Project role fields

**Project** - The project this role is associated with.

**User** - The user on the this role is associated with.

**Project role** - Defines the level of control a user has on the project and its datasets. The roles are listed below from least to most permissive, with each role being a superset of the previous.

- **Read** - Allows a user holding this role to query the contents and recall datasets stored as part of this project.
• **Update** - Allows a user holding this role to create new datasets as part of the project and update the user metadata for datasets stored as part of the project. Includes capabilities defined in the read role.

• **Delete** - Allows a user holding this role to delete datasets stored as part of the project. Includes capabilities defined in the update role.

• **Owner** - A user holding this role is considered to be a project owner and can add, modify, and delete roles for other users on the project (except other owners). Includes capabilities defined in the delete role.

### 3.2.4.6. Edit project

In the *List Projects* window, click the notepad-and-pen button next to the desired project to open the *Edit Project* window.

Click the **Save** button when finished.

See *Project fields* section for project field descriptions.

- See "Display project op limit" above on how to display project op limits.

- See "Edit project op limit" above on how to edit project op limits.

- See "Display project role" above on how to display project roles.

- See "Edit project role" above on how to edit project roles.

- See "Delete project role" above on how to delete project roles.

### 3.2.4.7. Delete project

In the *List Projects* window, select the eraser button next to the desired project to delete it.
A project cannot be deleted when it contains managed data sets.

If the project being deleted is the default project for any user, then that user’s default project is set to NULL.
3.2.5. System configuration

3.2.5.1. List system configuration

Select Configure > System in the HSB administrative GUI:

![List System Configuration window](image)

The List System Configuration window is displayed:

![Show System Configuration window](image)

3.2.5.2. Display system configuration

From the List System Configuration window, click the magnifying glass button next to the desired system configuration to open the Show System Configuration window.
Certificate Serial Number - Unique ID assigned to SSL Certificate at creation time.
Certificate Invalid Before Date - Start date for valid certificate.
Certificate Invalid After Date - End date for valid certificate.

See System fields section for system field descriptions.

3.2.5.3. Edit system configuration

From the List System Configuration window, select the notepad-and-pen button next to the desired system configuration to open the Edit System Configuration window.
Click the Save button when finished.

See System fields section for system field descriptions.

3.2.5.4. System configuration fields

System Distinguished Name - The domain name of the server the system is hosted on. The maximum size is 256 characters.

Endpoint Cred Check Interval - Time interval in seconds after which endpoint credentials are verified.

Maximum User Requests - The maximum total number of requests that can be in the system at one time.

Copier Key Lifetime - Time interval in minutes before renegotiating a shared key with the copier.

System Debug Logging Level - The log level used when capturing log messages. A more verbose log level also logs the lower levels. There are five log levels list below, in order from the most to the least verbose.

- **DEBUG4** - The most verbose level of logging. Logs entering and exiting functions.
- **DEBUG3** - Occasional trace logging of actions less verbose than DEBUG4.
• **DEBUG2** - Occasional trace logging of semi-significant actions such as a picking a resource to run request on.

• **DEBUG1** - Logs significant actions such as executing a request.

• **DEBUG** - Logs significant errors.

• **OFF** - Logs messages of INFO or higher. Will only log the most critical of errors.

**Password Expiration Interval** - Time interval in days that passwords remain valid.

**SSL certificate fields**

- **Certificate Serial Number** - The unique identifier created when the certificate was created.
- **Certificate Invalid Before Date** - The date the certificate became valid.
- **Certificate Invalid After Date** - The date the certificate expires.
- **Certificate Valid Days** - Required. Number of days the certificate is required.
Configuring HSB
instance components

3.3. Monitor menu

3.3.1. Events

3.3.1.1. List event messages

Select Monitor > Events in the HSB administrative GUI:

The List Event Messages window is displayed:

<table>
<thead>
<tr>
<th>Event Timestamp</th>
<th>Status</th>
<th>Request/Job Identifier</th>
<th>Message Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-01-29 17:25:49:776266</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 17:15:49:609333</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 17:05:49:329209</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 16:55:49:399964</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 16:45:49:250000</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 16:35:49:110020</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 16:25:49:893483</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 16:15:49:887200</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 16:05:49:729152</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 15:55:49:290428</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 15:45:49:182512</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 15:35:49:029790</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 15:25:47:683017</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 15:15:47:762165</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 15:05:47:648167</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
<tr>
<td>2020-01-29 14:55:47:503028</td>
<td>SUCCESS</td>
<td>None</td>
<td>Credential for user admin expired on endpoint fileSystem01</td>
</tr>
</tbody>
</table>

3.3.1.2. Display event message

From the List Event Messages window, click the magnifying glass button next to the desired message to open the Show Event Message window.
Event Timestamp - The moment in time this event took place. The timestamp is displayed as UTC time.

Message Identifier - A unique identifier for this message. It is generated when the event is generated.

Request/Job Identifier - If this event message was related to a request, then the request ID is here. Otherwise, it is "None".

Status - The completion status for the request. It can be one of the following:

- SUCCESS - The request completed successfully.
- FAILURE - The system was unable to complete the request and the number of failures exceeded the retry limit for the system.
- ABORTED - The request was aborted by the user or administrator.

Project Name - If this event message was related to a project, the project name is here. Otherwise, it is "None".

Dataset Name - If this event message was related to a dataset, the dataset name is here. Otherwise, it is "None".

Source Endpoint(s) - If this event message was related to transferring data, the source URIs for the data transfer is here. Otherwise, it is "None".

Destination Endpoint(s) - If this event message was related to transferring data, the destination URIs for the data transfer is here. Otherwise, it is "None".

Message Text - The event message text. It contains details relating to the event.
3.3.2. Requests

3.3.2.1. List requests

Select **Monitor > Requests** in the HSB administrative GUI:

![List Requests window](image)

The *List Requests* window is displayed:

![List Requests columns](image)

If you are viewing a multiple copy request, there is a plus sign to the left of the request ID. Clicking this expands the request and reveals the request’s children.

![Expanded request](image)

If you are viewing an expanded multiple copy request, there is a minus sign to the left of the request ID. Clicking this hides the child requests.

![Expanded request with minus sign](image)

The following columns are displayed:

**Id**

The request identifier. Generated when the request is created.

**Type**

The type of the request. It can be one of the following:

- **CREATE**: This request creates an MDS.
Configuring HSB instance components

- **DELETE**: This request deletes an MDS.
- **RECALL**: This request recalls an MDS from a repository endpoint to a static one.
- **ABORT**: This request aborts an existing scheduled request. **COMPLETE** requests and **ABORTED** requests cannot be aborted.

**State**
State of the request; this can be one of the following:

- **WAIT**: The request is waiting until it can allocate all of the resources it requires. In this state, no resources have been acquired.
- **HOLD**: The request holds some, but not all, of the resources it requires.
- **RUN**: The request holds all of the resources it requires and is executing on the copier resource.
- **SUSPENDED**: The request has stopped execution and is waiting to be resumed.
- **PENDING**: This parent request has completed but it is waiting for its child request(s) to complete.
- **COMPLETE**: The request has completed and is about to be deleted from the request queue.
- **ABORT_LOOMING**: The request is about to be aborted.
- **ABORT_PENDING**: The request is about to be aborted, pending the completion of associated delete request.
- **ABORTED**: The request has been aborted and is about to be deleted from the request queue.

**Project**
The name of the project that is running the request.

**Dataset**
The name of the dataset that the request is interacting with.

**Bytes Copied**
The number of bytes copied. Used for create and recall requests.

**Failures**
The number of failures the request has encountered since it was put in the queue.

**Priority (Project)**
The priority of the project. Priorities are integers between "0" and "9", inclusive where 0 is low priority and 9 is high priority. Requests are processed in priority order. Admin priority takes precedence over project priority when determining the priority order.

**Priority (Admin)**
The priority assigned to request by the administrator. Priorities are integers between "0" and "9", inclusive where 0 is low priority and 9 is high priority. Requests are processed in priority order. Admin priority takes precedence over project priority when determining the priority order.
Age
The amount of time that has passed since the request was added to the request queue. Displayed in MM:SS format.

Display request
To display additional request details, select the plus between the **Id** column and the **Type** column. The following request details are displayed:

**User**
The username of the user that created the request.

**Created**
The date the request was created. Displayed in MM/DD/YYYY HH:MM:SS (AM/PM) format.

**IOStarted**
The date the request started to transfer data with its assigned copier.

**IOUpdated**
The date the request last transferred data with its assigned copier.

**Policy**
The name of the storage policy associated with this request.

**CopyLevel**
The copy level used by this request. In the case of a recall request, this can change if the request fails over to a new copy level.

**Fragments**
The total number of fragments in the request.

**Parity**
The number of blocks to use for parity generation.

**AbortId**
The request ID of the target request that is to be aborted. This is only used for abort requests.

**Suspending**
Flag to mark if the request is in the process of suspending. If **true**, the request is in the process of suspending; otherwise, the flag is **false**.

**Resources**
List of resources (copiers, endpoints, and project operation limits) used by the request.

**Request actions**
The following actions can be performed from the *List Requests* window:

**Abort**
Aborts all selected requests of the **RECALL** or **CREATE** request type that are not in the **COMPLETE**, **ABORT_LOOMING**, **ABORT_PENDING**, or **ABORTED** state.
If an aborted request has a **Request type** of **RECALL**, the **State** changes to **ABORT_LOOMING** and then changes to **ABORTED**. An event is logged for the abort and the request is then deleted from the queue.

If an aborted request has a **Request type** of **CREATE**, additional delete requests are created as part of the abort process. These delete requests remove all data and metadata created by the aborted create requests. While this process is occurring the create request is in the **ABORT_PENDING** state. Once this process completes the create request changes to the **ABORTED** state. An event is logged for the abort and the request is then deleted from the queue.

**Suspend**

Suspends all selected requests that are not in **RUN**, **ABORT_LOOMING** or **ABORTED** state. The **State** changes to **SUSPENDED**.

**Resume**

Resumes all selected requests that are in **SUSPENDED** state.

**Select All**

Selects all requests.

**Deselect All**

Deselects all requests.
3.3.3. Copiers

3.3.3.1. List copiers

Select Monitor > Copiers in the HSB administrative GUI:

The List Copiers window is displayed:

3.3.3.2. Monitor copier controls

Four control buttons appear above the list of copiers. They allow the administrator to control the copier states.

A single copier may be selected with a left mouse click. Multiple copiers may be selected with a CTRL+Click.

When a copier is selected, the monitor view does not refresh.

Online - Set the selected copiers to the ONLINE state.

Offline - Set the selected copiers to the OFFLINE state.

Select All - Select all the copiers.

Deselect All - Deselect all selected copiers.

3.3.3.3. Monitor copier fields

Id - System defined numeric identifier for the copier.
If you are viewing a copier, there is a plus sign to the right of the copier ID. Clicking this expands the request and reveals the detailed copier information.

If you are viewing the detailed copier information, there is a minus sign to the right of the copier ID. Clicking this hides the detailed copier information.

**Name** - Administrator defined name of the copier.

**State** - The status of the copier. The possible states are:

- **ONLINE**: The copier is running normally.
- **OFFLINE**: The administrator has stopped the copier from the monitor screen.
- **DOWN**: The system encountered an error when accessing the copier. The copier tries to come back online without administrator intervention.
- **BROKEN**: The system encountered an error when accessing the copier that requires administrator intervention. The copier can’t recover on its own.
- **DISABLED**: The administrator has disabled the copier from the copier configuration screen.

If a create or delete request requires a copier that is **DOWN**, the request continuously retries. If a create or delete request requires a copier that is **BROKEN**, **DISABLED**, or **OFFLINE**, the request is suspended.

If a recall request requires a copier that is **DOWN**, **BROKEN**, **DISABLED**, or **OFFLINE**, the request fails over, if it can. If it cannot failover, the request is suspended.

**Host** - The host where the copier is running.

**Port** - The port number on the host where the copier is configured to run.

**Logging Level** - The current logging level of the copier. The possible logging levels are:

- **OFF**: The copier logging is turned off.
- **INFO**: Capture general information in the logs about state changes to the copier and information on the length of time it takes a create or recall dataset operation to run.
- **DEBUG**: Capture internal events that can be helpful in troubleshooting the copier operation.
- **TRACE**: Capture procedure entry and exit events during the copier operation.

**Last Failure** - The time of the last error on the copier.

**Usage Info** - The copier limits and number of running requests:

- **Limit** - The configured limit for the maximum number of requests that can be running on this copier simultaneously.

<table>
<thead>
<tr>
<th>Name</th>
<th>State</th>
<th>Host</th>
<th>Port</th>
<th>Logging Level</th>
<th>Last Failure</th>
<th>Usage Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treehog_Copier</td>
<td>ONLINE</td>
<td>sora.charbake.lbm.com</td>
<td>8181</td>
<td>TRACE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Configuring HSB
instance components

- **Current** - The current number of requests running on the copier.
3.3.4. Endpoints

3.3.4.1. List endpoints

Select Monitor > Endpoints in the HSB administrative GUI:

The List Endpoints window is displayed:

<table>
<thead>
<tr>
<th>On-Line</th>
<th>Off-Line</th>
<th>Select All</th>
<th>Deselect All</th>
<th>Last Failure</th>
<th>Usage Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>Name</td>
<td>State</td>
<td>URL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>enpd11</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk2/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>12</td>
<td>enpd12</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk2/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>13</td>
<td>enpd13</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk2/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>14</td>
<td>enpd14</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk2/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>15</td>
<td>enpd15</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk2/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>16</td>
<td>enpd16</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk2/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>1</td>
<td>enpd1</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>2</td>
<td>enpd2</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>3</td>
<td>enpd3</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>4</td>
<td>enpd4</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>5</td>
<td>enpd5</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>6</td>
<td>enpd6</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>7</td>
<td>enpd7</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>8</td>
<td>enpd8</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>9</td>
<td>enpd9</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
<tr>
<td>10</td>
<td>enpd10</td>
<td>DOWN</td>
<td>filesystem://icb1.dearake.bm.com/storage/disk1/treenorg...</td>
<td>1/29/2020, 11:30 AM</td>
<td>None 0</td>
</tr>
</tbody>
</table>

3.3.4.2. Monitor endpoint controls

Four control buttons appear above the list of endpoints. They allow the administrator to control the endpoint states.

A single endpoint may be selected with a left mouse click. Multiple endpoints may be selected with a CTRL+Click.

When an endpoint is selected, the monitor view does not refresh.

**Online** - Set the selected endpoints to the **ONLINE** state.

**Offline** - Set the selected endpoints to the **OFFLINE** state.
Configuring HSB instance components

The **OFFLINE** state will override the **BROKEN** and **DOWN** states. It will not affect the **DISABLED** state. Changing the copier state for an endpoint will not affect the **OFFLINE** state.

**Select All** - Select all the endpoints.

**Deselect All** - Deselect all selected endpoints.

### 3.3.4.3. Monitor endpoint fields

**Endpoint Id** - System defined numeric identifier for the endpoint.

| 1 | Skylnet Repository | ONLINE | filesystem:/skynet/dealake/bm.com:8181/TreeFrogEndpoints | 27/07/2020, 13:52:38 PM | None | 0 |

If you are viewing an endpoint, there is a plus sign to the right of the endpoint ID. Clicking this expands the request and reveals the detailed endpoint information.

If you are viewing the detailed endpoint information, there is a minus sign to the right of the endpoint ID. Clicking this hides the detailed endpoint information.

**Endpoint Name** - Administrator-defined name of the endpoint.

**State** - The status of the endpoint. The possible states are:

- **ONLINE**: The endpoint is running normally.
- **UNUSED**: The endpoint is not assigned to a copier.
- **DEGRADED**: The endpoint is running but has issues on some of the configured copiers.
- **OFFLINE**: The administrator has stopped the endpoint from the monitor screen.
- **DOWN**: All the copiers supporting this endpoint are down or the system encountered an error when accessing the endpoint. The copier tries to bring the endpoint back online without administrator intervention.
- **BROKEN**: The system encountered an error when accessing the endpoint that requires administrator intervention. The endpoint can’t recover on its own.
- **DISABLED**: The administrator has disabled the endpoint from the endpoint configuration screen.

If a create or delete request requires a endpoint that is **DOWN**, the request continuously retries. If a create or delete request requires a endpoint that is **BROKEN**, **DISABLED**, or **OFFLINE**, the request is suspended. If a recall request requires a endpoint that is **DOWN**, **BROKEN**, **DISABLED**, or **OFFLINE**, the request failovers if it can. If it cannot failover, the request is suspended.
Configuring HSB instance components

**URL** - The protocol, host, port, and path for the endpoint.

**Last Failure** - The time of the last error on the endpoint.

**Usage Info** - The endpoint limits and running requests.

- **Limit** - The configured limit for the max number of requests that can be running on this endpoint simultaneously.

- **Current** - The current number of requests running on the endpoint.

**Requests** - The list of requests running on this endpoint.

### 3.3.4.4. Monitor endpoint state at copiers fields

Double clicking on an endpoint will access the list of supporting copiers and the state of the endpoint at that copier. This information is used to determine the overall endpoint state.

You can see in the example image above that the Workload HPSS Repo3 endpoint is in the degraded state. That is because the copiers Bastrop and Yuma are in the **DOWN** state. Because of the endpoint status is degraded, HSB will not schedule requests that require the Workload HPSS Repo3 endpoint on the Bastrop or Yuma copiers.

**State at Copier** - The list of copiers supporting the selected endpoint

**State** - The status of the endpoint at a copier. The possible states are:

- **ONLINE**: The endpoint is running normally.
- **OFFLINE**: The administrator has stopped the endpoint from the monitor screen.
- **DOWN**: The copier is not online or the system encountered an error when accessing the endpoint. The copier will try to bring the endpoint back online without the administrator intervention.
- **BROKEN**: The system encountered an error when accessing the endpoint that requires administrator intervention. The endpoint can’t recover on its own.
- **DISABLED**: The administrator has disabled the endpoint from the endpoint configuration screen

**Last Access** - The time the copier last accessed the endpoint

**Last IO Failure** - The time of the last error on the endpoint from the copier.
3.3.5. Servers

3.3.5.1. List servers

Select Monitor > Servers in the HSB administrative GUI:

The List Servers window is displayed:

The system currently supports only a single server instance.

- **HPSS Treefrog Server** - Treefrog server

This window allows you to monitor the status of the server.

Each server is comprised of a number of server components. Each server component can be monitored for status.

Click the control box to the left of the server name to show or hide the server component status.

- **HPSS Treefrog Database** - Db2 engine supporting the server
- **HPSS Treefrog ZooKeeper Service** - Zookeeper service supporting the server
- **HPSS Treefrog Request Processor Service** - Request processing service
- **HPSS Treefrog Web Server** - Web services
- **HPSS Treefrog Cred Expirer** - Credential expirer service

3.3.5.2. Monitor servers fields

**Name** - System defined name of this server or server component.
### Configuring HSB

#### instance components

<table>
<thead>
<tr>
<th><strong>Host</strong></th>
<th>Administrator defined host running the server and server components.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pid</strong></td>
<td>Process identifier of the server or server component, if applicable.</td>
</tr>
<tr>
<td><strong>State</strong></td>
<td>The status of the server or server component. The possible states are:</td>
</tr>
<tr>
<td></td>
<td>• ACTIVE: The server or server component is running normally.</td>
</tr>
<tr>
<td></td>
<td>• INACTIVE: The server or server component has been stopped.</td>
</tr>
<tr>
<td></td>
<td>• FAILED: The server or server component has experience an unexpected failure.</td>
</tr>
<tr>
<td><strong>State Updated</strong></td>
<td>The time of the last status change.</td>
</tr>
</tbody>
</table>

The `systemctl` command can be used to activate (start) or deactivate (stop) the treefrog server.

```
$ sudo -E systemctl start treefrogd
```
### 3.4. Tools menu

#### 3.4.1. Datasets

##### 3.4.1.1. List datasets

Select **Tools > Datasets** in the HSB administrative GUI:

![List Datasets window](image)

The **List Datasets** window is displayed:

![List Datasets content](image)

A dataset is a collection of data that is managed by HSB.

*See Dataset fields section for dataset field descriptions.*

#### 3.4.1.2. Display dataset

In the **List Datasets** window, click the magnifying glass button next to the desired dataset to open the **Show Dataset** window.

![Show Dataset window](image)
See Dataset fields section for dataset field descriptions.

List fragment chunks

In the Show Dataset window click on "List Fragment Chunks":

A fragment chunk is a subset of the data that makes up a dataset. Fragments are generated based on storage policy copy settings. See Storage policy copy fields for details about which fields affect fragmentation behavior.

See Fragment chunk fields section for fragment chunk field descriptions.

Dataset objects

In the Show Dataset window click on Dataset Objects:
A dataset object is a representation of a source object that is contained in this dataset. So in the above picture our dataset contains ten files from our static endpoint.

### 3.4.1.3. Dataset fields

- **Identifier** - ID number that uniquely identifies the dataset. Automatically generated upon dataset creation.
- **Name** - The name of the dataset.
- **Version** - The version of this dataset.
- **Creator** - The creator of this dataset.
- **Project Identifier** - The unique identifier of the project governing this dataset.
- **State** - The current state of the dataset. It can be one of the following:
  - **PENDING** - The dataset is in the process of being created.
  - **CREATED** - The dataset has been created.
- **Policy Identifier** - The unique identifier of the storage policy used to create this dataset.
- **Expiration Date** - The date this dataset expires. If the date is None, the dataset never expires.
- **Last Verification Date** - The date this dataset was last verified. If the date is None, the dataset has never been verified.
- **Dataset Size** - The combined size of all objects in the dataset. The unit of measure is in bytes.
- **Creation Date** - The timestamp (utc) when the dataset was created. The creation date is set after the manifest of objects has been successfully loaded in the database.

### Dataset object fields

- **Identifier** - The ID number of this dataset object.
- **Type** - The type of dataset object. It can be one of the following:
  - **DATA**: file or object data.
  - **FOLDER**: folder or container.
Configuring HSB instance components

**Path** - The full path of the dataset object on the source endpoint.

**Name** - The name of the dataset object on the source endpoint.
3.4.2. Fragment chunks

A fragment chunk is a subset of the data that makes up a fragment. Fragments are generated based on storage policy copy settings. See Storage policy fields for details about which fields affect fragmentation behavior.

3.4.2.1. List fragment chunks

Select Tools > Fragments/Chunks in the HSB administrative GUI:

The List Fragment Chunks window is displayed:

See Fragment chunk fields section for information on fragment chunks fields.

3.4.2.2. Display fragment chunk

In the List Fragment Chunks window, click the magnifying glass button next to the desired fragment chunk to open the Show Fragment Chunk window.
Configuring HSB instance components

See Fragment chunk fields section for information on fragment chunks fields.

3.4.2.3. Edit fragment chunk

In the List Fragment Chunks window, click the notepad-and-pen button next to the desired fragment chunk to open the Edit Fragment Chunk window.

See Fragment chunk fields section for information on the fragment chunks fields.

With the 1.1 release of HSB, enabling or disabling a fragment chunk does not affect the ability to recall data from the fragment chunk. Setting the fragment chunk to DISABLED can be used as a method for flagging issues with a particular fragment chunk.

3.4.2.4. Fragment chunk fields

Copy ID - The identifier of the storage policy copy used to create this fragment chunk.
Fragment Number - The identifier of this fragment within the dataset.
Fragment Chunk Type - The type of the fragment chunk. It can be one of the following:

- **DAT**: Contains data.
Configuring HSB
instance components

- **ECC**: Error correction code used for parity and recovering corrupted datasets.

**Fragment Chunk Starting Offset** - The starting offset of this fragment chunk. The number of bytes between the start of the fragment and the start of this fragment chunk.

**Fragment Chunk State** - The state of the fragment chunk. It can be one of the following:

- **NORMAL**: The current fragment state is nominal.
- **INTEGRITY_FAILURE**: The fragment’s checksum does not match the checksum stored in the database. This means that the data is possibly corrupted.
- **IO_FAILURE**: The last read attempt resulted in an I/O failure.
- **MISSING**: The last read failed because the chunk could not be located or opened.

**Fragment Chunk Enabled** - Specifies whether or not the chunk is enabled for recall. It can be one of the following:

- **ENABLED**: Available for recall.
- **DISABLED**: Not available for recall.

With the 1.1 release of HSB, enabling or disabling a fragment chunk does not affect the ability to recall data from the fragment chunk. Setting the fragment chunk to **DISABLED** can be used as a method for flagging issues with a particular fragment chunk.

**Fragment Chunk Length** - The total size of the fragment chunk in bytes.

**Chunk Location URI** - The location of the chunk.

**Last Accessed** - The date this fragment chunk was last accessed. Formatted as YYYY-MM-DD HH:MM:SS.MMMMMM.

**Volume Identifier** - Identifier for the tape volume that this fragment chunk is stored on. If blank, this fragment chunk is not stored on a tape volume.

**Fragment Checksum Algorithm** - The algorithm used to calculate this fragment’s checksum. Currently only MD5 checksums are supported.

**Fragment Checksum** - The checksum calculated for this fragment using the fragment checksum algorithm.

**Storage Endpoint Identifier** - Unique identifier for the endpoint that contains this fragment.
3.4.3. Download Linux client installer

Select **Tools > Download Linux Client Installer** in the HSB administrative GUI:

Depending on the browser you are using and its settings, the download either starts immediately or the browser will display an open or download dialog, as shown below.

This window indicates the filename (that is, `client_installer`) of the client install package. Beneath the filename, the file type is displayed as well as the server you are downloading the install package from. Since this is a shell script, leave the "Save File" option selected and select "OK" to download the install package to the downloads directory on your computer. In order to execute the shell script, you need to change the permissions of the script to "execute".
3.4.4. Download windows client installer

Select Tools > Download Windows Client Installer ZIP Package in the HSB administrative GUI:

Depending on the browser you are using and its settings, the download either starts immediately or the browser will display an open or download dialog, as shown below.

This window indicates the filename (that is, windows_client_install.zip) of the client install package. Beneath the filename, the file type is displayed as well as the server you are downloading the install package from. Since this is a Zip archive, your computer may have a default application associated with this file type. If a default application is present and you would like to open the Zip file with that application, select "Open with" and select "OK" to open the Zip file with the application. Otherwise, select the "Save File" option and select "OK" to download the install package to the downloads directory on your computer. In order to execute the installer, unzip the contents of the Zip file to an empty directory on your computer and run the batch file windows_client_install.bat.
3.5. Help

3.5.1. About

Select Help > About in the HSB administrative GUI:
3.6. HSB instance management

3.6.1. Starting

To bootstrap the user space, most Linux distributions use the systemd initialization system. Subsequent to booting, it is used to manage system processes. It is a replacement for the System V and BSD init systems.

Each of the following HSB system services are configured in systemd:

- treefrog_ced   HSB credentials expiration manager
- treefrog_copier HSB copier
- treefrog_dbd    HSB database engine
- treefrog_rpd    HSB request processor
- treefrog_wsd    HSB web server
- treefrog_zkd    HSB zookeeper

The HSB service can be started by issuing these two commands in the following order:

```bash
> systemctl start treefrogd
> systemctl start treefrog_copier
```

Starting individual services, other than the two above, should only be performed by developers or under the guidance of a HPSS support representative.

3.6.2. Stopping

The HSB service can be stopped by issuing these two commands in the following order:

```bash
> systemctl stop treefrogd
> systemctl stop treefrog_copier
```

Stopping individual services, other than the two above, should only be performed by developers or under the guidance of a HPSS support representative.

3.6.3. Database

The HSB system uses an IBM Db2 database to manage application metadata. The database can be interacted with outside of the HSB interfaces to perform normal maintenance activities, such as:

- database backups
  - see `db2_fullbackup` tool for details.
- table maintenance

Other activities should be done under the guidance of an HPSS support representative.
3.6.4. Monitoring

The health and status of the HSB system can be monitored in several places:

- The various HSB log files described in the Section 1.1.10, “HSB logging” section of the system overview
- The Db2 diagnostics log file db2diag.log, which is usually located in the instance home directory under SQLLIB/db2dump/DIAG0000/
- The HSB administrative GUI can monitor the health and status of the following:
  - Copiers
  - Storage endpoints
  - Requests
  - Events
- HSB uses ABRT (Automatic Bug detection and Reporting Tool) to monitor for crashes

3.6.5. Using the abrt tools

HSB uses ABRT to monitor for crashes. The abrtd process will run at all times and is automatically started by RHEL. It will detect crashes for programs and the kernel. When a crash is detected a crash report directory is generated that includes any core files, and syslogs that might be relevant to the crash.

HSB expects the standard configuration of ABRT that comes with the base RHEL install. Configuration changes or extra packages won’t stop HSB crash report generation but could result in the crash reports being in a different location. For example, if automatic reporting is configured, the crash reports might have already been tarballed and copied to a central system.

Administrators can use either a CLI or GUI to manage crash reports. This document assumes that you are using the CLI. The CLI commands that will normally be used are abrt-cli and report-cli.

The abrt-cli command is used to manage any crashes on the system. Entering abrt-cli list will list all of the crash reports available. It will also tell you if the the crash has been reported.

```bash
abrt-cli list
id 17b8e6ef4cba5f18650e2120747fa67690c04a6
reason:        sleep killed by SIGSEGV
time:           Tue 21 Jul 2020 02:57:52 PM CDT
cmdline:        sleep 100
package:        coreutils-8.22-18.el7
uid:            0 (root)
count:          1
Reported:       scp://hero/tmp/abrt.tar
                scp://hero/tmp/ccpp-2020-07-21-14:57:52-38171.tar.gz
```

The crash report directory contains all of the information captured about the crash including logs, environment, and core files.
cd /var/spool/abrt/ccpp-2020-07-21-14:57:52-38171
[root@hero /var/spool/abrt/ccpp-2020-07-21-14:57:52-38171]# ls
abrt_version    count            limits      pkg_epoch        reported_to
analyzer        dso_list         machineid   pkg_fingerprint  runlevel
architecture    environ          maps        pkg_name         sosreport.tar.xz
cgroup          event_log        open_fds    pkg_release      time
cmdline         executable       os_info     pkg_vendor       type
comment         global_pid       os_release  pkg_version      uid
component       hostname         package     proc_pid_status  username
core_backtrace  kernel          package     proc_pid_status  username
coredump        last_occurrence  pkg_arch    reason           var_log_messages

The report-cli tool can be used to report the crash. There are many report tools available but report_Uploader is very useful, since it will generate a tarball of the crash report directory and SCP|FTP it to our desired location. Here is an example of using the report-cli command:

Report Uploader URL: scp://root@localhost/tmp/

No changes were detected in the report
Compressing data
Sending /var/tmp/ccpp-2020-07-21-08:38:48-20099.tar.gz to scp://localhost
Please enter user name for 'scp://localhost': Treefrog
Please enter password for 'scp://root@localhost':
Sending /var/tmp/ccpp-2020-07-21-08:38:48-20099.tar.gz to scp://localhost
Successfully created scp://localhost/tmp/ccpp-2020-07-21-08:38:48-20099.tar.gz

To clean up a crash report you can use the command abrt-cli delete.

### 3.7. HSB performance tuning

#### 3.7.1. Database connections

HSB uses SQLAlchemy to interface with the database and there are settings available to control the database connection pool maintained by the application. These settings can be increased so that the system is better able to handle a greater level of concurrent activity. The settings are SQLALCHEMY_POOL_SIZE and SQLALCHEMY_MAX_OVERFLOW and can be found in the server configuration file /var/treefrog/server.conf.

**SQLALCHEMY_POOL_SIZE** - The maximum number of connections held open by SQLAlchemy in the connection pool.

**SQLALCHEMY_MAX_OVERFLOW** - The number of connections that can be created after the pool reached its maximum size. When those additional connections are returned to the pool, they are disconnected and discarded.

The following example show a configuration that allows for five database connections with an overflow size of ten additional connections.

<root>$ cat /var/treefrog/server.conf

SQLALCHEMY_ECHO = False
SQLALCHEMY_POOL_SIZE = 5
SQLALCHEMY_MAX_OVERFLOW = 10
Configuring HSB
instance components

The following example shows a configuration that allows for 20 database connections with an overflow size of 40 additional connections.

```
<root>$ cat /var/treefrog/server.conf
SQLALCHEMY_ECHO = False
SQLALCHEMY_POOL_SIZE = 20
SQLALCHEMY_MAX_OVERFLOW = 40
...
```

To pick up changes to the server configuration file, HSB must be restarted.
Example:

```
$ systemctl restart treefrogd
```

### 3.7.2. Controlling copier memory usage for large dataset requests

The memory footprint of the HSB copier during the dataset create and recall process can become significant depending on the number of requests running and how many objects are in each request. For each create and recall request the HSB copier maintains a list of the objects associated with the request in memory. By default, if a dataset contains more than 10,000 objects the copier will start writing the list to a scratch disk to restrain memory usage. To provide some control over the amount of memory usage by the copier, the default value of 10,000 can be adjusted by adding and setting the `obj-list-limit` parameter to the copier configuration file `/etc/treefrog/copier.conf`.

The following example shows a configuration that sets the `obj-list-limit` parameter to 5,000.

```
<root>$ cat /etc/treefrog/copier.conf
... obj-list-limit=5000 ...
```

To pick up changes to the copier configuration file, the copier must be restarted.
Example:

```
$ systemctl restart treefrog_copier
```

The scratch disk location for the cached object lists is `/var/lib/treefrog/tmp`. This space is best allocated on fast access disk and the amount of storage should be based on the average number of objects per dataset and the average number of dataset creates/recalls being handled by the copier. When these lists are backed to disk, they occupy about (1.8 * 2) 2.6KiB per file contained in each dataset currently being created or recalled. For example, if you expect two 1 million file dataset creates or recalls to be simultaneously processed by a copier, that copier would need at least 2 * (1000000 * 2.6KiB)=~5GiB allocated for this file system.
Chapter 4. HSB tools

4.1. Tools

4.1.1. HSB import utility

The HSB import utility, known as treefrog_import, allows you to find archive files or archive objects on a specific static endpoint that are import candidates and import them into HSB as a dataset. In order to be able find and import archive files or objects on a specific static endpoint the endpoint must be added as an import source to at least one policy. Refer to the Storage policy configuration section for more information regarding the configuration of an import source.

The treefrog_import utility can currently only be run on the HSB server.

4.1.1.1. Usage

To invoke the utility:

```
> treefrog_import [-h] [--host HOST] [-u USER] [-pw PW] [--stash]
  [-c CERT] [-v] [-proj PROJECT] [-pol POLICY]
  [-ep ENDPOINT] [-ds DATASET]
  [--archive-type ARCHIVE_TYPE] [--mgmt-type MGMT_TYPE]
  [--can-access-archive] [--skip-lookup] [-i INPUT]
  {import,find} [source]
```

4.1.1.2. Positional arguments

find

Allows you to find archive files or archive objects that are candidates for an import.

import

Allows you to import archive files or archive objects into HSB as a dataset.

source

The import source URL.

4.1.1.3. Optional arguments

-h, --help

show this help message and exit

--host HOST

server host. Defaults to the hostname of the HSB server.

-u USER, --user USER

account name. Defaults to user account currently logged on.

-pw PW

account password
--stash  
use cookie stash file. Defaults to /<homedir>/.cookie where "homedir" is the value of the $HOME environment variable.

-c CERT, --cert CERT  
PEM file with server credential. Defaults to /var/treefrog/ssl/ca.crt.

-v, --verbose  
output additional execution information to the console. This option can be used multiple times to increase the level of detailed reported.

-proj PROJECT, --project PROJECT  
project name

-pol POLICY, --policy POLICY  
storage policy name

-ep ENDPOINT, --endpoint ENDPOINT  
source endpoint name

-ds DATASET, --dataset DATASET  
dataset name

--archive-type ARCHIVE_TYPE  
source archive type: TAR or HTAR

--mgmt-type MGMT_TYPE  
USER, SHARED, SYSTEM. Currently only USER is supported.

--can-access-archive  
if needed, access the archive file during import

--skip-lookup  
skip the dataset lookup as part of the find command

-i INPUT, --input INPUT  
source input file

-exit-on-error  
ext import (file input) processing on error

--dry-run  
process import, but do not issue the import command

-s, --silent  
do not show progress indicator

-n, --nice  
make an effort to throttle processing based on a relative sampling of server and database response times

-l LOG, --log LOG  
log output to the specified file
-j, --json
log only un-imported archive files/objects. The entries are logged in a JSON format suitable to used as input for the `import` command

### 4.1.1.4. Example usage

1. Find archive files or objects on a static endpoint whose path is `/TreeFrogEndpoints/TreeFrogFS`:

   ```sh
   > treefrog_import -v -u <username> -pw <password> find /TreeFrogEndpoints/TreeFrogFS
   ```

2. Import an archive file whose path is `filesystem://sora.clearlake.ibm.com/TreeFrogEndpoints/TreeFrogFS/dir1/test.tar`:

   ```sh
   > treefrog_import -v -u <username> -pw <password> -proj <projectname> -pol <policyname> --can-access-archive import filesystem://sora.clearlake.ibm.com/TreeFrogEndpoints/TreeFrogFS/dir1/test.tar
   ```

3. Find archive files or objects on a static endpoint and log the un-imported files or objects to a log file that can be used with the `import` command.

   ```sh
   > treefrog_import -v -u <username> -pw <password> --json --log <logfilename> find /TreeFrogEndpoints/TreeFrogFS
   ```

4. Use an input file generated by the `find` command to import a list of un-imported archive files or objects.

   ```sh
   > treefrog_import -v -u <username> -pw <password> -proj <projectname> -pol <policyname> --can-access-archive --input <logfilename> import
   ```
# Appendix A. Glossary of terms and acronyms

<table>
<thead>
<tr>
<th><strong>ABRT</strong></th>
<th>Automatic Bug detection and Reporting Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Catalog</strong></td>
<td>A complete list of the files, objects, directories, containers and chunks that comprise a Managed Dataset. This list is stored in a catalog file.</td>
</tr>
<tr>
<td><strong>Class of Service</strong></td>
<td>A set of storage system characteristics used to group HPSS bitfiles with similar logical characteristics and performance requirements together. A Class of Service is supported by an underlying hierarchy of storage classes.</td>
</tr>
<tr>
<td><strong>Chunk</strong></td>
<td>Contiguous data within a managed dataset. Fragments are composed of one or more chunks. Fragments can be broken into multiple chunks to facilitate device storage capacity limitations.</td>
</tr>
<tr>
<td><strong>Copier</strong></td>
<td>Component of the HSB service that creates and recalls managed datasets and lists contents of storage endpoints.</td>
</tr>
<tr>
<td><strong>COS</strong></td>
<td>Class of Service</td>
</tr>
<tr>
<td><strong>CRC</strong></td>
<td>Cyclic Redundancy Check</td>
</tr>
<tr>
<td><strong>Credential expirer</strong></td>
<td>Component of the HSB server that monitors endpoint credentials and notifies the user when credentials are older than the configured expiration period.</td>
</tr>
<tr>
<td><strong>Db2</strong></td>
<td>A relational database system, a product of IBM Corporation, used by HSB to store and manage HSB system metadata.</td>
</tr>
<tr>
<td><strong>Directory, Container</strong></td>
<td>The container components of a file system and object store, respectively.</td>
</tr>
<tr>
<td><strong>DNS</strong></td>
<td>Domain Name Service</td>
</tr>
<tr>
<td><strong>DOE</strong></td>
<td>Department of Energy</td>
</tr>
<tr>
<td><strong>ECC</strong></td>
<td>Error Correction Code</td>
</tr>
<tr>
<td><strong>EOM</strong></td>
<td>End of Media</td>
</tr>
<tr>
<td><strong>File, Object</strong></td>
<td>Data components of a file system and object store, respectively.</td>
</tr>
<tr>
<td><strong>File family</strong></td>
<td>An attribute of an HPSS file that is used to group a set of files on a common set of tape virtual volumes.</td>
</tr>
<tr>
<td><strong>Fragment</strong></td>
<td>Logically contiguous data within a managed dataset. Managed datasets can be fragmented into some number of relatively equal pieces to facilitate increased transfer performance via concurrent, parallel transfers and provide redundancy via the generation of parity fragments.</td>
</tr>
<tr>
<td><strong>FTP</strong></td>
<td>File Transfer Protocol</td>
</tr>
<tr>
<td><strong>GUI</strong></td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td><strong>HADR</strong></td>
<td>Db2 High Availability Disaster Recovery</td>
</tr>
<tr>
<td><strong>HPSS</strong></td>
<td>High Performance Storage System</td>
</tr>
</tbody>
</table>
Glossary of terms and acronyms

**HPSS bitfile**  A file stored in HPSS, represented as a logical string of bits unrestricted in size or internal structure. HPSS imposes a size limitation in 8-bit bytes, based upon the maximum size in bytes that can be represented by a 64-bit unsigned integer.

**HPSS Storage Broker (HSB)**  High Performance Storage System Storage Broker

**HSB service**  On-premise service that allows users to copy data between defined storage systems in a high-performance manner.

**HTTP**  Hyper Text Transmission Protocol

**IBM**  International Business Machines Corporation

**IEEE**  Institute of Electrical and Electronics Engineers

**Instance**  Instance of a HSB project name space.

**I/O**  Input/Output

**IP**  Internet Protocol

**ISA**  Intel ® Intelligent Storage Acceleration

**JRE**  Java Runtime Environment

**LAN**  Local Area Network

**LANL**  Los Alamos National Laboratory

**LBP**  Logical Block Protection

**LDAP**  Lightweight Directory Access Protocol

**LLNL**  Lawrence Livermore National Laboratory

**LTO**  Linear Tape-Open. A half-inch open tape technology developed by IBM, HP, and Seagate.

**Manifest**  A listing of the files and objects that comprise a managed dataset.

**MAC**  Mandatory Access Control

**Managed Data Set (MDS)**  Immutable collection of files or objects managed by the HSB service. Managed data sets are intended as a mechanism to group logically associated data and emphasize the desirable characteristics of high-latency, high-capacity storage. User-defined metadata can be associated with each managed dataset to facilitate efficient location and retrieval.

**Name space**  An organization of projects and managed datasets, so that these components can be referred to by name.

**NASA**  National Aeronautics and Space Administration

**NASM**  Netwide Assembler is an assembler for the x86 CPU architecture.

**NERSC**  National Energy Research Supercomputer Center

**NIS**  Network Information Service

**NLS**  National Language Support

**NSL**  National Storage Laboratory

**ORNL**  Oak Ridge National Laboratory

**PFTP**  Parallel extensions to File Transfer Protocol supported by HPSS
<table>
<thead>
<tr>
<th>Glossary of terms and acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PFTPD</strong></td>
</tr>
<tr>
<td><strong>POSIX</strong></td>
</tr>
<tr>
<td><strong>Project</strong></td>
</tr>
<tr>
<td><strong>RAO</strong></td>
</tr>
<tr>
<td><strong>Repository storage endpoint</strong></td>
</tr>
<tr>
<td><strong>Request processor</strong></td>
</tr>
<tr>
<td><strong>Server</strong></td>
</tr>
<tr>
<td><strong>SNIA</strong></td>
</tr>
<tr>
<td><strong>SNL</strong></td>
</tr>
<tr>
<td><strong>SSL</strong></td>
</tr>
<tr>
<td><strong>Storage Policy</strong></td>
</tr>
<tr>
<td><strong>Static storage endpoint</strong></td>
</tr>
<tr>
<td><strong>TCP/IP</strong></td>
</tr>
<tr>
<td><strong>TLS</strong></td>
</tr>
<tr>
<td><strong>UDA</strong></td>
</tr>
<tr>
<td><strong>URI</strong></td>
</tr>
<tr>
<td><strong>User</strong></td>
</tr>
<tr>
<td><strong>UUID</strong></td>
</tr>
<tr>
<td><strong>Web services</strong></td>
</tr>
<tr>
<td><strong>XML</strong></td>
</tr>
<tr>
<td><strong>YASM</strong></td>
</tr>
</tbody>
</table>
Appendix B. Install package contents

B.1. x86_64 Architecture

The contents include for x86_64 architectures:

HSB server wheel

• TreefrogServer-<version>-<release>-py27-none-any.whl

Db2 server and license files

• v11.1.4fp4a_linuxx64_universal_fixpack.tar.gz

• license files:
  • db2ese_c.lic - Db2 Enterprise Server license
  • db2hc.lic - Db2 High Capacity license
  • iidr.lic - Infosphere Data Replication license

HSB server RPM files

• python2-pip-8.1.2-5.el7.noarch.rpm

Zookeeper installation material

• zookeeper-3.5.4-beta.tar.gz

HSB server prerequisite wheels

• Babel-2.6.0-py2.py3-none-any.whl
• certifi-2017.7.27.1-py2.py3-none-any.whl
• cffi-1.11.5-cp27-cp27mu-manylinux1_x86_64.whl
• chardet-3.0.4-py2.py3-none-any.whl
• click-6.7-py2.py3-none-any.whl
• clickclick-1.2.2-py2.py3-none-any.whl
• colorama-0.3.9-py2.py3-none-any.whl
• enum34-1.1.6-py2-none-any.whl
• flake8-2.6.0-py2.py3-none-any.whl
Install package contents

- Flask_Babel-0.11.2-py2.py3-none-any.whl
- Flask_WTF-0.14.2-py2.py3-none-any.whl
- funcsigns-1.0.2-py2.py3-none-any.whl
- future-0.17.1-cp27-none-any.whl
- idna-2.8-py2.py3-none-any.whl
- jsonschema-2.6.0-py2.py3-none-any.whl
- kazoo-2.6.0-py2.py3-none-any.whl
- mccabe-0.5.3-py2.py3-none-any.whl
- mock-2.0.0-py2.py3-none-any.whl
- pbr-5.1.1-py2.py3-none-any.whl
- pip-9.0.1-py2.py3-none-any.whl
- pycodestyle-2.0.0-py2.py3-none-any.whl
- pycparser-2.19-py2.py3-none-any.whl
- pycryptoo-2.6.1-cp27-cp27mu-linux_x86_64.whl
- pyflakes-1.2.3-py2.py3-none-any.whl
- PyJWT-1.7.1-py2.py3-none-any.whl
- python_dateutil-2.7.2-py2.py3-none-any.whl
- requests-2.21.0-py2.py3-none-any.whl
- setuptools-28.8.0-py2.py3-none-any.whl
- swagger_spec_validator-2.4.1-py2.py3-none-any.whl
- typing-3.6.2-py2-none-any.whl
- urllib3-1.24.1-py2.py3-none-any.whl
- Werkzeug-0.12.2-py2.py3-none-any.whl
- PyYAML-3.13-cp27-cp27mu-linux_x86_64.whl
- blinker-1.4-py2-none-any.whl
- Flask_SQLAlchemy-2.0-py2-none-any.whl
- Jinja2-2.7.2-py2-none-any.whl
Install package contents

- Flask_AppBuilder-1.8.1-py2-none-any.whl
- ibm_db-2.0.8-cp27-cp27mu-linux_x86_64.whl
- ibm_db_sa-0.3.2-py2-none-any.whl
- Flask_Login-0.2.11-py2-none-any.whl
- strict-rfc3339-0.7.tar.gz
- Flask_OpenID-1.2.5-py2-none-any.whl
- Flask_Testing-0.7.1-py2-none-any.whl
- WTForms-2.1-py2.py3-none-any.whl
- SQLAlchemy-1.1.10-cp27-cp27mu-linux_x86_64.whl
- python-openid-2.2.5.tar.gz
- connexion-1.0.129-py2-none-any.whl
- Flask-0.10.1-py2-none-any.whl
- Flask_Mail-0.9.1-py2-none-any.whl
- itsdangerous-0.22-py2-none-any.whl
- MarkupSafe-0.11.tar.gz
- functools32-3.2.3-2.tar.gz
- pathlib-1.0.1.tar.gz
- ijson-3.1.tar.gz

B.2. ppc64le Architecture

HSB server wheel

- TreefrogServer-<version>-<release>-py27-none-any.whl

Db2 server and license files

- v11.1.4fp4a_linuxppc64le_universal_fixpack.tar.gz

license files:

- db2ese_c.lic - Db2 Enterprise Server license
- db2hc.lic - Db2 High Capacity license
- iidr.lic - Infosphere Data Replication license
Install package contents

HSB server RPM files

- python2-pip-8.1.2-5.el7.noarch.rpm

XLC

- XL_C_C_FOR_LINUX_13.1.6_PRODUCT.tar.gz

Zookeeper installation material

- zookeeper-3.5.4-beta.tar.gz

HSB server prerequisite wheels

- Babel-2.6.0-py2.py3-none-any.whl
- certifi-2017.7.27.1-py2.py3-none-any.whl
- cffi-1.12.3-cp27-cp27mu-linux_ppc64le.whl
- chardet-3.0.4-py2.py3-none-any.whl
- click-6.7-py2.py3-none-any.whl
- clickclick-1.2.2-py2.py3-none-any.whl
- colorama-0.3.9-py2.py3-none-any.whl
- enum34-1.1.6-py2-none-any.whl
- flake8-2.6.0-py2.py3-none-any.whl
- Flask_Babel-0.11.2-py2.py3-none-any.whl
- Flask_WTF-0.14.2-py2.py3-none-any.whl
- funcsig3-1.0.2-py2.py3-none-any.whl
- future-0.17.1-cp27-none-any.whl
- idna-2.8-py2.py3-none-any.whl
- jsonschema-2.6.0-py2.py3-none-any.whl
- kazoo-2.6.0-py2.py3-none-any.whl
- mccabe-0.5.3-py2.py3-none-any.whl
- mock-2.0.0-py2.py3-none-any.whl
- pbr-5.1.1-py2.py3-none-any.whl
- pip-9.0.1-py2.py3-none-any.whl
- pycodestyle-2.0.0-py2.py3-none-any.whl
Install package contents

- pycparser-2.19-py2.py3-none-any.whl
- pycrypto-2.6.1-cp27-cp27mu-linux_ppc64le.whl
- pyflakes-1.2.3-py2.py3-none-any.whl
- PyJWT-1.7.1-py2.py3-none-any.whl
- python_dateutil-2.7.2-py2.py3-none-any.whl
- requests-2.21.0-py2.py3-none-any.whl
- setuptools-28.8.0-py2.py3-none-any.whl
- swagger_spec_validator-2.4.1-py2.py3-none-any.whl
- typing-3.6.2-py2-none-any.whl
- urllib3-1.24.1-py2.py3-none-any.whl
- Werkzeug-0.12.2-py2.py3-none-any.whl
- PyYAML-3.13-cp27-cp27mu-linux_ppc64le.whl
- blinker-1.4-py2-none-any.whl
- Flask_SQLAlchemy-2.0-py2-none-any.whl
- Jinja2-2.7.2-py2-none-any.whl
- Flask_AppBuilder-1.8.1-py2-none-any.whl
- ibm_db-2.0.8-cp27-cp27mu-linux_ppc64le.whl
- ibm_db_sa-0.3.2-py2-none-any.whl
- Flask_Login-0.2.11-py2-none-any.whl
- strict-rfc3339-0.7.tar.gz
- Flask_OpenID-1.2.5-py2-none-any.whl
- Flask_Testing-0.7.1-py2-none-any.whl
- WTFORMS-2.1-py2.py3-none-any.whl
- SQLAlchemy-1.1.10-cp27-cp27mu-linux_ppc64le.whl
- python-openid-2.2.5.tar.gz
- connexion-1.0.129-py2-none-any.whl
- Flask-0.10.1-py2-none-any.whl
- Flask_Mail-0.9.1-py2-none-any.whl
Install package contents

- itsdangerous-0.22-py2-none-any.whl
- MarkupSafe-0.11.tar.gz
- functools32-3.2.3-2.tar.gz
- pathlib-1.0.1.tar.gz
- ijson-3.1.tar.gz
Appendix C. Database storage estimation tips

C.1. Space requirements

Database Sizing Factors

The configuration metadata is a very small percentage of the overall space requirements and is not factored into the space requirements. Storage space required by the database varies by installation, but will be driven by the following factors:

1. The number of managed datasets.

2. The number of copies defined in the storage policies used by the managed datasets. This affects the number of rows in the MDS_CATALOG table. There is a row in this table for every copy of a managed dataset.

3. The number of fragments defined in the storage policies used by the managed datasets. This affects the number of rows in the MDS_FRAG_CHUNK table. There is a row in this table for every copy of a managed dataset with addition rows based on the fragmentation settings defined by the storage policy copy.

4. The amount of user metadata ingested.

User metadata is stored in Db2 as an XML object. Each XML object can be as large as 2 GiB and multiple XML objects can be stored per managed dataset.

Additional indexes can be defined to optimize site-specific searching.

User Metadata can significantly increase the storage requirements and workload on the database. If you expect to store significant amounts of user metadata or plan on running significant or frequent searches on user metadata, contact HPSS support. Support personnel can review storage and query expectations and suggest potential configuration optimizations.

5. The number of directories, containers, files, or objects stored in the managed datasets. This affects the number of rows in the MDS_MANIFEST table. There is a row in this table for every directory, container, file, and object stored in each managed dataset.

6. The retention period defined for manifest associated with the managed dataset. Manifest rows can be purged from the database to reduce the space requirements and limit growth. The manifest data is also stored in catalog files at the storage endpoints and can be retrieved/reloaded into the database if needed after being purged. This does not include user metadata associated with the managed dataset. The retention period is a site-specific setting.

7. Tablespace density:

   a. The table spaces are the objects that actually store table data.
b. They are managed using a database concept called "Managed By Automatic Storage" where the database manager extends tablespace storage containers as needed.

c. When the available space in a tablespace is used up, the amount of space used per object is lower (higher density) than immediately following a tablespace expansion request (lower density).

i. The variance in tablespace density before and after expansion will depend on the expansion size.

ii. The larger the table, the less important this becomes when trying to estimate space usage.

8. Database compression efficiency:

a. The data in the larger tables is compressed using the database’s compression logic algorithm. The database uses three forms of compression to reduce the storage requirements of the database data:

   **Classic row compression**
   A dictionary-based compression algorithm to replace recurring strings with shorter symbols within data rows.

   **Adaptive compression**
   Improves upon the compression rates that can be achieved using Classic Row Compression by itself. Adaptive compression incorporates classic row compression; however, it also works on a page-by-page basis to further compress data.

   **Index compression**
   The degree of compression achieved will vary based on the type of index you are creating, as well as the data the index contains. For example, the database manager can compress an index with a large number of duplicate keys by storing an abbreviated format of the record identifier (RID) for the duplicate keys. In an index where there is a high degree of commonality in the prefixes of the index keys, the database manager can apply compression based on the similarities in prefixes of index keys.

b. The compressed data will remain compressed as it is retrieved from storage (Tablespaces) and placed in memory (Bufferpools) and when the row updates are recorded in the Transaction Logs. See the Transaction Logging Requirements

For details on table data compression visit: https://www.ibm.com/support/knowledgecenter/SSEPGG_11.5.0/com.ibm.db2.luw.admin.dbobj.doc/doc/c0007306.html

For details on index compression visit: https://www.ibm.com/support/knowledgecenter/SSEPGG_11.5.0/com.ibm.db2.luw.admin.dbobj.doc/doc/c0054539.html

**Example**

The following tables are provided as examples to illustrate the space used by randomly generated test data. Each site’s storage requirements will vary:
### Database storage estimation tips

Configuration related metadata:

<table>
<thead>
<tr>
<th>Table</th>
<th>Table</th>
<th>Row Count</th>
<th>Index Space (KiB)</th>
<th>Data Space (KiB)</th>
<th>Total Space (KiB)</th>
<th>Space per (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copier</td>
<td></td>
<td>1,000</td>
<td>512</td>
<td>384</td>
<td>896</td>
<td>917</td>
</tr>
<tr>
<td>Storage Endpoints</td>
<td></td>
<td>1,000</td>
<td>640</td>
<td>640</td>
<td>1,280</td>
<td>1,310</td>
</tr>
<tr>
<td>Copier Endpoints</td>
<td></td>
<td>100,000</td>
<td>4,992</td>
<td>2,816</td>
<td>7,808</td>
<td>79</td>
</tr>
<tr>
<td>Storage Policies</td>
<td></td>
<td>1,000</td>
<td>384</td>
<td>384</td>
<td>768</td>
<td>786</td>
</tr>
<tr>
<td>Storage Policy Copies</td>
<td></td>
<td>5,000</td>
<td>1,024</td>
<td>1,152</td>
<td>2,176</td>
<td>445</td>
</tr>
<tr>
<td>Storage Policy Copy Endpoints</td>
<td></td>
<td>46,183</td>
<td>2,560</td>
<td>1,536</td>
<td>4,096</td>
<td>90</td>
</tr>
<tr>
<td>Users</td>
<td></td>
<td>10,000</td>
<td>1,152</td>
<td>3,072</td>
<td>4,224</td>
<td>432</td>
</tr>
<tr>
<td>User Project Assignment</td>
<td></td>
<td>99,940</td>
<td>3,712</td>
<td>2,816</td>
<td>6,528</td>
<td>66</td>
</tr>
</tbody>
</table>

Managed Data Set related metadata:

<table>
<thead>
<tr>
<th>Table</th>
<th>Table</th>
<th>Row Count</th>
<th>Index Space (KiB)</th>
<th>Data Space (KiB)</th>
<th>Total Space (KiB)</th>
<th>Space per (bytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projects</td>
<td></td>
<td>10,000</td>
<td>7,168</td>
<td>6,144</td>
<td>13,312</td>
<td>891</td>
</tr>
<tr>
<td>Project Storage Policy Assignment</td>
<td></td>
<td>99,486</td>
<td>4,096</td>
<td>3,072</td>
<td>7,168</td>
<td>73</td>
</tr>
<tr>
<td>Managed Datasets</td>
<td></td>
<td>998,837</td>
<td>97,280</td>
<td>60,416</td>
<td>157,696</td>
<td>161</td>
</tr>
<tr>
<td>Managed Dataset Catalogs</td>
<td></td>
<td>4,994,185</td>
<td>100,352</td>
<td>241,664</td>
<td>342,016</td>
<td>70</td>
</tr>
<tr>
<td>Managed Dataset Fragments</td>
<td></td>
<td>49,838,687</td>
<td>3,299,328</td>
<td>4,685,824</td>
<td>7,985,152</td>
<td>164</td>
</tr>
<tr>
<td>Managed Dataset Manifests</td>
<td></td>
<td>100,000,171</td>
<td>22,378,496</td>
<td>17,212,416</td>
<td>39,590,912</td>
<td>405</td>
</tr>
</tbody>
</table>

Database space allocation and utilization:

<table>
<thead>
<tr>
<th>Total space allocated for data/indexes (KiB)</th>
<th>Total space used (KiB)</th>
<th>Percent Utilized</th>
</tr>
</thead>
</table>
Space allocated to the database will increase automatically as the space allocated is consumed (when percent utilized reaches 100%). The increase in amount of space allocated can vary with each space request made by the database.

Event message tables are not shown in the example data above.

## C.2. Transaction logging

Database logging is an important part of a highly available database solution design. Database logs make it possible to recover from a failure. These logs keep a record of database changes. If a database needs to be restored to a point beyond the last full offline backup, logs are required to roll the data forward to the point of failure.

By default the HSB setup utility `treefrog_setup` configures the database for **archival** logging.

For details on archival logging see: https://www.ibm.com/support/knowledgecenter/SSEP GG_11.5.0/com.ibm.db2.luw.admin.ha.doc/doc/c0051344.html

See the following database configuration parameter descriptions for details on configuring database transaction logging and space allocations:

- `NEWLOGPATH`
- `MIRRORLOGPATH`
- `LOGPRIMARY`
- `LOGSECOND`
- `LOGFILSIZ`
- `LOGBUFSZ`
- `LOGARCHMETH1`
- `LOGARCHMETH2`
- `LOGARCHCOMPR1`
- `LOGARCHCOMPR2`
- `NUM_DB_BACKUPS`
- `AUTO_DEL_REC_OBJ`
- `REC_HIS_RETENTN`
Transaction logging space requirements can be affected by the database compression efficiency.
Appendix D. Db2 database configuration parameters

Configuration parameters that are not included in the following list are set to their default values. For additional detail on Db2 database configuration parameters go to the Db2 Knowledge Center: https://www.ibm.com/support/knowledgecenter/SSEPGG_11.5.0/com.ibm.db2.luw.admin.config.doc/doc/c0060795.html

For custom database configurations contact HPSS support.

D.1. NEWLOGPATH

Default value: /var/treefrog/database/log

This parameter allows you to specify a string of up to 242 bytes to change the location where the log files are stored.

The NEWLOGPATH and MIRRORLOGPATH should reference different storage devices.

D.2. MIRRORLOGPATH

Default value: /var/treefrog/database/mirrorlog

This parameter allows you to specify a string of up to 242 bytes to change the location where the log files are stored.

D.3. LOGPRIMARY

Default value: 10

This parameter allows you to specify the number of primary log files to be preallocated. The primary log files establish a fixed amount of storage allocated to the recovery log files.

When the database is first activated the primary log files are created in the log paths specified in NEWLOGPATH and MIRRORLOGPATH. Each logfile will have space pre-allocated from the file system. The amount of space pre-allocated depends on the configuration parameter LOGFILSZ.

D.4. LOGSECOND

Default value: -1
This parameter specifies the number of secondary log files that are created and used for recovery log files. The secondary log files are created only as needed.

A value of \(-1\) indicates the database is configured with infinite active log space. There is no limit on the size or the number of in-flight transactions running on the database. LOGPRIMARY and LOGFILSZ still control the number and size of log files the database manager should keep in the active log path.

**D.5. LOGFILSZ**

Default value: 25000

Unit of measure: 4KiB page

This parameter defines the size of each primary and secondary log file. The default size of each log file created will be: \(25000 \times 4\text{KiB} = 102400000\) bytes

**D.6. LOGBUFSZ**

Default value: 16384

This parameter allows you to specify the amount of the database heap (defined by the dbheap parameter) to use as a buffer for log records before writing these records to disk.

**D.7. LOGARCHMETH1**

Default value: /var/treelfrog/database/logarchive1

This parameter specifies the media type and location of the primary destination for logs that are archived from the current log path.

The filesystem supporting this path should have enough space allocated to handle the number of transaction log files expected to be generated and archived over the period of time defined by the site. This should coincide with the age and number of database backups kept on the system for a timely recovery. For example, if the goal is to keep a week’s worth of database backups online, then the archived log files that have occurred since the oldest backup should be kept online as well. This number will vary and depends on the transaction workload the database is performing. More insert, update, or delete operations generate more log files.

**D.8. LOGARCHMETH2**

Default value: /var/treelfrog/database/logarchive2

This parameter specifies the media type and location of the secondary destination for logs that are archived from either the current log path or the mirror log path. See LOGARCHMETH1 for file system details.
D.9. LOGARCHCOMPR1

Default value: ON

This parameter specifies whether the log files written to the primary archive destination for logs are compressed.

D.10. LOGARCHCOMPR2

Default value: ON

This parameter specifies whether the log files written to the secondary archive destination for logs are compressed.

D.11. NUM_DB_BACKUPS

Default value: 4

This parameter specifies the number of full database backups to retain for a database.

D.12. REC_HIS_RETENTN

Default value: 0

This parameter specifies the number of days that historical information on backups are retained.

When set to "0" and AUTO_DEL_REC_OBJ is set to ON, automated history file pruning and recovery object deletion are carried out based on the timestamp of the oldest backup maintained by the NUM_DB_BACKUPS database configuration parameter.

D.13. AUTO_DEL_REC_OBJ

Default value: ON

This parameter specifies whether database log files, backup images, and load copy images should be deleted when their associated recovery history file entry is pruned.

D.14. SELF_TUNING_MEM

Default value: ON

This parameter determines whether the memory tuner will dynamically distribute available memory resources as required between memory consumers that are enabled for self-tuning.

D.15. PCKCACHESZ

Default value: AUTOMATIC
Db2 database configuration parameters

This parameter is allocated out of the database shared memory, and is used for caching of sections for static and dynamic SQL and XQuery statements on a database.

**D.16. LOCKLIST**

Default value: AUTOMATIC

This parameter indicates the amount of storage that is allocated to the lock list. There is one lock list per database and it contains the locks held by all applications concurrently connected to the database.

**D.17. MAXLOCKS**

Default value: AUTOMATIC

This parameter defines a percentage of the lock list held by an application that must be filled before the database manager performs lock escalation.

**D.18. SHEAPTHRES_SHR**

Default value: AUTOMATIC

This parameter represents a soft limit on the total amount of shared sort memory reservation available to sort heap-based operations.

**D.19. SORTHEAP**

Default value: AUTOMATIC

This parameter defines the maximum number of private or shared memory pages that an operation that requires sort heap memory allocates.

**D.20. DATABASE_MEMORY**

Default value: AUTOMATIC

This parameter specifies the size of the database memory set.

**D.21. AUTO_REVAL**

Default value: DEFERRED

This configuration parameter controls the revalidation and invalidation semantics.

**D.22. AUTO_MAINT**

Default value: ON
This parameter is the parent of all the other automatic maintenance database configuration parameters set during HSB setup:

**AUTO_DB_BACKUP**
Default Value: OFF

This automated maintenance parameter enables or disables automatic backup operations for a database.

**AUTO_TBL_MAINT**
Default Value: ON

This parameter is the parent of table maintenance parameters:

**AUTO_RUNSTATS**
Default value: ON

This automated table maintenance parameter enables or disables automatic table RUNSTATS operations for a database.

It is a parent to **AUTO_STMT_STATS**.

**AUTO_STMT_STATS**
Default value: ON

This parameter enables and disables the collection of real-time statistics. It is a child of the auto_runstats configuration parameter.

**AUTO_REORG**
Default value: OFF

This automated table maintenance parameter enables or disables automatic table and index reorganization for a database.

**AUTO_STATS_VIEW**
Default value: ON

This parameter enables and disables automatic statistic collection on statistical views.

**AUTO_SAMPLING**
Default value: ON

This parameter controls whether automatic statistics collection uses sampling when collecting statistics for a large table.
Appendix E. Db2 database backup tool

This tool provides the capability to back up the Db2 database managing the HSB application metadata.

E.1. Features

- **Online** or **Offline** backups
- Backup image compression
- Local or remote secondary backup image location
- Uses syslog
- File system space utilization check

E.2. Usage

Usage: db2_fullbackup.ksh

- **Required:**
  - -i <DB2 instance owner>
  - -p <primary backup directory>
- **Optional:**
  - -s <secondary backup directory>

Examples:

```
> db2_fullbackup.ksh -p /db2_backup1 -s /db2_backup2 -i treefrog
```
```
db2_fullbackup.ksh -p /db2_backup1 -s hm2:/db2_backup2 -i treefrog
```
```
db2_fullbackup.ksh -p /db2_backup1 -i treefrog
```

All databases in the database directory will be backed up.

If a database has been activated an online backup will be performed, otherwise an offline backup will be performed.

If the secondary backup is on another host, use <host>:<directory> format. The `scp` command will be used to copy the database backup image from the primary backup
location to the secondary location on the remote host. It is recommended that the HSB instance owner user id have password-less ssh configured between the HSB server and the remote host.

E.3. Installation location

/var/treefrog/bin/*db2_fullbackup.ksh*
Appendix F. Create TLS certificates

The steps in the following sections provide an example of how to create the certificates necessary for securely communicating with an openLDAP directory service using TLS or SSL. These steps are just an example and will vary depending on the type of directory service being used and the version. These steps are helpful when trying to configure a security domain using the administrative GUI.

F.1. LDAP server configuration steps

1. Create a temporary directory:

   > mkdir /tmp/certs
   > cd /tmp/certs

2. Create the CA private key:

   > openssl genrsa -des3 -out ldapclient-key.pem 2048
   Generating RSA private key, 2048 bit long modulus
   .............................................
   ...........................................+++
   .............................................
   e is 65537 (0x10001)
   Enter pass phrase for ldapclient-key.pem:
   Verifying - Enter pass phrase for ldapclient-key.pem:

3. Create the CA certificate:

   > openssl req -new -key ldapclient-key.pem -x509 -days 1095
   -out ldapservcer-cacerts.pem
   Enter pass phrase for ldapclient-key.pem:
   You are about to be asked to enter information that will be incorporated
   into your certificate request.
   What you are about to enter is what is called a Distinguished Name or a DN.
   There are quite a few fields but you can leave some blank
   For some fields there will be a default value,
   If you enter ".", the field will be left blank.
   -----BEGIN CERTIFICATE-----
   Country Name (2 letter code) [XX]:US
   State or Province Name (full name) []:Texas
   Locality Name (eg, city) [Default City]:Houston
   Organization Name (eg, company) [Default Company Ltd]:IBM
   Organizational Unit Name (eg, section) []:HPSS
   Common Name (eg, your name or your server's hostname) []:linuxg1CA.clearlake.ibm.com
   Email Address []:root@linuxg1.clearlake.ibm.com
   -----END CERTIFICATE-----

4. Create a server private key:

   > openssl genrsa -out ldapservcer-key.pem
   Generating RSA private key, 2048 bit long modulus
   .............................................
   .............................................+++
   .............................................+++
5. Create a certificate signing request:

```bash
> openssl req -new -key ldapserver-key.pem -out server.csr
```

You are about to be asked to enter information that will be incorporated into your certificate request.
What you are about to enter is what is called a Distinguished Name or a DN.
There are quite a few fields but you can leave some blank
For some fields there will be a default value,
If you enter ".", the field will be left blank.
```
-----
Country Name (2 letter code) [XX]:US
State or Province Name (full name) []:Texas
Locality Name (eg, city) [Default City]:Houston
Organization Name (eg, company) [Default Company Ltd]:IBM
Organizational Unit Name (eg, section) []:HPSS
Common Name (eg, your name or your server's hostname) []:linuxg1.clearlake.ibm.com
Email Address []:root@linuxg1.clearlake.ibm.com
```

Please enter the following 'extra' attributes to be sent with your certificate request
A challenge password []:
An optional company name []:

6. Create the server certificate (signed by your own CA certificate):

```bash
> openssl x509 -req -days 2000 -in server.csr -CA ldapserver-cacerts.pem -CAkey ldapclient-key.pem -CAcreateserial -out ldapserver-cert.pem
```

Signature ok
subject=/C=US/ST=Texas/L=Houston/O=IBM/OU=HPSS/CN=linuxg1.clearlake.ibm.com/
emailAddress=root@linuxg1.clearlake.ibm.com
Getting CA Private Key
Enter pass phrase for ldapclient-key.pem:

7. Create and copy the CA certificate, server certificate, and server private key to the final location:

```bash
> mkdir /etc/openldap/certs
> cp ldapserver-cacerts.pem /etc/openldap/certs
> cp ldapserver-cert.pem /etc/openldap/certs
> cp ldapserver-key.pem /etc/openldap/certs
> chown ldap:ldap /etc/openldap/certs/*
```

8. Copy the CA certificate to each client:

```bash
> scp ldapserver-cacerts.pem <ldap client>:/tmp/ldapclient-cacerts.pem
```

9. Create an input file for ldapmodify to modify the LDAP configuration:

```bash
> cat add-tls.ldif
```

```bash
dn: cn=config
changetype: modify
add: olcTLSCACertificateFile
```

```bash
olcTLSCACertificateFile: /etc/openldap/certs/ldapserver-cacerts.pem
```

```bash
add: olcTLSCertificateKeyFile
```

```bash
```
Create TLS certificates

Create TLS certificates:

```
olcTLSCertificateFile: /etc/openldap/certs/ldapserver-cert.pem
   - add: olcTLSCertificateFile
olcTLSPrivateKeyFile: /etc/openldap/certs/ldapserver-key.pem
   - add: olcTLSPrivateKeyFile
olcTLSCipherSuite
   - olcTLSCipherSuite: HIGH:MEDIUM:+SSLv2
```

10. Modify the LDAP configuration:

```
> ldapmodify -Y EXTERNAL -H ldapi:/// -f add-tls.ldif
```

If the variables in the input file already exist in the LDAP configuration, just update the
input file and replace "add:" with "replace:".

11. Restart the LDAP server:

```
> systemctl restart slapd
```

12. Remove the temporary directory:

```
> cd /tmp
> rm -rf /tmp/certs
```

F.2. LDAP client configuration steps

1. Copy the CA certificate to the final location:

```
> cp /tmp/ldapclient-cacerts.pem /etc/openldap/cacerts/ldapclient-cacerts.pem
```

The permissions on the certificate should be set to read for all.

F.3. HSB security domain configuration steps

1. The "Host Name" in the HSB security domain should match the host name used configuring the
server certificate. For example, if the "Common Name" specified during certificate creation is
"linuxg1.clearlake.ibm.com" then the "Host Name" should be "linuxg1.clearlake.ibm.com".

2. The "Host Port" in the HSB security domain should indicate the ldaps port number. Port 636 is the
standard port for LDAP TLS and SSL communication.

3. The "Allow Self-Signed Certificates" checkbox should be unchecked.

4. The "CA Certificate Path" should be the CA certificate file path. For example, `/etc/openldap/
cacerts/ldapclient-cacerts.pem`.
Appendix G. References

1. HPSS Error Messages Reference Manual, current release
2. HPSS User’s Guide, current release
4. HPSS Storage Broker Installation Guide, current release
5. HPSS Storage Broker Administrator’s Guide, current release
6. SNIA Self-contained Information Retention Format (SIRF v1.0)
Appendix H. Developer acknowledgments

HPSS Storage Broker is a product of a government-industry collaboration. The project approach is based on the premise that no single company, government laboratory, or research organization has the ability to confront all of the system-level issues that must be resolved for significant advancement in high-performance storage system technology.

HPSS Storage Broker development was performed jointly by IBM Worldwide Government Industry, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, NASA Langley Research Center, Oak Ridge National Laboratory, and Sandia National Laboratories.