

1. DuraCloud Release 2.1	2
1.1 DuraCloud Chunker Tool	2
1.2 Logging Configuration	5
1.3 DuraCloud REST API	6
1.3.1 REST API Examples Using curl	26
1.4 Release Notes	34
1.5 DuraCloud Sync Tool	41
1.6 DuraCloud Java Clients	47
1.7 DuraCloud Administration	48
1.8 DuraCloud Services	49
1.9 DuraCloud Features	54
1.10 DuraCloud Retrieval Tool	55
1.11 DuraCloud Stitcher Tool	59
1.12 DuraCloud Upload Tool	60
1.12.1 Upload Tool Help	61
1.13 Known Issues	63
1.14 Building DuraCloud Software from Source	64
1.15 Deploying DuraCloud from Binaries	70
1.16 DuraCloud Security	71
1.17 Service Development Guide	77

DuraCloud Release 2.1

User Documentation

- [Release Notes](#)
- [Known Issues](#)
- [DuraCloud Features](#)
- [DuraCloud Services](#)
- [DuraCloud REST API](#)
- [DuraCloud Security](#)
- [DuraCloud Java Clients](#)
- [DuraCloud Upload Tool](#)
- [DuraCloud Sync Tool](#)
- [DuraCloud Retrieval Tool](#)
- [DuraCloud Chunker Tool](#)
- [DuraCloud Stitcher Tool](#)
- [DuraCloud Administration](#)

Developer Documentation

- [Deploying DuraCloud from Binaries](#)
- [Building DuraCloud from Source](#)
- [Service Development Guide](#)
- [Logging Configuration](#)

DuraCloud Chunker Tool

Introduction

The Chunker Tool is a utility which was created in order to provide a simple way to copy files from a local file system to DuraCloud in a "one-off" manner. Actually, although the common case is to use this tool to copy one or more files to DuraCloud, it may also be run to copy files to another location on the local file system.

Download

[Download the Chunker Tool from the Downloads page.](#)

Operational notes

- If you want to jump directly into using the tool, download it from the link above and run the following command

```
java -jar chunk-{version}-driver.jar
```

The resulting usage statement (detailed below) should be enough to help you get started.

- The Chunker Tool allows you to copy multiple local files and directories into a single space within DuraCloud. The names of the objects which are added to DuraCloud will contain all of the directory elements in the path starting from the first element below the base directory down to the individual file names.

Using the Chunker Tool

- To run the Chunker Tool, open a terminal or command prompt and navigate to the directory where the Chunker Tool is located and run the above command.
- The following options are available when running the Chunker Tool

Short Option	Long Option	Arguments	Description
-a	--add	<f t s>	<p>add content from directory:<f> to space or directory:<t> of maximum chunk size:<s>, where the chunk size must have a unit suffix of K,M, or G</p> <p>—</p> <p>If the -c option is provided, the destination space <t> will be interpreted as the name of a space in the DuraCloud account found at the host:port provided in the -c option, otherwise the destination space will be interpreted as a directory on the local file system.</p>
-c	--cloud-store	<host:port>	<p>use cloud store found at <host>:<port> as content destination</p>
-d	--dir-filter	<l>	<p>limit processed directories to those listed in file-list:<l></p> <p>—</p> <p>If the -d option is not used, all directories under the base source directory provided in the -a option will be included. The file specified by this option is expected to contain a list of directory names each on there own line. The list is converted to an OrFileFilter from Apache Commons IO</p>

-f	--file-filter	<l>	<p>limit processed files to those listed in file-list:<l></p> <p>—</p> <p>The file specified by this option is expected to contain a list of file names each on there own line. The list is converted to an OrFileFilter from Apache Commons IO</p>
-g	--generate	<outFile numBytes>	<p>generate test data to <outFile> of <size> bytes</p> <p>—</p> <p>This option does not copy any files, it only generates test data files of the size specified in the give argument.</p>
-i	--ignore-large-files	no args	<p>if this option is set, files over the chunk size specified in the 'add' option will be ignored.</p>
-p	--password	<password>	<p>password of duracloud instance</p>
-u	--username	<username>	<p>username of duracloud instance</p>
-x	--exclude-chunk-md5s	no args	<p>if this option is set, chunk MD5s will NOT be preserved in the manifest</p> <p>—</p> <p>It is expected that this option is rarely used, but in certain situations where the MD5s of the segments of a file that needed to be chunked because the parent file was larger than the limit set in the -a option, not generating these MD5s improves performance.</p>

Creating your own Chunks

If you are interested in creating chunked files in DuraCloud using your own tools, you may do so by adhering to the XML schema used by DuraCloud to create chunks.

[Download the Chunker XSD from the Downloads page](#)

Logging Configuration

Introduction

The logging framework used in the DuraCloud application is [SLF4J](#) with the [LogBack](#) implementation statically bound at runtime.

See the [LogBack](#) website for a detailed description of the configuration options.

The application also contains bridges for both Log4J and Commons-Logging which translates any underlying, dependency libraries which are configured to write to these frameworks into the SLF4J API. The effect is that all logging is channeled through the SLF4J configuration.

General Usage

- By default, if no configuration file is found by LogBack, the logging level is set to "DEBUG" and the appender is set to "STDOUT"
- When starting any DuraCloud application, a LogBack configuration file may be specified by using the following system variable

```
java -Dlogback.configurationFile={path-to-logging-configuration-file}
-jar any-application
```

- Additionally, LogBack will use the file named "logback.xml" found at the top of the classpath for configuration
- [An example logback.xml file can be found on the Downloads page](#)

```

1<?xml version="1.0" encoding="UTF-8"?>
2
3<configuration >
4  <!--<configuration debug="true" scan="true">-->
5  <jmxConfigurator/>
6  <property name="LOG_FILENAME"
value="/home/duraspace/logs/duracloud-osgi.log" />
7
8  <appender name="DURACLOUD"
class="ch.qos.logback.core.rolling.RollingFileAppender">
9    <File>${LOG_FILENAME}</File>
10   <encoder>
11     <pattern>%-14p %d{yyyy/MM/dd HH:mm:ss} [%t] (%F:%L\\) [%M(\\)]
- %m%n</pattern>
12   </encoder>
13   <rollingPolicy
class="ch.qos.logback.core.rolling.FixedWindowRollingPolicy">
14     <maxIndex>5</maxIndex>
15     <FileNamePattern>${LOG_FILENAME}.%i</FileNamePattern>
16   </rollingPolicy>
17   <triggeringPolicy
class="ch.qos.logback.core.rolling.SizeBasedTriggeringPolicy">
18     <MaxFileSize>20MB</MaxFileSize>
19   </triggeringPolicy>
20 </appender>
21 <appender name="STDOUT"
class="ch.qos.logback.core.ConsoleAppender">
22   <encoder>
23     <pattern>%-14p %d{yyyy/MM/dd HH:mm:ss} [%t] (%F:%L\\) [%M(\\)]
- %m%n</pattern>
24   </encoder>
25 </appender>
26 <logger name="org.duracloud" level="DEBUG" additivity="false">
27   <appender-ref ref="DURACLOUD"/>
28 </logger>
29 <root level="WARN">
30   <appender-ref ref="STDOUT"/>
31 </root>
32</configuration>

```

- Notes on the above logback.xml file
 - on line 4, the attribute "debug" applies to displaying configuration information when LogBack starts up if set to "true"
 - on line 4, the attribute "scan" configures LogBack to re-read the given logback.xml every 60 seconds (by default) for updates
 - on line 26, the attribute "additivity" configures the given logger to inherit the configuration of the parent logger, in this case, the root logger
 - on line 26, if the "additivity" attribute were set to "true", all "DURACLOUD" log output would also log to "STDOUT"


DuraCloud REST API


DuraCloud REST API methods:


- [Notes](#)
- [All Applications](#)
 - [Initialize Security Users](#)
- [DuraStore](#)
 - [Initialize Stores](#)
 - [Is Initialized](#)
 - [Get Stores](#)
 - [Get Spaces](#)
 - [Get Space](#)
 - [Get Space Properties](#)
 - [Get Space ACLs](#)
 - [Create Space](#)
 - [Set Space Properties](#)
 - [Set Space ACLs](#)
 - [Delete Space](#)
 - [Get Content](#)
 - [Get Content Properties](#)
 - [Store Content](#)
 - [Copy Content](#)
 - [Set Content Properties](#)
 - [Delete Content](#)
 - [Get Tasks](#)
 - [Perform Task](#)
 - [Tasks](#)
- [DuraService](#)
 - [Initialize Services](#)
 - [Is Initialized](#)
 - [Get Services](#)
 - [Get Service](#)
 - [Get Deployed Service](#)
 - [Get Deployed Service Properties](#)
 - [Deploy Service](#)
 - [Update Service Configuration](#)
 - [UnDeploy Service](#)
- [DurAdmin](#)
 - [Initialize Application](#)
 - [Is Initialized](#)
- [DuraBoss](#)
 - [Initialize Application](#)
 - [Is Initialized](#)
 - [Get Latest Storage Report](#)
 - [Get Storage Report List](#)
 - [Get Storage Report](#)
 - [Get Storage Report Info](#)
 - [Start Storage Report](#)
 - [Cancel Storage Report](#)
 - [Schedule Storage Report](#)

- [Cancel Storage Report Schedule](#)
- [Get Deployed Services Report](#)
- [Get Completed Services Report](#)
- [Get Completed Services Report List](#)
- [Get Services Report](#)
- [Get Executor Status](#)
- [Get Supported Executor Actions](#)
 - [Executor Actions](#)
- [Perform an Executor Action](#)
- [Shutdown Executor](#)
- [Create Initial Audit Log](#)
- [Get Audit Logs](#)
- [Shutdown Auditor](#)
- [Get Content Manifest](#)

Notes

 Each of the methods below has specific security requirements. See [DuraCloud Security](#) for more information

 Due to [an issue](#) which does not properly handle requests redirected from http to https, it is recommended that all REST API requests use https directly.

 Examples calling the API defined below with the Unix utility "curl" can be found [here](#)

All Applications

Security Initialization REST Methods

Initialize Security Users

- Purpose: Allows the initialization of authorized users
- Request:
 - POST <https://host:port/durastore/security>
 - POST <https://host:port/duraservice/security>
 - POST <https://host:port/duradmin/security>
 - POST <https://host:port/durareport/security>
- Request Body: XML similar to:

```
<?xml version="1.0" encoding="UTF-8"?>
<dur:security-users schemaVersion="0.2" xmlns:dur="duracloud.org">
  <security-user>
    <username>username-0</username>
    <password>password-0</password>
    <enabled>true</enabled>
    <accountNonExpired>true</accountNonExpired>
    <credentialsNonExpired>true</credentialsNonExpired>
    <accountNonLocked>true</accountNonLocked>
    <grantedAuthorities>ROLE_USER</grantedAuthorities>
  </security-user>
  <security-user>
    <username>username-1</username>
    <password>password-1</password>
    <enabled>false</enabled>
    <accountNonExpired>false</accountNonExpired>
    <credentialsNonExpired>false</credentialsNonExpired>
    <accountNonLocked>false</accountNonLocked>
    <grantedAuthorities>ROLE_USER ROLE_ADMIN</grantedAuthorities>
  </security-user>
</dur:security-users>
```

- Response Code: 200 (on success)
- Response Body: "Initialization Successful" (on success)

DuraStore

Purpose: DuraStore is the application through which DuraCloud manages storage. The DuraStore REST API provides access to storage by mediating the underlying storage provider APIs to allow access to multiple cloud storage options through a single API.

Initialization REST Methods

Initialize Stores

- Purpose: Allows the initialization of storage provider accounts
- Request: POST <https://host:port/durastore/init>
- Request Body: XML similar to:

```
<storageProviderAccounts>
  <storageAcct ownerId='0' isPrimary='true'>
    <id>1</id>
    <storageProviderType>AMAZON_S3</storageProviderType>
    <storageProviderCredential>
      <username>username</username>
      <password>password</password>
    </storageProviderCredential>
  </storageAcct>
</storageProviderAccounts>
```

- Response Code: 200 (on success)
- Response Body: "Initialization Successful" (on success)

Is Initialized

- Purpose: Performs a check to determine if the DuraStore application has been initialized
- Request: GET <https://host:port/durastore/init>
- Response Code: 200 (if the application has been initialized), 503 (if the application has NOT been initialized)
- Response Body: Text indicating whether initialization has occurred.

Store REST Methods

Get Stores

- Purpose: Provides a listing of available storage providers accounts (without credentials)
- Request: GET <https://host:port/durastore/stores>
- Parameters: None
- Response Code: 200 (on success)
- Response Body: XML similar to:

```
<storageProviderAccounts>
  <storageAcct isPrimary='true'>
    <id>1</id>
    <storageProviderType>AMAZON_S3</storageProviderType>
  </storageAcct>
  <storageAcct isPrimary="false">
    <id>2</id>
    <storageProviderType>RACKSPACE</storageProviderType>
  </storageAcct>
</storageProviderAccounts>
```

Space REST Methods

Get Spaces

- Purpose: Provides a listing of all of the spaces that a customer has created
- Request: GET <https://host:port/durastore/spaces> ? (storeID)
- Response Code: 200 (on success)
- Response Body: XML similar to:

```
<spaces>
  <space id="space1" />
  <space id="space2" />
</spaces>
```

Get Space

- Purpose: Provides a listing of the contents of a space along with space properties
- Request: GET <https://host:port/durastore/spaceID> ? (storeID) (prefix) (maxResults) (marker)
 - storeID (optional) - ID of the content storage provider to query (default is primary store)
 - prefix (optional) - Only retrieve content ids with this prefix (default is all content ids)
 - maxResults (optional) - The maximum number of content IDs to return in the list (default is 1000)

note: the maximum allowable value for maxResults is 1000. Any larger value will be reduced to 1000.
 - marker (optional) - The content ID marking the last item in the previous set (default is the first set of ids)
- Response Code: 200 (on success)
- Response Body: XML similar to:

```
<space id="space1">
  <item>Image 1</item>
  <item>Image 2</item>
</space>
```

- Response Headers: All available space properties, example:

```
x-dura-meta-space-count: 65
x-dura-meta-space-access: OPEN
x-dura-meta-space-created: Mon, 01 Jan 2000 08:00:00 EST
x-dura-meta-custom-property: Custom Property Value
```

Get Space Properties

- Purpose: Provides all space properties
- Request: HEAD <https://host:port/durastore/spaceID> ? (storeID)
- Response Code: 200 (on success)
- Response Headers: Same as for Get space (above)

Get Space ACLs

- Purpose: Provides all space ACLs, with values of 'r' (read) and 'w' (read/write)
- Request: HEAD <https://host:port/durastore/acl/spacelD> ? (storeID)
- Response Code: 200 (on success)
- Response Headers: All available space ACLs, example:

```
x-dura-meta-acl-user0: WRITE
x-dura-meta-acl-user1: WRITE
x-dura-meta-acl-group-curators: READ
```

Create Space

- Purpose: Creates a new space
- Request: PUT <https://host:port/durastore/spacelD> ? (storeID)
- Request Headers: Properties about the space, example:

```
x-dura-meta-space-access: OPEN
x-dura-meta-custom-property: Custom Property Value
```

- Response Code: 201 (on success)
- Response Headers: Location of the new space (i.e. the URL used to create the space), example:

```
Location: https://myhost:8080/durastore/spacel
```

Set Space Properties

- Purpose: Updates the properties associated with a space
- Request: POST <https://host:port/durastore/spacelD> ? (storeID)
- Request Headers: Same as Create space (above)
- Response Code: 200 (on success)
- Response Body: "Space \$spacelD updated successfully" (on success)

Set Space ACLs

- Purpose: Updates the ACLs associated with a space
- Request: POST <https://host:port/durastore/acl/spacelD> ? (storeID)
- Request Headers: For 'user' ACLs the header prefix must be 'x-dura-meta-acl-' and for 'groups' the header prefix must be 'x-dura-meta-acl-group-'. Allowable values for ACL headers are: 'READ' and 'WRITE'.

Example:

```
x-dura-meta-acl-user0: WRITE
x-dura-meta-acl-user1: WRITE
x-dura-meta-acl-group-curators: READ
```

- Response Code: 200 (on success)
- Response Body: "Space \$spacelD ACLs updated successfully" (on success)

Delete Space

- Purpose: Deletes a space
- Request: DELETE <https://host:port/durastore/spaceID> ? (storeID)
- Response Code: 200 (on success)
- Response Body: "Space \$spaceID deleted successfully" (on success)

Content REST Methods

Get Content

- Purpose: Retrieves a piece of content along with its properties
- Request: GET <https://host:port/durastore/spaceID/contentID> ? (storeID) (attachment)
 - if attachment param value is true, a Content-Disposition header is included with the response
- Response Code: 200 (on success)
- Response Body: The content stream
- Response Headers: All available content properties, example:

```
Content-Type: text/plain
Content-Length: 5732
Content-MD5: 3456709234785097473839202
ETag: 3456709234785097473839202
x-dura-meta-content-name: Testing Content
x-dura-meta-content-owner: JSmith
```

Get Content Properties

- Purpose: Retrieves the properties of a piece of content without the content itself
- Request: HEAD <https://host:port/durastore/spaceID/contentID> ? (storeID)
- Response Code: 200 (on success)
- Response Headers: Same as Get content (above)

Store Content

- Purpose: Adds a piece of content to the store
- Request: PUT <https://host:port/durastore/spaceID/contentID> ? (storeID)
- Request Body: Content to be added
- Request Headers: Properties about the content, example:

```
Content-Type: text/plain
Content-MD5: 4cd56e137a93alaccb43c5d32f4afffb
x-dura-meta-content-name: Testing Content
x-dura-meta-content-owner: JSmith
```

- Response Code: 201 (on success)
- Response Headers:
 - MD5 checksum of stored content
 - ETag of stored content
 - Location of the new content (i.e. the URL used to create the content), example:

```
Content-MD5: 4cd56e137a93alaccb43c5d32f4afffb
ETag: 4cd56e137a93alaccb43c5d32f4afffb
Location: https://myhost:8080/durastore/space1/content1
```

- Usage Notes
 - When the optional Content-MD5 header is included, the final checksum of the stored file is compared against the MD5 value included in the header to ensure that the file was stored correctly. If the header is not included, an MD5 checksum is computed as the file is transferred to storage, and that value is used in the final comparison.

Copy Content

- Purpose: Copies a piece of content from a source space to a destination space within a given store
- Request: PUT <https://host:port/durastore/spaceID/contentID> ? (storeID)
- Request Body: must not exist
- Request Headers: Copy source, example:

```
x-dura-meta-copy-source: space-id/content-id
```

- Optional Request Headers: Copy source store, example:

```
x-dura-meta-copy-source-store: storeId
```

- Response Code: 201 (on success)
- Response Headers:
 - MD5 checksum of stored content
 - ETag of stored content
 - Location of the new content (i.e. the URL used to create the content), example:

```
Content-MD5: 4cd56e137a93alaccb43c5d32f4afffb
ETag: 4cd56e137a93alaccb43c5d32f4afffb
Location: https://myhost:8080/durastore/space1/content1
```

- Usage Notes
 - The properties associated with the source content item are copied to the destination content item.
 - The source and destination spaces may be the same.
 - Including the optional header indicates that the copy action should retrieve the source file from a space in the specified storage provider. This allows for copying a file from one storage provider to another.


Set Content Properties

- Purpose: Updates the properties associated with a piece of content
- Request: POST <https://host:port/durastore/spaceID/contentID> ? (storeID)
- Request Headers: Same as Store content (above)
- Response Code: 200 (on success)
- Response Body: "Content \$contentID updated successfully"

Delete Content

- Purpose: Removes a piece of content from the store
- Request: DELETE <https://host:port/durastore/spaceID/contentID> ? (storeID)
- Response Code: 200 (on success)
- Response Body: "Content \$contentID deleted successfully"

Task REST Methods

-  Tasks are used to perform storage provider actions which cannot be performed in a generic manner across multiple providers.

Get Tasks

- Purpose: Provides a listing of all of the supported tasks for a given provider. Note that if no storeID parameter is included, the task listing is provided for the primary storage provider.
- Request: GET <https://host:port/durastore/task> ? (storeID)
- Response Code: 200 (on success)
- Response Body: XML similar to:

```
<list>
  <string>task1</string>
  <string>task2</string>
</list>
```

Perform Task

- Purpose: Performs a particular task. Note that most tasks can be performed by only one storage provider type.
- Request: POST <https://host:port/durastore/task/taskName> ? (storeID)
- Request Body: Parameters for task. Each task will expect parameters in a specific format, see task listing for more details.
- Response Code: 200 (on success)
- Response Body: Response value for task, format varies by task.

Tasks

taskName	Storage Provider	Name	Description	Request Body	Response Body
enable-streaming	Amazon S3	Enable Streaming task	Enables RTMP streaming for all files within a DuraCloud space through the use of Amazon's Cloudfront streaming capability. This task may take up to 15 minutes to complete.	Name of the space for which streaming is to be enabled	Text indicating the results of the task, including the streaming host

disable-streaming	Amazon S3	Disable Streaming task	Disables streaming by removing the ability for Cloudfront to access files within a space. This does not remove the streaming distribution, only disables its use, so enabling streaming on the same space again can be performed much more quickly. Some content in the space may continue to be available for streaming up to 24 hours after streaming has been disabled.	Name of the space for which streaming is to be disabled	Text indicating the results of the task
delete-streaming	Amazon S3	Delete Streaming task	Removes a streaming distribution created by the enable-streaming task. This task should be performed after performing the disable-streaming task. This task may take up to 15 minutes to complete, after which no content in the space will be available for streaming.	Name of the space for which streaming is to be deleted	Text indicating the results of the task

run-hadoop-job	Amazon S3	Run Hadoop Job task	Runs a hadoop job using Amazon's Elastic Map Reduce feature. A JAR which implements the hadoop interfaces is expected to have already been loaded into S3. This JAR is used to execute the hadoop job.	A map serialized into XML which includes, at a minimum, values for jarContentId, sourceSpaceId, destSpaceId, and workSpaceId.	A map serialized into XML which includes the jobFlowId
describe-hadoop-job	Amazon S3	Describe Hadoop Job task	Retrieves information about a hadoop job running in Amazon's Elastic Map Reduce	The Job Flow ID	A map serialized into XML which includes information about the running job
stop-hadoop-job	Amazon S3	Stop Hadoop Job task	Stops a hadoop job running in Amazon's Elastic Map Reduce	The Job Flow ID	A map serialized into XML which includes a results key with a value of either success or failure
noop	Amazon S3	Test task	Provides a simple way to test the calling of tasks	Body content is ignored	Text indicating successful task completion

DuraService

Purpose: DuraService is the application through which DuraCloud manages services. The DuraService REST API provides the means by which services available in the DuraCloud service repository are deployed, configured, and undeployed.

Resources: [XML schema which define the service configuration can be found on the Downloads page](#)

Initialization REST Methods

Initialize Services

- Purpose: Initializes the DuraService application
- Request: POST <https://host:port/duraservice/init>
- Request Body: XML similar to:

```
<servicesConfig>
  <primaryServiceInstance>
    <host>[PRIMARY-SERVICE-INSTANCE-HOST]</host>

  <servicesAdminPort>[PRIMARY-SERVICES-ADMIN-PORT]</servicesAdminPort>

  <servicesAdminContext>[PRIMARY-SERVICES-ADMIN-CONTEXT]</servicesAdmin
Context>
  </primaryServiceInstance>
  <userStorage>
    <host>[USER-STORAGE-HOST-NAME]</host>
    <port>[USER-STORAGE-PORT]</port>
    <context>[USER-STORAGE-CONTEXT]</context>
    <msgBrokerUrl>[USER-STORAGE-MSG-BROKER-URL]</msgBrokerUrl>
  </userStorage>
  <serviceStorage>
    <host>[SERVICES-STORAGE-HOST-NAME]</host>
    <port>[SERVICES-STORAGE-PORT]</port>
    <context>[SERVICES-STORAGE-CONTEXT]</context>
    <spaceId>[SERVICES-STORAGE-SPACE-ID]</spaceId>
  </serviceStorage>
  <serviceCompute>
    <type>AMAZON_EC2</type>
    <imageId>[MACHINE-IMAGE-ID]</imageId>
    <computeProviderCredential>
      <username>[USERNAME]</username>
      <password>[PASSWORD]</password>
    </computeProviderCredential>
  </serviceCompute>
</servicesConfig>
```

- Response Code: 200 (on success)
- Response Body: "Initialization Successful" (on success)

Is Initialized

- Purpose: Performs a check to determine if the DuraService application has been initialized
- Request: GET <https://host:port/duraservice/init>
- Response Code: 200 (if the application has been initialized), 503 (if the application has NOT been initialized)
- Response Body: Text indicating whether initialization has occurred.

Service REST Methods

Get Services

- Purpose: Retrieves a listing of services, along with their configuration options
- Request: GET <https://host:port/duraservice/services> ? (show)
 - Parameter options for show (optional)
 1. available (default) - Includes only services which have not been deployed but are available for deployment
 2. deployed - Includes only services which have been deployed and started
- Response Code: 200 (on success)
- Response Body: XML list of services (see service config xsd)

Get Service

- Purpose: Retrieves information about a particular service including description, configuration options, and all deployments
- Request: GET <https://host:port/duraservice/serviceID>
- Response Code: 200 (on success)
- Response Body: XML service (see service config xsd)

Get Deployed Service

- Purpose: Retrieves information about a deployed service including description, configuration options, and a single deployment indicating the configuration options in use
- Request: GET <https://host:port/duraservice/serviceID/deploymentID>
- Response Code: 200 (on success)
- Response Body: XML service (see service config xsd)

Get Deployed Service Properties

- Purpose: Retrieves the runtime properties of a deployed service
- Request: GET <https://host:port/duraservice/serviceID/deploymentID/properties>
- Response Code: 200 (on success)
- Response Body: XML service (simple xml Map serialization)

Deploy Service

- Purpose: Deploys and starts an available service
- Request: PUT <https://host:port/duraservice/serviceID> ? (serviceHost)
 - Parameter value for serviceHost (optional) should indicate the services host on which the service should be deployed. Default is the primary customer host.
- Request Body: XML user configuration indicating the config selections for the service (see user config portion of service config xsd)
- Response Code: 201 (on success)
- Response Header: Location header indicates the URL at which information about the deployed service can be retrieved (the URL for a get deployed service call) which includes the deploymentID

Update Service Configuration

- Purpose: Updates the configuration of a deployed service
- Request: POST <https://host:port/duraservice/serviceID/deploymentID>
- Request Body: Updated XML user configuration indicating the config selections for the service (see user config portion of service config xsd)
- Response Code: 200 (on success)

UnDeploy Service

- Purpose: Stops and Undeploys a deployed service
- Request: DELETE <https://host:port/duraservice/serviceID/deploymentID>
- Response Code: 200 (on success)

DurAdmin

Purpose: DurAdmin is the user-facing application through which DuraCloud exposes DuraStore and DuraService functionality. The DurAdmin REST API provides the means by which DurAdmin is initialized.

Initialization REST Methods

Initialize Application

- Purpose: Allows the initialization of duradmin
- Request: POST <https://host:port/duradmin/init>
- Request Body: XML similar to:

```
<duradminConfig>
  <durastoreHost>[host]</durastoreHost>
  <durastorePort>[port]</durastorePort>
  <durastoreContext>durastore</durastoreContext>
  <duraserviceHost>[host]</duraserviceHost>
  <duraservicePort>[port]</duraservicePort>
  <duraserviceContext>duraservice</duraserviceContext>
</duradminConfig>
```

- Response Code: 200 (on success)
- Response Body: "Initialization Successful" (on success)

Is Initialized

- Purpose: Performs a check to determine if the DurAdmin application has been initialized
- Request: GET <https://host:port/duradmin/init>
- Response Code: 200 (if the application has been initialized), 503 (if the application has NOT been initialized)
- Response Body: Text indicating whether initialization has occurred.

DuraBoss

Purpose: DuraBoss provides administrative control over a variety of activities that run over the storage and services managed by DuraCloud. DuraBoss consists of four major applications:

- Reporter - generates reports relating to the status of the storage and services within DuraCloud
- Executor - manages actions which automate functions within DuraCloud, primarily the scheduling and running of DuraCloud services
- Auditor - maintains audit logs for all spaces within DuraCloud, ensuring that all additions, update, and deletions are recorded and made available
- Manifest - provides manifests in various formats for the content that resides in DuraCloud spaces

Resources: [XML schema which defines the expected transfer data for storage and service reporting can be found on the Downloads page](#)

Initialization REST Methods

Initialize Application

- Purpose: Allows the initialization of duraboss
- Request: POST <https://host:port/duraboss/init>
- Request Body: XML similar to:

```
<durabossConfig>
  <reporterEnabled>[true|false]</reporterEnabled>
  <executorEnabled>[true|false]</executorEnabled>
  <auditorEnabled>[true|false]</auditorEnabled>
  <durastoreHost>[host]</durastoreHost>
  <durastorePort>[port]</durastorePort>
  <durastoreContext>durastore</durastoreContext>
  <duraserviceHost>[host]</duraserviceHost>
  <duraservicePort>[port]</duraservicePort>
  <duraserviceContext>duraservice</duraserviceContext>
  <notificationConfig>
    <type>EMAIL</type>
    <username>[username for notification system]</username>
    <password>[password for notification system]</password>
    <originator>[from email address]</originator>
    <admin>[administrator email address]</admin>
  </notificationConfig>
</durabossConfig>
```

- Response Code: 200 (on success)
- Response Body: "Initialization Successful" (on success)

Is Initialized

- Purpose: Performs a check to determine if the DuraBoss application has been initialized
- Request: GET <https://host:port/duraboss/init>
- Response Code: 200 (if the application has been initialized), 503 (if the application has NOT been initialized)
- Response Body: Text indicating whether initialization has occurred.

Reporter: Storage Report REST Methods

Get Latest Storage Report

- Purpose: Provides the most current storage report in XML format
- Request: GET <https://host:port/duraboss/report/storage>
- Response Code: 200 (on success)
- Response Body: XML, defined by the storage report XSD

Get Storage Report List

- Purpose: Provides a list of all storage report IDs
- Request: GET <https://host:port/duraboss/report/storage/list>
- Response Code: 200 (on success)
- Response Body: XML, defined by the storage report XSD

Get Storage Report

- Purpose: Provides a specific storage report based on the provided report ID
- Request: GET <https://host:port/duraboss/report/storage/reportID>
- Response Code: 200 (on success)
- Response Body: XML, defined by the storage report XSD

Get Storage Report Info

- Purpose: Provides a information about the current status of the storage reporting system
- Request: GET <https://host:port/duraboss/report/storage/info>
- Response Code: 200 (on success)
- Response Body: XML, defined by the storage report XSD

Start Storage Report

- Purpose: Starts a storage report if one is not already running
- Request: POST <https://host:port/duraboss/report/storage>
- Response Code: 200 (on success)
- Response Body: "Report Started" (on success), or ""Report Already In Progress" (if a report is already in progress)

Cancel Storage Report

- Purpose: Cancels a running storage report
- Request: DELETE <https://host:port/duraboss/report/storage>
- Response Code: 200 (on success)
- Response Body: "Storage report cancelled"

Schedule Storage Report

- Purpose: Schedules a time for a storage report to be run
- Request: POST <https://host:port/duraboss/report/storage/schedule> ? (startTime) (frequency)
 - startTime: time (in milliseconds since the epoch) to begin the next storage report
 - frequency: time (in milliseconds) to wait between running reports (minimum value is 600000)
- Response Code: 200 (on success)
- Response Body: "Storage reports scheduled" (on success)

Cancel Storage Report Schedule

- Purpose: Cancels all entries on the storage report schedule
- Request: DELETE <https://host:port/duraboss/report/storage/schedule>
- Response Code: 200 (on success)
- Response Body: "Storage Reports schedule cancelled"

Reporter: Service Report REST Methods

Get Deployed Services Report

- Purpose: Provides a listing of the services which are currently deployed
- Request: GET <https://host:port/duraboss/report/service/deployed>
- Response Code: 200 (on success)
- Response Body: XML, defined by the service report XSDs

Get Completed Services Report

- Purpose: Provides a listing of the most recent completed services
- Request: GET <https://host:port/duraboss/report/service> ? (limit)
 - Parameter value for limit (optional) should indicate the maximum number of services to return in the report (default is 20, max is 1000)
- Response Code: 200 (on success)
- Response Body: XML, defined by the service report XSDs

Get Completed Services Report List

- Purpose: Provides a listing of all service report IDs
- Request: GET <https://host:port/duraboss/report/service/list>
- Response Code: 200 (on success)
- Response Body: XML, defined by the service report XSDs

Get Services Report

- Purpose: Provides a specific services report based on the provided report ID
- Request: GET <https://host:port/duraboss/report/service/reportID>
- Response Code: 200 (on success)
- Response Body: XML, defined by the service report XSDs

Executor REST Methods

Get Executor Status

- Purpose: Provides a status of the Executor, which is the collected status of all Action Handlers
- Request: GET <https://host:port/duraboss/exec>
- Response Code: 200 (on success)
- Response Body: XML serialization of status map

Get Supported Executor Actions

- Purpose: Provides a listing of the actions which the Executor can perform
- Request: GET <https://host:port/duraboss/exec/action>
- Response Code: 200 (on success)
- Response Body: XML serialization of the action set

Executor Actions

Action Name	Description	Request Body
start-bit-integrity	Instructs the bit integrity handler to begin the process at a given time and at a given frequency	start-time,frequency (where start-time is epoch date in milliseconds and frequency is number of milliseconds)
cancel-bit-integrity	Instructs the bit integrity handler to stop performing bit integrity checks and gracefully shut down	None
start-streaming	Starts the streaming service, so that media streaming can begin	None
stop-streaming	Shuts down the streaming service, stops streaming of all media	None
start-streaming-space	Begin streaming all content in the given space	Name of the space to stream
stop-streaming-space	Stop streaming content in the given space	Name of the space to end streaming

Perform an Executor Action

- Purpose: Performs a specific Executor action based on the provided actionName
- Request: POST <https://host:port/duraboss/exec/actionName>
- Response Code: 200 (on success)

Shutdown Executor

- Purpose: Requests that the Executor perform a graceful shutdown
- Request: DELETE <https://host:port/duraboss/exec>
- Response Code: 200 (on success)

Auditor REST Methods

Create Initial Audit Log

- Purpose: Requests the creation of initial audit logs, and removal of any existing audit logs
- Request: POST <https://host:port/duraboss/audit>
- Response Code: 202 (on acceptance)

Get Audit Logs

- Purpose: Provides a listing of the audit logs for the provided spaceId
- Request: GET <https://host:port/duraboss/audit/spaceId>
- Response Code: 200 (on success)
- Response Body: Plain text listing of audit log contentIds

Shutdown Auditor

- Purpose: Requests that the Auditor perform a graceful shutdown
- Request: DELETE <https://host:port/duraboss/audit>
- Response Code: 200 (on success)
- Response Body: Plain text of "auditor shutting down"

Content Manifest REST Methods

Get Content Manifest

- Purpose: Requests the content manifest for the provided spaceId, in the requested format, as of the provided date, and for the provided storeId
- Request: GET <https://host:port/duraboss/manifest/spaceId> ? (format) & (date) & (storeId)
 - Parameter options for format (optional)
 1. tsv (default) - Produces content manifest in tab-separated-value format
 2. bagit - Produces content manifest in [BagIt](#) format
 - Parameter options for date (optional, current datetime is default)
 1. Any formatted date of the form: 'yyyy-MM-dd'T'HH:mm:ss.sss' or 'yyyy-MM' or any form in between these two
 - Parameter options for storeId (optional, default is primary store)
- Response Code: 200 (on success)
- Response Body: Plain text content manifest in the requested format

REST API Examples Using curl

- [Convenience Variables](#)
- [DuraStore Notes](#)
- [DuraStore](#)
 - [Get Stores](#)
 - [Get Spaces](#)
 - [Create Space](#)
 - [Store Content](#)
 - [Get Space](#)
 - [Set Space Properties](#)
 - [Get Space Properties](#)
 - [Get Content](#)
 - [Set Content Properties](#)
 - [Get Content Properties](#)
 - [Delete Content](#)
 - [Delete Space](#)

- [DuraService Notes](#)
- [DuraService](#)
 - [Get Services](#)
 - [Deploy Service](#)
 - [Get Service](#)
 - [Get Deployed Service](#)
 - [Get Deployed Service Properties](#)
 - [Update Service Configuration](#)
 - [UnDeploy Service](#)
- [DuraBoss Report API](#)
 - [Get Latest Storage Report](#)
 - [Get Storage Report List](#)
 - [Get Storage Report Info](#)
 - [Start Storage Report](#)
 - [Stop Storage Report](#)
 - [Schedule Storage Report \(to begin Jan 1, 2020 at 01:01:01 and repeat every 10 min\)](#)
 - [Cancel Storage Report Schedule](#)
 - [Get Deployed Services Report](#)
 - [Get Completed Services Report](#)
 - [Get Services Report List](#)
- [DuraBoss Auditor API](#)
 - [Start the Auditor](#)
 - [Retrieve Log for a Space](#)
 - [Stop the Auditor](#)
- [DuraBoss Executor API](#)
 - [Init Status](#)
 - [Get Executor Status](#)
 - [Get Executor's Supported Actions](#)
 - [Start/Stop Streaming Service](#)
 - [Start/Stop Streaming on a Particular Space](#)
 - [Start Bit Service Integrity](#)
 - [Cancel Bit Integrity Service](#)
- [DuraBoss Manifest API](#)
 - [Get Default Manifest](#)
 - [Get Default Manifest with Specific Format \(BagIt\)](#)
 - [Get Default Manifest with Specific store-id](#)
 - [Get Default Manifest with Specific As-Of-Date](#)

Convenience Variables

The curl commands below can be used directly if you define the following variables in your shell

```

host=<duracloud-hostname>
space-0=<any-name>
space-1=<any-name>
user=<username>
pwd=<password>
file=<any-file-name>

```

DuraStore Notes

The curl commands in the DuraStore section expect a test file for uploads.

```
echo hello > ${file}
```

Note that if the target of a content or space retrieval (GET) has access permissions set to "OPEN", then the "-u" option in the curl commands is not required.

DuraStore

Get Stores

```
curl -u ${user}:${pwd} https://${host}/durastore/stores
```

Get Spaces

```
curl -u ${user}:${pwd} https://${host}/durastore/spaces  
curl -u ${user}:${pwd} https://${host}/durastore/spaces?storeID=1
```

Create Space

```
curl -u ${user}:${pwd} -X PUT https://${host}/durastore/${space-0}  
curl -u ${user}:${pwd} -H "x-dura-meta-city: arlington" -H  
"x-dura-meta-state: va" -X PUT  
https://${host}/durastore/${space-1}?storeID=1
```

Store Content

```
curl -u ${user}:${pwd} -T ${file}  
https://${host}/durastore/${space-0}/test.txt  
curl -u ${user}:${pwd} -T ${file}  
https://${host}/durastore/${space-0}/item.txt
```

Get Space

```
curl -u ${user}:${pwd} https://${host}/durastore/${space-0}  
curl -u ${user}:${pwd} https://${host}/durastore/${space-1}?storeID=1  
curl -u ${user}:${pwd} https://${host}/durastore/${space-0}?prefix=test
```

Set Space Properties

```
curl -u ${user}:${pword} -H "x-dura-meta-country: usa" -X POST
https://${host}/durastore/${space-0}
```

Get Space Properties

```
curl -u ${user}:${pword} -I https://${host}/durastore/${space-0}
curl -u ${user}:${pword} -I https://${host}/durastore/${space-1}?storeID=1
```

Get Content

```
curl -u ${user}:${pword} https://${host}/durastore/${space-0}/test.txt
curl -u ${user}:${pword}
https://${host}/durastore/${space-0}/test.txt?storeID=0&attachment=true
```

Set Content Properties

```
curl -u ${user}:${pword} -X POST -H "x-dura-meta-color: green"
https://${host}/durastore/${space-0}/test.txt
```

Get Content Properties

```
curl -u ${user}:${pword} -I https://${host}/durastore/${space-0}/test.txt
```

Delete Content

```
curl -u ${user}:${pword} -X DELETE
https://${host}/durastore/${space-0}/test.txt
```

Delete Space

```
curl -u ${user}:${pword} -X DELETE https://${host}/durastore/${space-0}
curl -u ${user}:${pword} -X DELETE
https://${host}/durastore/${space-1}?storeID=1
```

The commands in the DuraService section below demonstrate the deployment, inspection, reconfiguration, and undeployment of a service.

As of release 0.8.0, the Bit Integrity Checker has service id of '0' (noted in variable below).

The actual deployment id is dynamically generated based on the number of previous deployments within a given DuraCloud application.

After deploying the service, use the "Get Deployed Service" call to determine the specific *deployment id* of your deployed service.

```
bitintegrity=0
deployment=<determined-by-inspection>
```

The following two service configuration files are provided for deployment and reconfiguration of the Bit Integrity Checker.

They are based on the spaces that were created in the DuraStore section, but can be modified to indicate other configuration options or to execute over alternate spaces.

1. [configuration file 0](#)
2. [configuration file 1](#)

DuraService

Get Services

```
curl -u ${user}:${pword} https://${host}/duraservice/services
curl -u ${user}:${pword} https://${host}/duraservice/services?show=deployed
```

Deploy Service

```
curl -u ${user}:${pword} -X PUT -T deploy.xml
https://${host}/duraservice/${bitintegrity}
```

Get Service

```
curl -u ${user}:${pword} https://${host}/duraservice/${bitintegrity}
```

Get Deployed Service

```
curl -u ${user}:${pword}
https://${host}/duraservice/${bitintegrity}/${deployment}
```

Get Deployed Service Properties

```
curl -u ${user}:${pword}
https://${host}/duraservice/${bitintegrity}/${deployment}/properties
```

Update Service Configuration

```
curl -u ${user}:${pword} -X PUT -T deploy-new.xml
https://${host}/duraservice/${bitintegrity}
```

UnDeploy Service

```
curl -u ${user}:${pword} -X DELETE
https://${host}/duraservice/${bitintegrity}/${deployment}
```

DuraBoss Report API

Get Latest Storage Report

```
curl -u ${user}:${pword} https://${host}/duraboss/report/storage
```

Get Storage Report List

```
curl -u ${user}:${pword} https://${host}/duraboss/report/storage/list
```

Get Storage Report Info

```
curl -u ${user}:${pword} https://${host}/duraboss/report/storage/info
```

Start Storage Report

```
curl -u ${user}:${pword} -X POST https://${host}/duraboss/report/storage
```

Stop Storage Report

```
curl -u ${user}:${pword} -X DELETE https://${host}/duraboss/report/storage
```

Schedule Storage Report (to begin Jan 1, 2020 at 01:01:01 and repeat every 10 min)

```
curl -u ${user}:${pword} -X POST
https://${host}/duraboss/report/storage/schedule?startTime=1577840461000\&f
requency=600000
```

Cancel Storage Report Schedule

```
curl -u ${user}:${pword} -X DELETE
https://${host}/duraboss/report/storage/schedule
```

Get Deployed Services Report

```
curl -u ${user}:${pword} https://${host}/duraboss/report/service/deployed
```

Get Completed Services Report

```
curl -u ${user}:${pword} https://${host}/duraboss/report/service
```

Get Services Report List

```
curl -u ${user}:${pword} https://${host}/duraboss/report/service/list
```

DuraBoss Auditor API

Start the Auditor

```
curl -u ${user}:${pword} -X POST https://${host}/duraboss/audit
```

Retrieve Log for a Space

```
curl -u ${user}:${pword} https://${host}/duraboss/audit/${space}
```

Stop the Auditor


```
curl -u ${user}:${pword} -X DELETE https://${host}/duraboss/audit
```

DuraBoss Executor API

Init Status

```
curl -u ${user}:${pword} https://${host}/duraboss/init
```

Get Executor Status

```
curl -u ${user}:${pword} https://${host}/duraboss/exec
```

Get Executor's Supported Actions

```
curl -u ${user}:${pword} https://${host}/duraboss/exec/action
```

Start/Stop Streaming Service

```
curl -u ${user}:${pword} -X POST  
https://${host}/duraboss/exec/start-streaming  
curl -u ${user}:${pword} -X POST  
https://${host}/duraboss/exec/stop-streaming
```

Start/Stop Streaming on a Particular Space

```
curl -u ${user}:${pword} -X POST  
https://${host}/duraboss/exec/start-streaming-space -d "${space}"  
curl -u ${user}:${pword} -X POST  
https://${host}/duraboss/exec/stop-streaming-space -d "${space}"
```

Start Bit Service Integrity

```
curl -u ${user}:${pword} -X POST  
https://${host}/duraboss/exec/start-bit-integrity -d  
"${firstRunInEpochMilliseconds},${periodInMilliseconds}"
```

Cancel Bit Integrity Service

```
curl -u ${user}:${pword} -X POST
https://${host}/duraboss/exec/cancel-bit-integrity
```

DuraBoss Manifest API

Get Default Manifest

```
curl -u ${user}:${pword} https://${host}/duraboss/manifest/space-id
```

Get Default Manifest with Specific Format (BagIt)

```
curl -u ${user}:${pword}
https://${host}/duraboss/manifest/space-id?format=bagit
```

Get Default Manifest with Specific store-id

```
curl -u ${user}:${pword}
https://${host}/duraboss/manifest/space-id?storeID=${storeId}
```

Get Default Manifest with Specific As-Of-Date

```
curl -u ${user}:${pword}
https://${host}/duraboss/manifest/space-id?date=${data}
```

Release Notes

Release 2.1.0

Released: August 9, 2012

The primary features of release 2.1.0 are:

- Duplicate on Change service updates
 - Duplication is now configured at the space level, allowing for much finer grained selection of which content will be automatically copied to a secondary provider, and which provider that will be.
- SDSC Storage Provider
 - The SDSC storage provider connection was tested and improved, in close cooperation with SDSC personnel, to ensure its readiness for production status.
- Security
 - The DuraCloud security configuration has been extended to only allow Administrative users to perform service execution.

For more details about specific changes in release 2.1.0, see the [JIRA issue tracker](#).

Release 2.0.0

Released: April 17, 2012

The primary features of release 2.0.0 are:

- Integrated History Reports
 - Reports illustrating both current and historical views of the content stored in DuraCloud have now been integrated directly into the display for each individual space.
- Automated Service Execution
 - Bit Integrity Checks
 - Bit Integrity checking is now an automated function. Each content item in each space in each storage provider will be checked regularly to verify that it has not changed since it arrived in DuraCloud. No user intervention is required for this to occur. As each space is checked, the display will update to indicate that the bit integrity of the contained content has been verified.
 - Media Serving
 - Streaming content from DuraCloud now requires only a single button click. Each space now provides the option to turn on or off streaming, with no need to configure, deploy, or re-deploy a service.
- Audit Logging
 - Logging is now being generated and captured which tracks the events occurring within DuraCloud. This provides a permanent record of when content is added, updated, or removed.
- Manifest Generation
 - Building on the work of the audit log, a content manifest can now be requested for any space within DuraCloud. This manifest, which can be in multiple formats, describes the content that resides in the space the moment that the manifest is requested.

For more details about specific changes in release 2.0.0, see the [JIRA issue tracker](#).

Release 1.3.1

Released: Jan 20, 2012

The primary features of release 1.3.1 are:

- Improved service reporting
 - The information within service reports can now be viewed in tabular form directly from within the service details area.
 - Items which are considered error cases that occur as part of service execution are now included in an independent error report that is available for viewing upon service completion.
- Simplified public access
 - Setting the contents of a space to be publicly viewable is now accomplished by simply granting read access to the "public" group.
- CloudSync service
 - The latest version of [CloudSync](#), a utility for managing the movement of content between DuraCloud and a Fedora repository, can now be run as a service within DuraCloud

For more details about specific changes in release 1.3.1, see the [JIRA issue tracker](#).

Release 1.3

Released: Dec 14, 2011

The primary features of release 1.3 are:

- Improved access control
 - Administrators can now define access control lists for each space, indicating which users and groups have read or write access to the content within that space.
 - Users now see only the spaces in their spaces listing which they have access to view. This includes all Open spaces, which remain available for public read access.
 - Users now only see options to perform add, edit, or delete actions in spaces where they have write permissions.
- Content copy across providers
 - Files stored in DuraCloud can now be easily copied individually to another storage provider via both the REST API and the web interface.
- Email notification on service completion
 - After a service in DuraCloud completes, an email is now sent to the user who launched the service, notifying them that the service has completed, and providing details about the results of the service.

For more details about specific changes in release 1.3, see the [JIRA issue tracker](#).

Release 1.2

Released: Oct 30, 2011

The primary features of release 1.2 are:

- [Upload Tool](#)
 - Provides a graphical method for transferring large numbers of files to DuraCloud.
 - Allows for the selection of both files and folders for transfer, and presents a visual indication of the transfer progress.
 - Can be run from the DuraCloud UI by choosing the "Add Many Items" button when a space is selected.
 - Can also be download and run locally.
- SDSC Storage Provider
 - An initial beta release of the new storage provider integration for connecting to the SDSC Cloud storage system.

For more details about specific changes in release 1.2, see the [JIRA issue tracker](#).

Release 1.1

Released: Sept 30, 2011

The primary features of release 1.1 are:

- Media Streamer service
 - Service now allows content from multiple spaces to be streamed.
 - Service now recognizes when new content is added to spaces which are being streamed and starts streaming on those files.
- Duplicate on Change service
 - Service updated to provide greater assurance of file transfer and produce an output report which details the actions of the service and the results of those actions.
- Content item copy and rename capability
 - New copy operation is able to copy, move, and rename content items within a space and between spaces.
- Stitch Tool

- New utility for large files which were "chunked" (split into multiple small files) when placed into storage. This new utility will recombine all of the pieces of a file to re-produce the original file in local storage.
- Stitch capabilities incorporated into the Retrieval Tool, allowing any content which has been "chunked" to be reconstituted on retrieval.
- UI updates
 - Improvements to the user interface in order to provide simpler and more useful feedback for actions being performed throughout the application.
- Properties
 - Use of the term 'properties' has replaced 'metadata' to describe the name/value pairs which can be associated with spaces and content. This update helps to clarify the purpose and capability of this attached information.
- Local service repositories
 - Service repositories can now reside within the same storage container as is used by DuraStore, allowing for simpler configuration of stand-alone DuraCloud instances.
- Initialization endpoint
 - New REST API endpoint (/init) for application initialization.
- Tools tab
 - New tools tab on the dashboard provides convenient links for downloading DuraCloud client tools.

For more details about specific changes in release 1.1, see the [JIRA issue tracker](#).

Release 1.0

Released: July 29, 2011

The primary features of release 1.0 are:

- Storage Reporting
 - A new feature which provides detailed information about the amount of data you have stored in DuraCloud, as well as the kinds of data you have stored in DuraCloud. Reports are generated automatically and the information can be viewed using the DuraCloud dashboard or downloaded for processing using new REST API methods.
- Service Reporting
 - A new feature which provides detailed information about services which are currently running and which have run previously in your DuraCloud account. Reports are generated automatically as services are run and the information can be viewed using the DuraCloud dashboard or downloaded for processing using new REST API methods.
- Service Dependencies
 - Any service can now define a dependency on another service.
 - "System" services, which need to be installed prior to other services being deployed, are now installed on-the-fly only when needed.
- Improved service feedback
 - The information provided by running services is now more complete and more consistent with other DuraCloud services.
- Improved character set support
 - Content IDs can now consist of any characters which can be properly encoded using UTF-8 (with the exception of "reserved" characters mentioned [here](#))
- Image Viewer URL stability
 - The URLs for images made available by the Image Server service will now stay consistent across restarts of the service and restarts of the DuraCloud instance.

For more details about specific changes in release 1.0, see the [JIRA issue tracker](#).

Release 0.9

Released: April 27, 2011

The primary features of release 0.9 are:

- The Duplicate on Ingest service is now the Duplicate on Change service.
 - This service now supports all of the same on-ingest features as before, but it now also performs duplication of all update and delete actions as well. This allows the primary and secondary cloud stores to be kept completely in sync.
- The Bulk Bit Integrity Service has been improved.
 - This service has been updated and verified to properly handle spaces with up to 1 million items
 - The second step of the MD5 verification, which used to run locally on the instance, has been moved to hadoop, allowing the service to complete much more quickly for large data sets.
- User management functions have been removed, as they are now performed by the [DuraCloud Management Console](#).
 - As a convenience, administrators are still able to see the list of users and their roles within the DuraCloud Administrator UI.
- Service outputs have been made more consistent.
 - All DuraCloud services which produce an output file now store that file in the *x-service-out* space.
 - Services which produce log files store those logs in the *x-service-work* space.
 - The names of the output files have been made more consistent, making it simpler to determine which files correspond to which service deployment.
- Password security has been improved.
 - All passwords used within DuraCloud are now immediately pushed through a hashing function before being stored, so that no user passwords are transferred or stored as clear text.
- A ServiceClient is now available, to compliment the StoreClient and make it easier to make direct API calls to manipulate DuraCloud services.

For more details about specific changes in release 0.9, see the [JIRA issue tracker](#).

Release 0.8

Released: Jan 26, 2011

The primary features of release 0.8 are:

- Simplified services
 - The listing of services has been better organized, to make finding the service you would like to run simpler.
 - All services now require you to set fewer options, simplifying the deployment process.
 - Bulk services (Image Transformer - Bulk, Bit Integrity Checker - Bulk, and Duplicate on Demand) now provide a *standard* configuration mode which handles the setting of server type, and number of servers used to perform the job, so that you no longer have to make those choices.
 - The output location for services has been set to the *x-service-out* space, which removes the need to set this value for each service, and provides a standard location to look for service output reports.
 - The work location for services has been set to the *x-service-work* space, which removes the need to set this value for each service, and provides a standard location to look for service logs and other run time artifacts.
- More reliable services
 - Several bugs which have caused services to fail have been resolved.
- Sync Tool command line flags now match those offered by the Retrieval Tool.
- UI updates which provide better visual cues for which storage provider is in use.
- A host of bug fixes and small tweaks

For more details about specific changes in release 0.8, see the [JIRA issue tracker](#).

Release 0.7

Released: Oct 28, 2010

The primary features of release 0.7 are:

- A new [Retrieval Tool](#), a companion to the existing [Sync Tool](#), which is a command-line tool for the retrieving content from DuraCloud spaces.
- A new [Bulk Bit Integrity Checker](#) service, which can be run over content stored in Amazon to create a listing of checksums calculated for each file. This new service pairs well with the [Bit Integrity Checker](#) service (previously known as the Fixity Service), allowing the heavy processing to be handled in parallel using Hadoop on an Amazon EC2 cluster, and the simpler checks and comparisons to be handled by the DuraCloud instance.
- A new [Duplicate on Demand](#) service, which can be used to copy files from the primary Amazon store into another storage provider. This service pairs well with the [Duplicate on Upload](#) service (previously known as the Replication Service) by performing the large up-front copy using Hadoop on an Amazon EC2 cluster, then allowing Duplicate on Upload to watch for and add new files as they are uploaded.
- Integration of a new storage provider: [Microsoft Windows Azure](#).

For more details about specific changes in release 0.7, see the [JIRA issue tracker](#).

Note that there have been [issues discovered](#) during testing of the Bulk Image Transformer (included in release 0.6 as the Bulk Image Conversion Service). If you choose to run this service, it is recommended that the size of images being used be kept under 100MB. The likelihood of success appears to increase with server size, and number of servers being set to 3 or more is recommended. If you do run this service, please note the data set and configuration and make us aware of the outcome.

Release 0.6

Released: Sept 03, 2010

The primary features of release 0.6 are:

- Addition of a new Fixity Service, which allows for bit integrity checking on content stored within DuraCloud. This service has many options to fit various usage needs. For more information, [see the Fixity Service page](#).
- Addition of a new Bulk Image Conversion Service, which, like the Image Conversion Service, allows for converting images into other formats. This new service, however, makes use of Hadoop in the background to run the conversion using multiple servers, allowing for much higher overall throughput.
- An updated handling of space metadata so that spaces with a large number of content items will not cause slow response times. Now spaces with more than 1000 items will initially show a value of 1000+ as the number of items in the space. DurAdmin, the administrative interface, will then calculate the total number of items on the fly.
- The Sync Tool has a new option (-e) which will cause the tool to exit once it has completed syncing rather than continually monitoring for changes. This makes it easier for administrators to include the Sync Tool in scripts which run daily or weekly to ensure all local content is moved to DuraCloud.
- DurAdmin now provides a way to delete groups of content items and spaces in one step.
- A host of bug fixes and small tweaks

For more details about specific changes in release 0.6, see the [JIRA issue tracker](#).

Release 0.5

Released: July 28, 2010

The primary feature of release 0.5 is the addition of a completely new administrative user interface. This UI, called DurAdmin like its predecessor, provides for easy access to the primary features of DuraCloud.

For more details about specific changes in release 0.5, see the [JIRA issue tracker](#).

Release 0.4

Released 0.4.1: June 30, 2010

- This build release is the first publicly available
- It primarily provides clean-up of projects and tests

Released 0.4: June 21, 2010

The primary features added in release 0.4 of DuraCloud were:

- Media Streaming Service
 - Provides a way to enable streaming for video and audio files as well as providing an example media player.
- Logging moved to SLF4J over Logback
 - Provides greater consistency in log output and greater flexibility in log configuration

For more details about specific changes in release 0.4, see the [JIRA issue tracker](#).

Release 0.3

Released: May 17, 2010

The primary features in the third pilot release of DuraCloud are:

- Security
 - All DuraCloud applications now require authentication prior to performing write activities
 - Read activities on 'closed' spaces also require authentication, but 'open' spaces allow anonymous read access
- Sync Tool
 - Provides a command line utility for keeping DuraCloud content synchronized with the local file system

Other improvements in the 0.3 release:

- Image Conversion Service
 - Adds an option to convert images to the (web standard) sRGB color space
 - Adds the capability to perform multiple conversions at once (providing the compute capacity is available) and provides more frequent activity feedback through the continual writing of the conversion output file
- DuraStore
 - Adds an option for users to provide MD5 checksum when adding content. This disables the in-transfer MD5 computation (providing improved performance) and compares the final MD5 computed by the storage provider with the user provided MD5.

For more details about specific changes in release 0.3, see the [JIRA issue tracker](#).

Release 0.2

Released: Feb 19, 2010

The second pilot release of DuraCloud focused on providing access to services which can be run over content, as well as improvements to the storage foundation provided by the first release.

Services available as of release 0.2:

- J2K service - serves J2K images, provides a J2K image viewer
- Image Conversion service - converts image files from one format to another
- Replication service - replicates content stored in one provider to another upon content upload
- Web Application Utility service - infrastructure service required by J2K service (allows for deployment of web applications)
- ImageMagick service - infrastructure service required by Image Conversion service (provides access to ImageMagick utilities)

Service functions available as of release 0.2:

- Services may be deployed with configuration
- Available and deployed services may be listed
- Deployed service configuration may be viewed and updated
- Deployed service properties may be viewed
- Deployed services may be undeployed and redeployed

New storage functions available as of release 0.2:

- Space content may be listed in chunks with an optional prefix filter
- Space and content metadata may be edited via the UI
- Space and content metadata tags may be added/removed via the UI

For more details about specific changes in release 0.2, see the [JIRA issue tracker](#). Note that while most items included in the release are listed in the tracker, we migrated to using JIRA while working on release 0.2, so issues completed prior to the migration are not included.

Release 0.1

Released: Nov 2, 2009

The first pilot release of DuraCloud laid the foundation for storage across underlying providers.

Through either the web-ui or via direct REST calls

- underlying providers may be listed
- spaces may be created/deleted
- content may be uploaded/downloaded/deleted
- metadata may be viewed
- metadata may be modified
 - *modification is fully supported through the REST API*
 - *modification is partially supported through the web-ui*

DuraCloud Sync Tool

Introduction

The Sync Tool is a utility which was created in order to provide a simple way to move files from a local file system to DuraCloud and subsequently keep the files in DuraCloud synchronized with those on the local system.

Download

[Download the sync tool from the Downloads page.](#)

How the Sync Tool Works

- When you run the Sync Tool for the first time, you must include DuraCloud connection information (host, port, username, password) as well as the space where you would like all of your files stored. You must also provide a list of directories which will be synced to DuraCloud and a directory for the Sync Tool to use for its own work.
- When the Sync Tool starts up, it will look through all of the files in each of the local content directories and add them to its internal queue for processing. Each of those files will then be written to your DuraCloud space. As this initial write is happening a listener is set up to watch for any file changes within each of the content directories. As a change occurs (a file is added, updated, or deleted), that change is added to the queue, and the appropriate action is taken to make the DuraCloud space consistent with the local file (i.e. the file is either written to the space or deleted from the space.)
- You can stop the Sync Tool at any time by typing 'x' or 'exit' on the command line where it is running. It will stop all listeners, complete any file transfers that are in progress, and close down.
- When you restart the Sync Tool, if you point it at the same work directory, it will pick up where it left off. While the Sync Tool is running, it is constantly writing backups of its internal queue, so it first reads the most current backup and begins processing the files there. It then scans the content directories to see if there are any files which have been added or updated since the last backup, and it also pulls a list of files from the DuraCloud space and scans that list to see if any local files have been deleted. Any changes detected are added to the internal queue, and the Sync Tool continues to run as usual.

Operational notes

- Restarting
 - You can perform a restart of the Sync Tool by using the -g command line option to point to the Sync Tool configuration file, which is written into the work directory (named synctool.config)
 - If you would like the Sync Tool to perform a clean start rather than a restart (i.e. you would like it to compare all files in the content directories to DuraCloud) you will need to either point it to a new work directory, or clear out the existing work directory.
 - The Sync Tool will perform a clean start (not a restart) if the list of content directories is not the same as the previous run. This is to ensure that all files in all content directories are processed properly.
- Getting a clean start
 - If you specifically do not want to restart from a previous run, and would like to ensure that the sync tool considers every file in all directories specified, you can use the -l (or --clean-start) command line option to indicate this desire.
 - A clean start will also occur by default whenever the host, destination space, destination store, or the list of content directories changes from one run of the tool to the next.
- Collisions
 - The Sync Tool allows you to sync multiple local directories into a single space within DuraCloud. Because of this, there is the possibility of file naming collisions, where two local files resolve to the same DuraCloud ID. If this happens, one file will be overwritten by the other. There are a few ways to ensure that this does not occur:
 - Ensure that the top level files and directories within the set of content directories do not have overlapping names.
 - Sync only a single directory to a space. You can run multiple copies of the Sync Tool, each over a single local directory, syncing to its own DuraCloud space.
- Work Directory - these files and directories can be found in the work directory (specified using the -w command line parameter)
 - Config Files
 - When the Sync Tool starts up, it writes the list of parameters and values provided by the user

on startup to a file called `synctool.config` in the work directory. This file can be used to restart the Sync Tool, using the `-g` parameter to point to the file's location. You can also restart the Sync Tool by indicating the same set of options as used originally. The `-g` parameter is for convenience only and is not required in any circumstance. Note that this file is overwritten each time the Sync Tool is run with a different set of parameters, so you may choose to copy the file elsewhere (or give it a new name) if you would like to keep a copy of a particular configuration set.

- You may also see a file named `synctool.config.bak` in the work directory which is used to compare against the current config in order to determine if a restart is possible. In order for a restart to occur, the list of content directories (`-c` parameter) must be the same as the previous execution of the tool, and there must be at least one changed list backup (see below.)
- Changed List Directory
 - While the Sync Tool is running it is constantly updating the list of files which have been changed (when starting the first time, this includes all files in the directories that need to be synced). In order to allow the Sync Tool to restart after it has been stopped, this list of files is continually backed up into the `changedList` directory. There is no reason to edit these files, but you may choose to delete the `changedList` directory along with the config files mentioned above to ensure that the Sync Tool does not attempt to perform a restart.
- Logs Directory
 - Information about what the Sync Tool is doing while it is running can be found in the `sync-tool.log` file. It is a good idea to monitor this file for errors and warnings as this information is not printed to the console.
 - The `duracloud.log` file is useful for application debugging when the information in the `sync-tool.log` file is insufficient to understand a problem.

Prerequisites

- You must have Java version 6 or above installed on your local system. If Java is not installed, you will need to [download](#) and install it. To determine if the correct version of Java is installed, open a terminal or command prompt and enter

```
java -version
```

The version displayed should be 1.6.0 or above. If running this command generates an error, Java is likely not installed.

- You must have downloaded the Sync Tool. It is available as a link near the top of this page.

Using the Sync Tool

- To run the Sync Tool, open a terminal or command prompt and navigate to the directory where the Sync Tool is located
- To display the help for the Sync Tool, run

```
java -jar synctool-{version}-driver.jar
```

- When running the Sync Tool for the first time, you will need to use these options:

Short Option	Long Option	Argument Expected	Required	Description	Default Value (if optional)
--------------	-------------	-------------------	----------	-------------	-----------------------------

-h	--host	Yes	Yes	The host address of the DuraCloud DuraStore application	
-r	--port	Yes	No	The port of the DuraCloud DuraStore application	443
-i	--store-id	Yes	No	The Store ID for the DuraCloud storage provider	The primary storage provider is used
-s	--space-id	Yes	Yes	The ID of the DuraCloud space where content will be stored	
-u	--username	Yes	Yes	The username necessary to perform writes to DuraStore	
-p	--password	Yes	Yes	The password necessary to perform writes to DuraStore	
-c	--content-dirs	Yes	Yes	A list of the directory paths to monitor and sync with DuraCloud. If multiple directories are included in this list, they should be separated by a space.	
-w	--work-dir	Yes	Yes	The state of the sync tool is persisted to this directory	

-f	--poll-frequency	Yes	No	The time (in ms) to wait between each poll of the sync-dirs	10000 (10 seconds)
-t	--threads	Yes	No	The number of threads in the pool used to manage file transfers	3
-m	--max-file-size	Yes	No	The maximum size of a stored file in GB (value must be between 1 and 5), larger files will be split into pieces	1
-d	--sync-deletes	No	No	Indicates that deletes performed on files within the content directories should also be performed on those files in DuraCloud; if this option is not included all deletes are ignored	Not set
-x	--exit-on-completion	No	No	Indicates that the sync tool should exit once it has completed a scan of the content directories and synced all files; if this option is included, the sync tool will not continue to monitor the content dirs	Not set

-l	--clean-start	No	No	Indicates that the sync tool should perform a clean start, ensuring that all files in all content directories are checked against DuraCloud, even if those files have not changed locally since the last run of the sync tool	Not set
----	---------------	----	----	---	---------

- When the Sync Tool runs, it creates a backup of your configuration in the work directory that you specify. When running the tool again, you can make use of this file to keep from having to re-enter all of the options specified on the initial run. In this case you need only a single option:

Short Option	Long Option	Argument Expected	Required	Description
-g	--config-file	Yes	Yes	Read configuration from this file (a file containing the most recently used configuration can be found in the work-dir, named synctool.config)

- An example for running the Sync Tool

```
java -jar synctool-{version}-driver.jar -w C:\tools\synctool\backup -c
C:\files\important -f 2000 -h test.duracloud.org -s important-dir-backup -t 5
-u myname -p mypassword
```

Runtime commands

- While the Sync Tool is running, these commands are available. Just type them on the command line where the tool is running. These commands are not available when running in exit-on-completion mode.

Short Command	Long Command	Description
x	exit	Tells the Sync Tool to end its activity and close

c	config	Prints the configuration of the Sync Tool (the same information is printed at startup)
s	status	Prints the current status of the Sync Tool
l <Level>	N/A	Changes the log level to <Level> (may be any of DEBUG, INFO, WARN, ERROR)
h	help	Prints the runtime command help

DuraCloud Java Clients

Download

[Download the Store Client, Service Client, Report Client, Executor Client, and Manifest Client from the Downloads page](#)

Available clients

StoreClient

DuraCloud provides access to files stored in cloud storage systems through an application called DuraStore. DuraStore is installed and running on your DuraCloud instance and can be accessed via a [REST](#) interface. In order to aid Java developers in communicating with DuraStore, a Java client, called StoreClient was written.

ServiceClient

DuraCloud provides management capabilities to deploy, undeploy, configure, and gain status of the services within DuraCloud through an application called DuraService. DuraService is installed and running on your DuraCloud instance and can be accessed via a [REST](#) interface. In order to aid Java developers in communicating with DuraService, a Java client, called ServiceClient was written.

ReportClient

DuraCloud provides reporting capabilities which assist in understanding and managing the content stored in DuraCloud and services run through DuraCloud. These capabilities are included as part of an application called DuraBoss. DuraBoss is installed and running on your DuraCloud instance and the reporting capabilities can be accessed via a [REST](#) interface. In order to aid Java developers in communicating with the reporting features of DuraBoss, a Java client, called ReportClient was written.

ExecutorClient

DuraCloud includes a component called the Executor, which performs actions over the storage and services maintained by DuraStore and DuraService. The Executor allows services to be run in an automated and scheduled fashion. Similar to the reporting functions noted above, the Executor is part of an application called DuraBoss. DuraBoss is installed and running on your DuraCloud instance and the Executor can be accessed via a [REST](#) interface. In order to aid Java developers in communicating with the Executor in DuraBoss, a Java client, called ExecutorClient was written.

ManifestClient

DuraCloud provides the ability to retrieve a manifest file for any space maintained in storage. These manifests list all of the content items and their MD5 checksums, which can be used to compare with your local file system, or with the contents of your DuraCloud space at a later date. This manifest capability is part of the DuraBoss application, which is installed and running on your DuraCloud instance, and can be accessed via a [REST](#) interface. In order to aid Java developers in retrieving manifest documents, a Java client, called ManifestClient was written.

Using the clients

To use any of the above clients, you will need all of the jars included in the libs directory of the download package to be available on your classpath. You will then be able to write code using the provided Javadocs to interact with the client.

Example code

For each of the clients listed above, an example Java class has been provided to assist in set up and testing, as well as a starting point for writing your client code. The example client (found in ExampleClient.java) includes a simple main class which performs a subset calls the client is capable of performing, and printing information to the console. To run an example:

- Extract the client zip file
- Update the HOST, PORT, USERNAME, and PASSWORD constant values in ExampleClient.java as needed to connect to your DuraCloud instance.
- Make sure Ant is available on your path and type "ant" to compile the example.
- Type "ant run" to run the example.

DuraCloud Administration

- [Naming restrictions](#)
- [Access Control Lists \(ACLs\)](#)

This document details some of the considerations of concern to a DuraCloud administrator.

Naming restrictions

1. Space names
 - a. The following restrictions apply to user-defined space names
 - only lowercase letters, numbers, periods, and dashes
 - no adjacent pair of "-" and/or "."
 - no number immediately following the last "."
 - between 3 and 42 characters
 - must start with a letter
 - may not end with a dash
 - b. Note: Users can provide space names through the [REST-API](#) that do not follow these conventions, but the space actually created will have a different name under the covers.
2. Reserved space names
 - a. Due to some specific operations exposed through the durastore [REST-API](#), the following names are unavailable as user-defined space names
 - init
 - stores
 - spaces
 - security
 - task

3. Content object names
 - a. The only restrictions are that a content object name
 - cannot include a question mark '?' character
 - cannot include a reverse solidus (backslash) '\' character
 - is limited to 1024 bytes (byte count is checked after URL and UTF-8 encoding)

Access Control Lists (ACLs)

Prior to DuraCloud v1.3.1, control over access to spaces was limited to setting the space-level permission to either *OPEN* or *CLOSED*. An *OPEN* space permitted anonymous reads of content within that space, whereas a *CLOSED* space disallowed any reads or writes of content within that space by unauthenticated users. As of DuraCloud v1.3.1, the space-level access control mechanisms have been richened.

1. Users and Groups
 - a. Access is granted to *users*, *groups*, or combinations thereof
 - b. *Users* are those with credentials to access an account
 - c. *Groups* are collections of users that are created in the Management Console
2. Rights
 - a. When assigning a space ACL, *users* and/or *groups* are granted one of two rights
 - i. *READ* allows reading any content within that space
 - ii. *WRITE* allows reading, adding, and modifying any content within that space
3. Public (anonymous) Access
 - a. There is a special group named 'public' that can only be granted *READ* access to a space
 - b. If the 'public' group has *READ* access, then unauthenticated (anonymous) reads of content are permitted on that space
4. Use
 - a. REST API can be used to programmatically create, update, and delete space ACLs
 - [Get Space ACLs](#)
 - [Set Space ACLs](#)
 - b. DurAdmin provides authorized users to update space ACLs in the web interface

DuraCloud Services

- [Introduction](#)
- [Duplicate on Change](#)
- [Duplicate on Demand](#)
- [Image Server](#)
- [Media Streamer](#)
- [Bit Integrity Checker](#)
- [Bit Integrity Checker - Tools](#)
- [Bit Integrity Checker - Bulk](#)
- [CloudSync](#)

Introduction

The DuraCloud application provides a set of services which can be deployed and used for a variety of purposes, primarily to process the content which has been loaded into DuraCloud storage. The following list of services describes how each service is expected to be used and the options available for tailoring the service to your needs.

If you start a service, you will receive an email when it completes processing. For each of the services below, if they are run independently, they will not auto-restart if they fail during processing. If you are made aware of a failed job state, you have the option to redeploy the service. Some services are run automatically by the DuraCloud Executor, and as such their state is managed by the Executor. In particular, these are the Media Streamer and the Bit Integrity Tools services.

 Not all services are available in all service plans.

Duplicate on Change

Description:

The Duplicate on Change service provides a way to ensure that the content stored in DuraCloud is synchronized between different storage providers. The Duplicate on Change service duplicates any changes made to spaces, content, or properties for the spaces it is configured to watch. This means that once the Duplicate on Change service is deployed, it notices all content that is added, updated, or deleted for each configured space in the watched DuraCloud provider and performs the same functions on the selected secondary provider. All content that is copied will be placed in an identically named space in the secondary storage location with the same property fields attached. The duplication provided by this service is one-way; only the provider that is selected to be watched is monitored for changes.

Note that this service only performs duplication of content after it has been deployed. Content that exists in the provider prior to this service being deployed will not be duplicated. To duplicate existing content, see Duplication on Demand.

Configuration Options:

1. Store to Watch: The primary storage location which DuraCloud will monitor for changes. When spaces, content, or properties are added, updated, or deleted in this store, the same actions will be taken for the configured spaces in a secondary store.
2. Space to store selection: Each space in the watched provider will be duplicated in the storage provider(s) selected. For each space, 0, 1, or more providers may be selected.

Duplicate on Demand

Description:

The Duplicate On Demand service provides a simple way to duplicate content from one space to another. This service is primarily focused on allowing the duplication of content from the primary storage provider to a secondary provider. To begin, a source space is chosen, along with a store and space to which content will be duplicated. The service then performs a copy of all content and properties in the source space to the duplication space, creating the space if necessary. When the service has completed its work, a results file will be stored in the chosen space and a set of files (primarily logs) created as part of the process will be stored in the work space.

Configuration Options:

1. Source Space: DuraCloud space where source files can be found
2. Copy to this store: DuraCloud store to which content will be copied
3. Copy to this space: DuraCloud space where content will be copied
4. Standard vs. Advanced configuration
 - a. Standard mode automatically sets up the service to be run
 - b. Advanced mode allows the user to configure the number and type of servers that will be used to run the job
 - i. Number of Server Instances: The number of servers to use to perform the duplication task.
 - ii. Type of Server: The type (size) of server used as perform the task. The larger the server, the faster the processing will occur. Larger servers also cost more than smaller servers to run. For more information, see the [Amazon EC2 documentation](#).

Service Output

All outputs of this service are placed in the system space, *x-service-out*.

1. duplicate-on-demand/duplicate-results-<date>.csv
 - Final report indicating which files were duplicated, as well as any failures encountered

Image Server

Description:

The Image Server provides a viewer for image files through use of the [Djatoka image server](#). While this service is geared towards serving JPEG 2000 images, it supports multiple image file types by converting them to JPEG 2000 format on the fly.

Note that the current implementation of this service requires that spaces be set to OPEN in order to use the viewer to view image files.

Configuration Options:

None

Media Streamer

Description:

The Media Streamer provides streaming capabilities for video and audio files. The service takes advantage of Amazon Cloudfront streaming, so files to be streamed must be within spaces on an Amazon provider.

Amazon Cloudfront streaming uses the Flash Media Server to host streaming files over RTMP. **File formats supported include MP3, MP4 and FLV among others.** For a full listing of supported file types see the [Flash Media Server documentation](#).

Configuration Options:

1. Source Media Space: The DuraCloud space(s) where the source video and audio files to be streamed are stored. The Media Streamer service attempts to stream all files in the selected space(s).

Integration Files

The following files are available as a bundle on the [downloads page](#).

They are intended as a starting point for integrating streaming media into your own website.

- player.swf - The flash-based video player [JWPlayer](#)
- playlist.xml - An example playlist which would include a list of items in your Source Media Space
- playlistplayer.html - An HTML file which uses JWPlayer to display the items in the playlist
- singleplayer.html - An HTML file which uses JWPlayer to display a single media file
- stylish.swf - A supplementary flash file used to style the JWPlayer
- swfobject.js - A javascript file ([available from here](#)) used to embed the JWPlayer on a web page
- viewer.js - A javascript file used to simplify the loading of JWPlayer

All of the above files are intended as examples only. Their purpose is give developers a starting point for embedding video streamed by DuraCloud on their own web pages.

i If you add files when the Media Streamer service is already running, they too will automatically be available for streaming.

Bit Integrity Checker

Description:

The Bit Integrity Checker provides the ability to verify that the content held within DuraCloud has maintained its bit integrity. There are two modes of operation.

Modes:

1. Verify integrity of a Space
2. Verify integrity of an item list

When running in the *Verify integrity of a Space* mode, the checker performs the following steps

- collect the content hash values for each item from the underlying storage provider
- stream through each item recalculating their hashes
- compare the two listings

When running in the *Verify integrity of an item list* mode, the checker performs the following steps

- stream through each item in the provided listing, recalculating their hashes
- compare the newly generated listing with the provided listing

Configuration Options:

1. Stores: The underlying storage provider over which the service will run
2. Space containing content items: The DuraCloud space in which the content items to be verified reside
3. Verify integrity of an item list mode
 - a. Input listing name: Name of the content item which contains the listing of items over which to run the service

Service Outputs

All outputs of this service are placed in the system space, *x-service-out*.

1. bitintegrity/fingerprints-gen-<spaceId>-<date>.csv
 - Interim listing generated with hash values from underlying storage provider
2. bitintegrity/fingerprints-<spaceId>-<date>.csv
 - Interim listing with hashes recalculated from content streams
3. bitintegrity/fixity-report-<spaceId>-<date>.csv
 - Final report with status of integrity check

Bit Integrity Checker - Tools

Description:

The Bit Integrity Checker Tools provide additional bit integrity checking utilities which can be used to perform specific integrity checking tasks.

Modes:

1. Generate integrity information for a Space
2. Generate integrity information for an item list
3. Compare two integrity reports

Configuration Options:

1. Mode 1 - *Generate integrity information for a Space*
 - a. Get integrity information from...

- i. The storage provider: Determine the file MD5 by asking the storage provider for its stored MD5 value
 - ii. The files themselves: Determine the file MD5 by retrieving them from the storage provider and computing the MD5
 - b. Stores: The underlying storage provider in which the following space resides
 - i. Space containing content items: The DuraCloud space in which the content items to be considered reside
- 2. Mode 2 - *Generate integrity information for an item list*
 - a. Get integrity information from...
 - i. The storage provider: Determine the file MD5 by asking the storage provider for its stored MD5 value
 - ii. The files themselves: Determine the file MD5 by retrieving them from the storage provider and computing the MD5
 - iii. Input listing name: Name of the content item which contains the listing of items over which to run the service
 - b. Stores: The underlying storage provider in which the following space resides
 - i. Space with input listing: The DuraCloud space in which the input listing file resides
- 3. Mode 3 - *Compare two integrity reports*
 - a. Input listing name: Name of the first content item which contains a listing of items to be compared to the second listing
 - b. Second input listing name: Name of the second content item which contains a listing of items to be compared to the first listing
 - c. Stores: The underlying storage provider in which the following spaces reside
 - i. Space with input listing: The DuraCloud space in which the first input listing file resides
 - ii. Space with second input listing: The DuraCloud space in which the second input listing file resides

Service Outputs

All outputs of this service are placed in the system space, *x-service-out*.

1. bitintegrity/fingerprints-<spaceId>-<date>.csv
 - Listing of hashes when running in *from space* or *from list* modes
2. bitintegrity/fixity-report-<listingId-0>-vs-<listingId-1>-<date>.csv
 - Comparison report of two hash listings

Bit Integrity Checker - Bulk

Description:

The Bulk Bit Integrity Checker provides a simple way to determine checksums (MD5s) for all content items in any particular space by leveraging an Amazon Hadoop cluster. This service is designed for large datasets (+10GB).

Configuration Options:

1. Space to verify: DuraCloud space where source files are stored
2. Service Mode
 - a. Verify integrity of a Space: Retrieves all items in a space, computes the checksum of each, and compares that value with the MD5 value available from the storage provider
 - b. Verify integrity from an item list: Retrieves all items listed in the item list, computes the checksum of each, and compares that value with the MD5 value provided in the item list
 - i. Space with input listing: The DuraCloud space in which the input listing file resides
 - ii. Input listing name: Name of the content item which contains the listing of items over which to run the service
3. Standard vs. Advanced configuration

- a. Standard mode automatically sets up the service to be run
- b. Advanced mode allows the user to configure the number and type of servers that will be used to run the job
 - i. Number of Server Instances: The number of servers to use to perform the duplication task.
 - ii. Type of Server: The type (size) of server used as perform the task. The larger the server, the faster the processing will occur. Larger servers also cost more than smaller servers to run. For more information, see the [Amazon EC2 documentation](#).

Service Outputs

All outputs of this service are placed in the system space, *x-service-out*.

1. bitIntegrity-bulk/bitIntegrity-report-<date>.csv
 - Final report with status of integrity check
2. bitIntegrity-bulk/bitIntegrity-results.csv
 - Interim listing with hashes recalculated from content streams

CloudSync

Description:

The CloudSync service starts and runs the CloudSync application, which provides capabilities to allow the backup and restore of content from a Fedora repository into DuraCloud. For more information about CloudSync, please refer to its the [CloudSync documentation](#).

DuraCloud Features

- **Application Features**
 - [Storage system REST API](#)
 - Space (top level folder) listing
 - Space creation
 - Space deletion
 - Space properties and tagging
 - Space access definition
 - Content (file) listing
 - Content storage
 - Content deletion
 - Content properties and tagging
 - [Service administration REST API](#)
 - Available service listing
 - Deployed service listing
 - Service deployment
 - Service configuration
 - Service undeployment
 - [Reporting system REST API](#)
 - Storage report listing
 - Storage report retrieval
 - Storage report starting and stopping
 - Storage report scheduling
 - Service report listing
 - Service report retrieval
 - [Executor REST API](#)
 - Scheduled execution of services
 - Execution status
 - [Auditor REST API](#)

- Retrieval of audit logs
- [Manifest REST API](#)
 - Retrieval of space-level manifests
- [Security requiring authentication on all DuraCloud applications](#)
- Web user interface (DurAdmin)
 - Access to all storage system capabilities, including space and content creation, updates, and deletes
 - Access to all service administration capabilities, including service deployment, configuration, and undeployment
 - Access to graphical depictions of the information contained in the storage and service reports
 - Bulk deletion of spaces and content items
 - User account administration
- **Core Services**
 - **[Media Streamer:](#)**
 - Provides streaming capabilities for video and audio files.
 - **[Duplicate on Change:](#)**
 - Provides a simple mechanism for synchronizing your content between storage providers. A running Duplicate on Change service will notice when a content item or space is added, updated, or deleted in one store and duplicate those changes in other stores.
 - **[Duplicate on Demand:](#)**
 - Provides a simple way to duplicate content (and its associated properties) from a space in the primary storage provider to a space in another provider.
 - **[Image Server:](#)**
 - Deploys an instance of the Adore Djabatoka web application which provides serving and viewing of JPEG2000 images. Note that in order to view images using the Image Server, the images must be in an OPEN space.
 - **[Bit Integrity Checker:](#)**
 - Provides the ability to verify that the content held within DuraCloud has maintained its bit integrity.
 - **[Bit Integrity Checker - Bulk:](#)**
 - Provides a simple way to determine checksums (MD5s) for all content items in any particular space by leveraging an Amazon Hadoop cluster. This service is designed for large datasets (+10GB).
 - **[Bit Integrity Checker - Tools:](#)**
 - Provides additional bit integrity checking utilities which can be used to perform specific integrity checking tasks.
- **Tools**
 - **[Upload Tool:](#)**
 - Provides a graphical interface for transferring files and folders to DuraCloud.
 - **[Sync Tool:](#)**
 - Provides a command line utility for keeping DuraCloud content synchronized with the local file system.
 - **[Retrieval Tool:](#)**
 - Provides a command line utility for transferring content stored in DuraCloud to the local file system.
 - **[Chunker Tool:](#)**
 - Provides a command line utility for transferring single files to DuraCloud. Files larger than a configurable threshold will be "chunked" (split into multiple files) prior to transfer.
 - **[Stitcher Tool:](#)**
 - Provides a command line utility for retrieving single "chunked" files from DuraCloud. As the file is retrieved, it is "stitched" (combined back into the original file).

DuraCloud Retrieval Tool

Introduction

The Retrieval Tool is a utility which is used to transfer (or "retrieve") digital content from DuraCloud to your local file system.

Download

[Download the retrieval tool from the Downloads page.](#)

How the Retrieval Tool Works

- When the Retrieval Tool starts up, it connects to DuraCloud using the connection parameters you provide and gets a list of content items in the spaces you indicate. It will then proceed to download the files from those spaces, each into a local directory named for the space, which is placed within the content directory.
- For each content item, the Retrieval Tool checks to see if there is already a local file with the same name. If so, the checksums of the two files are compared to determine if the local file is the same as the file in DuraCloud. If they match, nothing is done, and the Retrieval Tool moves on to the next file. If they do not match, the file from DuraCloud is retrieved.
- By default, when a local file exists and differs from the DuraCloud copy, the local file is renamed prior to the DuraCloud file being retrieved. If you would prefer that the local file simply be overwritten, you will need to include the overwrite command-line flag when starting the Retrieval Tool.
- As each content file is downloaded, a checksum comparison is made to ensure that the downloaded file matches the file in DuraCloud. If the checksums do not match, the file is downloaded again. This re-download will occur up to 5 times. If the checksums still do not match after the fifth attempt, a failure is indicated in the output file.
- As each file download completes, a new line is added to the retrieval tool output file in the work directory, indicating whether the download was successful or not. Files which did not change are not included in the output file.
- As the Retrieval Tool runs, it will print its status approximately every 10 minutes to indicate how many files have been checked and downloaded.
- Once all files are retrieved, the Retrieval Tool will print its final status to the command line and exit.
- As files are updated in DuraCloud, you can re-run the Retrieval Tool using the same content directory, and only the files which have been added or updated since the last run of the tool will be downloaded.

Operational notes

- Content Directory - the directory to which files will be downloaded. A new directory within the content directory will be created for each space.
- Work Directory - the work directory contains both logs, which give granular information about the process, and output files. A new output file is created for each run of the Retrieval Tool, and it stores a listing of the files which were downloaded.

Prerequisites

- You must have Java version 6 or above installed on your local system. If Java is not installed, you will need to [download](#) and install it. To determine if the correct version of Java is installed, open a terminal or command prompt and enter

```
java -version
```


The version displayed should be 1.6.0 or above. If running this command generates an error, Java is likely not installed.

- You must have downloaded the Retrieval Tool. It is available as a link near the top of this page.

Using the Retrieval Tool

- To run the Retrieval Tool, open a terminal or command prompt and navigate to the directory where the Retrieval Tool jar file is located
- To display the help for the Retrieval Tool, run

```
java -jar retrievaltool-{version}-driver.jar
```

- When running the Retrieval Tool, you will need to use these options:

Short Option	Long Option	Argument Expected	Required	Description	Default Value (if optional)
-h	--host	Yes	Yes	The host address of the DuraCloud DuraStore application	
-r	--port	Yes	No	The port of the DuraCloud DuraStore application	443
-u	--username	Yes	Yes	The username necessary to perform writes to DuraStore	
-p	--password	Yes	Yes	The password necessary to perform writes to DuraStore	
-i	--store-id	Yes	No	The Store ID for the DuraCloud storage provider	The default store is used

-s	--spaces	Yes	No	The space or spaces from which content will be retrieved. Either this option or -a must be included	
-a	--all-spaces	No	No	Indicates that all spaces should be retrieved; if this option is included the -s option is ignored	Not set
-c	--content-dir	Yes	Yes	Retrieved content is stored in this local directory	
-w	--work-dir	Yes	Yes	Logs and output files will be stored in the work directory	
-o	--overwrite	No	No	Indicates that existing local files which differ from files in DuraCloud under the same path and name should be overwritten rather than copied	Not set
-t	--threads	Yes	No	The number of threads in the pool used to manage file transfers	3

- An example for running the Retrieval Tool

```
java -jar retrievaltool-{version}-driver.jar -c content -h test.duracloud.org
-u myname -p mypassword -w work -s spacel spacel -o
```

DuraCloud Stitcher Tool

Introduction

The Stitcher Tool is a utility which provides a simple way to retrieve "chunked" files from DuraCloud. When files are moved to DuraCloud using either the [Chunker Tool](#) or the [Sync Tool](#) and they exceed a defined size limit, they are split (chunked) into multiple files for transfer. The Stitcher tool provides the means by which those files can be retrieved and combined to result in the original file. It should be noted that the Stitcher Tool is also embedded into the [Retrieval Tool](#). If you are using [Retrieval Tool](#), stitching will be automatically performed.

Download

[Download the Stitcher Tool from the Downloads page.](#)

Operational notes

- If you want to jump directly into using the tool, download it from the link above and run the following command

```
java -jar stitch-{version}-driver.jar
```

The resulting usage statement (detailed below) should be enough to help you get started.

- When using the Stitcher Tool, you need to know the ID of the manifest which was generated to list all of the chunks of the original file. If the chunking was done by either the Chunker or Sync tool, then the name of the manifest is the name of the original file (prefixed with any enclosing directory names) followed by ".dura-manifest".

Using the Stitcher Tool

- To run the Stitcher Tool, open a terminal or command prompt and navigate to the directory where the Stitcher Tool is located and run the above command.
- The following options are available when running the Stitcher Tool

Short Option	Long Option	Argument Expected	Required	Description	Default Value (if optional)
-m	--manifest-id	Yes	Yes	The ID of the manifest file used to contain the listing of content chunks	

-d	--to-dir	Yes	Yes	Retrieved and stitched content is stored in this local directory	
-s	--space-id	Yes	Yes	The space ID in which content and manifest files reside	
-i	--store-id	Yes	No	The store ID for the DuraCloud storage provider	The default store is used
-h	--host	Yes	Yes	The host address of the DuraCloud instance	
-r	--port	Yes	No	The port of the DuraCloud instance	443
-u	--username	Yes	Yes	The username necessary to perform writes to DuraCloud	
-p	--password	Yes	Yes	The password necessary to perform writes to DuraCloud	

DuraCloud Upload Tool

Introduction

The Upload Tool is a graphical utility which is used to transfer files from your local file system to DuraCloud. The Upload Tool can be run two ways:

1. By selecting the option to upload items into a space on the DuraCloud UI. This will open a new window and display the Upload Tool.
2. By downloading the Upload Tool (see below) and running it on your own machine (or any other machine you copy it to.)

Regardless of which method you choose for running the Upload Tool, the functions will be the same. You will be able to select files and folders to be uploaded, then once the upload has started, you will see the progress of the upload.

Download

[Download the upload tool from the Downloads page.](#)

Operational notes

- When starting the downloaded version of the Upload Tool, you will first be asked for some information necessary to connect to DuraCloud. The username and password required are those you would use to log into DuraCloud. The Space ID value is the space into which you would like the uploaded files to go. There is no need to enter any of this information when running the Upload Tool from the DuraCloud UI.
- When the Upload Tool starts running, it checks to see if any of the files you have selected already exist in the space. It does this by checking for files with the same name. Any name matches trigger a checksum verification to determine if the files differ. If the files are different, the file being uploaded will overwrite the file in the space. If the files are found to be the same, no transfer occurs.

Prerequisites

- You must have Java version 6 or above installed on your local system. If Java is not installed, you will need to [download](#) and install it. To determine if the correct version of Java is installed, open a terminal or command prompt and enter

```
java -version
```

The version displayed should be 1.6.0 or above. If running this command generates an error, Java is likely not installed.

Using the Sync Tool

- To start the Upload Tool from the DuraCloud UI, simply navigate to the space you would like to upload files to, and select the "Add Many Items" button.
- To start the downloaded Upload Tool, you can either simply double-click the downloaded file or execute the file from the command line.
 - To execute the Upload Tool via the command line, enter:

```
java -jar uploadtool-{version}-driver.jar
```

- If you choose to execute the Upload Tool via the command line, you have the option of providing connection parameters on the command line as well. If you enter these values on the command line, the first screen of the Upload Tool will be skipped. The connection parameters must be entered on the command line in this order (all values are required): host port username password spaceID. An example of running the upload tool using these parameters:

```
java -jar uploadtool-{version}-driver.jar institution.duracloud.org 443  
jsmith secretPassword myContent
```

Upload Tool Help

Upload Tool Help

- [Upload Tool Help](#)
 - [What is the Upload Tool?](#)
 - [The Upload Tool isn't working!](#)
 - [What should I see?](#)
 - [What should I do?](#)
 - [No Pop-Up](#)
 - [Pop-Up but not loading](#)
 - [It's still not working!](#)
 - [I need more help!](#)

What is the Upload Tool?

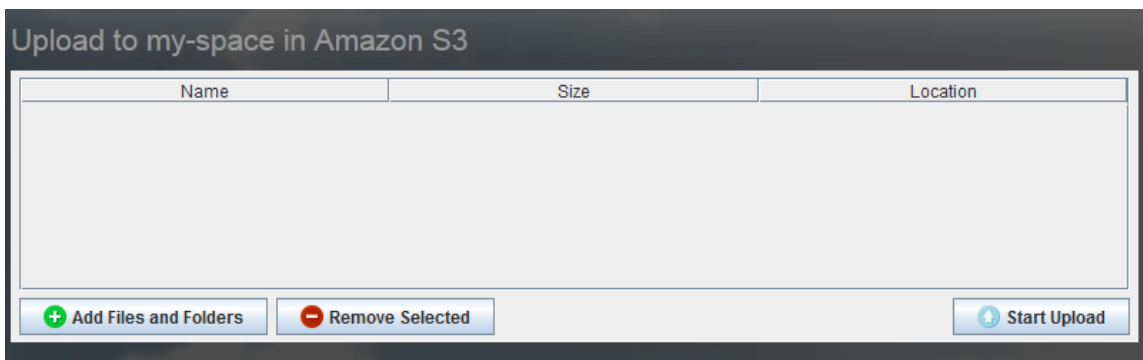
The Upload Tool is a small application which can be used to visually select files and folders, and monitor the transfer of files to DuraCloud. This tool is integrated into the DuraCloud software, and shows up when you select the "Add Items" button on the DuraCloud user interface.

The Upload Tool isn't working!

We are sorry that you're not able to use the Upload Tool at the moment. There are a variety of reasons why the Upload Tool may not be functioning properly. In the sections below, we'll offer some suggestions which may help to resolve the problem.

What should I see?

When the Upload Tool is properly loaded and ready for use, you should see a window which looks similar to the following:



What should I do?

No Pop-Up

If a window did not pop up when you selected the "Add Items" button, there is a good chance that your browser did not allow the pop-up which includes the Upload Tool to load. Depending on your browser, you may need to click a button at the top of the page to allow the pop-up to load, or add an exception to the pop-up blocker for the DuraCloud website.

Pop-Up but not loading

If the window popped up when you clicked the "Add Items" button, but the Upload Tool did not load into it, try the following actions. Note that once the Upload Tool loads properly, there is no reason to continue with further items in this list:

1. If there are any dialogs or selectors which ask for permission to load a Java applet, ensure that you have

- selected the option to allow the applet to load
2. Refresh the window by selecting the window, right clicking, and selecting "Reload".
 3. Close the pop-up window, and select "Add Items" again.
 4. Close your browser, open it again, log back into DuraCloud, then select "Add Items" again.
 5. Try a different browser. The recommended options are Mozilla Firefox and Google Chrome.
 6. Verify that the browser you are using have Java 7 available
 - a. **The Upload Tool requires Java 7 to function properly.** You can [test the Java version used by your browser here](#). If this page indicates that Java is not working properly, or that the version is less than 7, you will need to install the latest version of Java.

If it's still not working, consider trying these more advanced techniques:

1. Clear your browser's cache, restart your browser, and try the Upload Tool again.
2. Ensure that the Java plug-in within your browser is up-to-date and enabled. If an update was required, restart the browser and try the Upload Tool again.
3. Clear the applet cache
 - a. Open a terminal window. On Windows, this is called Command Prompt.
 - b. Type "javaws -uninstall" and hit Enter
 - c. Restart your browser, and try the Upload Tool again

It's still not working!

If none of the above actions worked, you may want to try running the Upload Tool directly from your machine. Directions for downloading and running the Upload Tool locally can be found on the [DuraCloud Upload Tool](#) page.

I need more help!

Finally, if you need further help either getting the Upload Tool to work for you or just determining the best strategy for transferring content into DuraCloud, you can submit a support request ticket through the [DuraCloud support system](#).

Known Issues

The following issues are known to exist in release 2.1.0 of DuraCloud:

Title	Issue	Tracker Item	Work Around
Incorrect redirects from HTTP to HTTPS	When performing PUT/POST/DELETE requests via the REST API, if the URL uses http:// rather than https:// the response is a 404.	DURACLOUD-255	Use https:// when performing requests via the REST API of hosted DuraCloud instances
Incorrect MIME types set on files	When uploading files via the sync tool which do not have an extension, the MIME type is always set to application/octet-stream	DURACLOUD-227	Set the MIME type of these files either through DurAdmin or via the REST API

Administrative files in storage report graphs	The storage report information displays all files in DuraCloud, including all of the administrative files	DURACLOUD-514	View reports for content spaces directly
Sync Tool and Java 7	When running the Sync Tool with Java 7, errors indicating "peer not authenticated" or "unrecognized_name" may occur. This is due to a changed requirement in Java 7.	DURACLOUD-717	Use the flag: "-Djsse.enableSNIExtension=false" when running the Sync Tool

More issues and planned improvements can be found on the [DuraCloud issue tracker](#).

Building DuraCloud Software from Source

- [Introduction](#)
- [Prerequisites](#)
- [Setting up DuraCloud](#)
 - [Build and deploy the DuraCloud web applications](#)
 - [Set up the OSGi services container](#)
 - [Initialize the DuraCloud applications](#)
 - [Test your installation](#)
- [Optional items](#)
 - [Code coverage](#)
 - [Service tests within OSGi \(Linux only\)](#)
 - [Logging](#)
- [DuraCloud internal tools](#)
 - [ServicesAdmin CLI](#)
 - [Building Java client packages](#)
- [Misc configuration/discussion](#)
 - [Services on Windows](#)

Introduction

i If you would prefer to install DuraCloud from a binary distribution, you can [find instructions for that process here](#).

DuraCloud application software is composed of many parts. A breakdown of the primary pieces is as follows:

- DuraStore - this web application provides the access to and management of storage resources, which includes handling the storage portion of the DuraCloud REST API
- StorageProviders - this set is made up of the StorageProvider interfaces and the implementations which connect to distinct cloud stores (currently Amazon S3, Rackspace CloudFiles, and Windows Azure)
- DuraService - this web application handles the deployment and management of services within DuraCloud, which includes handling the services portion of the DuraCloud REST API
- DuraBoss - this web application includes the reporter, executor, auditor, and manifest projects, each of which

expose their own distinct set of calls via the DuraCloud REST API

- DurAdmin - this web application is the UI front end of DuraCloud, it provides users with a view into the information available from the other applications. DurAdmin uses the REST APIs of the other applications to communicate with them.
- Services - the set of all deployable services, as well as the support projects that allow the DuraCloud services infrastructure to function
- Security - handles security for the DuraCloud applications
- Common - a set of projects which provide utilities for other portions of the codebase to reuse

The DuraCloud software, by its very nature, is designed to be integrated with underlying cloud storage providers. As may be expected, these integrations are necessary for the system to be properly exercised. In order for DuraCloud to connect to these underlying providers, appropriate credentials must be provided as part of the application initialization step. It is recommended that you acquire the necessary storage provider credentials prior to attempting to set up DuraCloud. Only one storage provider is required to run DuraCloud. The storage providers which are currently supported are:

- [Amazon S3](#)
- [Rackspace CloudFiles](#)
- [SDSC Cloud](#)
- [Windows Azure](#) (beta)

This guide lays out the steps necessary to begin using DuraCloud:

1. Build and deploy the DuraCloud web applications
2. Set up the OSGi services container
3. Initialize the DuraCloud applications
4. Test your installation


Although this document is written from a Linux environment perspective, analogous builds/installations have been tested in Windows (but may have limitations, as noted below). Any comments or feedback are welcomed.

Prerequisites

Software that must be installed on your system prior to building/using DuraCloud

1. Maven 2.2.1 or above
2. Tomcat 6.x or above
3. Java 6 (note: the djatoka service has compatibility issues with open-jdk)
4. Subversion

Setting up DuraCloud

 Any portions of the configuration below for which you need to include a replacement value will be written in all capital letters and included in brackets: [LIKE-THIS]

Build and deploy the DuraCloud web applications

1. Check out latest stable release from Subversion repository

```
svn co https://svn.duraspace.org/duracloud/tags/duracloud-2.1.0
```

2. Set environment variables

```
export JAVA_OPTS="-XX:MaxPermSize=256m"
export MAVEN_OPTS="-Xmx1024m -XX:MaxPermSize=1024m"
```

3. Configure Tomcat

a. Add to \$CATALINA_HOME/conf/tomcat-users.xml

```
<tomcat-users>
  <role rolename="manager" />
  <role rolename="admin" />
  <user username="[ANY-USERNAME]" password="[ANY-PASSWORD]"
roles="admin,manager" />
</tomcat-users>
```

4. Start tomcat

```
$CATALINA_HOME/bin/startup.sh
```

5. Configure Maven2

a. Add tomcat user to \$M2_HOME/conf/settings.xml

```
<servers>
  <server>
    <id>tomcat-server</id>
    <username>[ANY-USERNAME]</username>
    <password>[ANY-PASSWORD]</password>
  </server>
</servers>
```

6. Build

a. From top of source tree

```
mvn clean install
```

Set up the OSGi services container

This step assumes the successful completion of the previous build instructions

Linux/Mac

1. Start OSGi service container

```
cd //services/servicesadmin
mvn clean -f pom-run.xml pax:provision
cd runner
chmod +x run.sh
export BUNDLE_HOME=[DURACLOUD_HOME]/osgi-container
./run.sh
```

- Where [DURACLOUD_HOME] is a directory where the application has write access (can be same as <duracloud.home> set in Maven settings.xml above)
- The run.sh script will start an OSGi container and commandline interface to it
- The container starts with required bundles including the 'services-admin' installed

Windows

1. Set up OSGi service container

```
cd services/servicesadmin
mvn clean -f pom-run.xml pax:provision
cd runner
```

2. (Optional) Set the OSGi bundle storage location

```
set BUNDLE_HOME=[BUNDLE_HOME]
```

- Where [BUNDLE_HOME] is the full path to an empty directory where the osgi container content will be stored
- Open the run.bat file in the runner directory in a text editor and replace all instances of "\$BUNDLE_HOME" with "%BUNDLE_HOME%"
- Note: A directory called "\$BUNDLE_HOME" under the runner directory will be used as the default bundle home if one is not specified.

3. (Optional) Set up logging

- Download the logback.xml file [found here](#) into your bundle home directory.
- Open the logback.xml file in a text editor and edit the LOG_FILENAME property to point to a full file path (including file name) for a log file.
- Note: One benefit to performing this step will be faster start time for your OSGi container.

4. Start OSGi service container

```
run.bat
```

- The run.bat script will start an OSGi container and commandline interface to it
- The container starts with required bundles including the 'services-admin' installed

Once the OSGi services container is running, check to ensure that it was created properly

1. In the console where the "run" script was executed, an "osgi" prompt should be available. If it is not available, hitting enter should bring it up.
2. Type "ss" and hit enter. This should list all of the available bundles. This list should include 50 items, all of which are either in the ACTIVE or RESOLVED state.

Initialize the DuraCloud applications

1. Use the application initialization (app-config) utility to configure the deployed DuraCloud applications
 - a. Build app-config utility, from within the //app-config module

```
mvn assembly:assembly
```

- b. Run the app-config utility

```
java -jar target/app-config-[VERSION]-driver.jar <init.props>
```

- i. The init.props file is a configuration file which specifies all of the information necessary for the DuraCloud applications to run. An example of this file can be found at //app-config/src/main/resources/init.props. This file will need to be updated to match your environment.
- c. When the app-config utility completes successfully, the last line of output printed to the console will be the word "success". If this is not the case, check that your configuration file includes the correct information.

Test your installation

1. Once all of the above steps have been completed, your DuraCloud should be ready to test.
 - a. Go to <http://localhost:8080/duradmin> (change host or port as necessary).
 - b. Log in using the credentials provided in your configuration file
 - c. You should be able to view, add, update, and delete spaces and content in Spaces tab
 - d. You should be able to deploy services in the Services tab
2. Congratulations! You now have a functional DuraCloud installation.

Optional items

Code coverage

1. If you plan on using Clover, the following element needs to be added to your maven 'settings.xml'

```
<profiles>
  <profile>
    <id>profile-clover</id>
    <activation>
      <property>
        <name>profile</name>
        <value>clover</value>
      </property>
    </activation>
    <properties>
      <cloverLicense>[LOCATION-OF-clover.license-FILE]</cloverLicense>
    </properties>
  </profile>
</profiles>
```

2. To run clover

```
mvn clover2:instrument clover2:aggregate clover2:clover -Pprofile-clover
```

3. A report will be generated in the following directory:

//target/site/clover/

Service tests within OSGi (Linux only)

1. Assuming that the OSGi services container is set up and running (as described above), tests that deploy services into the OSGi environment may be run
 - a. From inside the //integration-test module

```
mvn install -PrunServicesAdminTests
```

Logging

1. DuraCloud uses the SLF4j logging framework backed by the LogBack implementation
2. By adding either a logback.xml or logback-test.xml file on the classpath, logging configuration can be customized

DuraCloud internal tools

ServicesAdmin CLI

1. This tool provides a commandline interface for interacting with the 'services-admin' installed in a running OSGi container (see notes above for starting the container)
2. To build and run the CLI, from within the //servicesadminclient module

```
mvn assembly:assembly
java -cp target/servicesadminclient-[VERSION]-cli.jar
```

Building Java client packages

1. To create a distributable zip of the storeclient, serviceclient, or reportclient which includes their dependencies, from within the project directory (//storeclient, //serviceclient, //reportclient) run

```
mvn install -Ppackage-client
```

2. The packaged zip will be found under the project's target directory

Misc configuration/discussion

Services on Windows

The following services do not function in a Windows deployment environment

- WebAppUtilService
- HelloWebappWrapper
- J2KService
- ImageMagickService

If you would like to run the ImageConversionService, you must install ImageMagick and have its /bin directory in your PATH, which is essentially what the ImageMagickService does in a linux environment.

Deploying DuraCloud from Binaries

i If you run into problems trying to install and run DuraCloud, the best place to ask questions is on the [DuraCloud Dev mailing list](#)

The steps below outline how to start up a DuraCloud instance from the binary distribution.

1. Download the binary distribution from the [downloads page](#)
2. Install Tomcat application server (versions 5.5.x and 6.0.x are known to work)
3. Deploy the web applications
 - a. Take the 4 .war files included in the binary distribution and copy them into the "webapps" directory under your tomcat installation.
 - b. Start tomcat, as part of the startup process, tomcat will unpack the wars and deploy them
4. Start up the OSGi container
 - a. Unpack the osgi-container zip file from the distribution package
 - b. Create an empty directory which will serve as the execution directory for the services
 - c. Create a BUNDLE_HOME environment variable pointing at the full path to the new execution directory
 - d. From a command prompt, step down into the runner directory under the unpacked osgi-container directory
 - e. Make the run.sh script executable

```
chmod +x run.sh
```

- f. execute run.sh
 - Note: Errors of the type "ClassNotFoundException: org.apache.jasper.servlet.JspServlet" can be safely ignored
 - g. Wait a few moments for startup to complete. Then hit enter to get a prompt and type "ss". You should see a list of 50 bundles which are in either the ACTIVE or RESOLVED states.
5. Initialize the DuraCloud applications
 - a. Open the init.properties file found in the distribution package

- b. Edit the values that are contained in brackets "[...]" to be appropriate for your environment. This is the step where most problems tend to show up, so feel free to ask questions regarding what values need to be included in this file.
- c. A few notes on editing this file:
 - The value of [host] will almost certainly be: localhost
 - You will need to have your own Amazon S3 account in order to connect. The "username" and "password" in the init.properties file, in the case of Amazon, refer to the Access Key ID and the Private Access Key that are used to make API connections to Amazon Web Services.
 - You probably want to comment out (using '#') the section starting with 'durastore.storage-acct.1', unless you also want to create a Rackspace CloudFiles account.
 - The version number below is currently, 2.1.0

```
duraservice.primary-instance.services-admin-context=org.duracloud.services.admin_[VERSION]
```

- Even though they are not in brackets, you will probably want to update the values of duraservice.service-storage.space-id and duraservice.service-storage.service-xml-id. I'd recommend these values:

```
duraservice.service-storage.space-id=duracloud-2-1-0-service-repo
duraservice.service-storage.service-xml-id=duracloud-2-1-0-service-repo.xml
```

- The security user settings at the bottom allow you to indicate the username and password users accounts that can access your local DuraCloud, you'll use these credentials to log in to DuraCloud after initialization

- d. Execute the app-config jar, passing in the init.properties file as a parameter

```
java -jar app-config.jar init.properties
```

- e. The very last line of output from the execution of the app-config process should be "success". If that's not the case, look at the output more closely to determine what error may have occurred.

6. Add services

- a. At this point, the applications are deployed and initialized. Point your browser to <http://localhost:8080/duradmin> and log in to DurAdmin.
- b. On the spaces tab, create a space called: duracloud-2-1-0-service-repo (assuming you took the naming suggestions above)
- c. Use the Add Items button to add all of the files included under the services directory in the distribution package to this new space.

This should be it. You should now be able to perform all space functions and deploy services

DuraCloud Security

Overview

The security approach is divided into two distinct spheres of responsibility

1. Channel security (encryption)

2. Application security (AuthN / AuthZ)

The configuration of any given user compute instance will consist of an Apache HttpServer layered on top of Tomcat.

1. Apache HttpServer

- All requests will come through Apache on port 443 (https) of the instance
- The requests will internally be unencrypted, where encryption exists, and redirected to tomcat as open text

2. Tomcat

- A defined set of resource endpoints will require AuthN and AuthZ
- Spring-security is being leveraged to wire AuthN and AuthZ across relevant resources

Channel Security Implementation

1. Apache HttpServer is configured to require all requests to the four DuraCloud web applications (/duradmin, /durastore, /duraservice, and /duraboss) go over https.
2. Below are the https enforcement rules configured in Apache

```
###
# ensure 'duradmin' uses https
###

RewriteCond %{REQUEST_URI} /duradmin
RewriteCond %{SERVER_PORT} !^443$
RewriteRule ^(.*)$ https://%{SERVER_NAME}$1 [R=301,L]

###
# try to require https for 'durastore', 'duraservice', & 'duraboss' for
# external requests
###

RewriteCond %{REQUEST_URI} ^(/durastore|/duraservice|/duraboss)
RewriteCond %{SERVER_PORT} !^443$
RewriteCond %{SERVER_NAME} !^localhost$
RewriteCond %{SERVER_NAME} !^127.0.0.1$
RewriteCond %{REMOTE_HOST} !^127.0.0.1$
RewriteCond ${local-ip-map:%{REMOTE_HOST}} !^localhost$
RewriteRule ^(.*)$ https://%{SERVER_NAME}$1 [R=301,L]
```

Application Security Implementation

The basic AuthN flow is as follows

1. User requests secured resource
2. If credentials not in request
 - response 401
3. Spring AuthenticationProvider performs AuthN
 - a. AuthProvider asks UserDetailsService for GrantedAuthorities for given Principal
 - b. notes
 - i. DuraCloud provides custom UserDetailsService implementation to return UserDetails of requesting Principal
 - ii. AbstractSecurityInterceptor permanently caches user AuthN decisions by default
4. Authentication object and "configuration attributes" are passed to AccessDecisionManager for AuthZ

Security Servlet Filters

DuraCloud leverages Spring's mechanism for wiring AuthN/Z into an application across servlet url patterns. The following access rules are placed across the durastore and duraservice REST-APIs:

Initialization REST Methods - Common across all applications	
Action	Role
Is Initialized	ROLE_ANONYMOUS
Initialize	ROLE_ROOT
Initialize Security Users	ROLE_ROOT

DuraStore REST Methods

Action	Role
Get Stores	ROLE_USER
Get Spaces	ROLE_ANONYMOUS if space ACL allows public read, else ROLE_USER
Get Space	ROLE_ANONYMOUS if space ACL allows public read, else ROLE_USER
Get Space Properties	ROLE_ANONYMOUS if space ACL allows public read, else ROLE_USER
Get Space ACLs	ROLE_ANONYMOUS if space ACL allows public read, else ROLE_USER
Create Space	ROLE_ADMIN
Set Space Properties	ROLE_USER
Set Space ACLs	ROLE_ADMIN
Delete Space	ROLE_ADMIN
Get Content	ROLE_ANONYMOUS if space ACL allows public read, else ROLE_USER
Get Content Properties	ROLE_ANONYMOUS if space ACL allows public read, else ROLE_USER
Store Content	ROLE_USER
Copy Content	ROLE_USER
Set Content Properties	ROLE_USER
Delete Content	ROLE_USER
Get Tasks	ROLE_ADMIN
Perform Task	ROLE_ADMIN

DuraService REST Methods

Action	Role
Get Services	ROLE_USER
Get Service	ROLE_USER
Get Deployed Service	ROLE_USER
Get Deployed Service Properties	ROLE_USER
Deploy Service	ROLE_ADMIN
Update Service Configuration	ROLE_ADMIN
UnDeploy Service	ROLE_ADMIN

DuraBoss REST Methods

Action	Role
Get Latest Storage Report	ROLE_ADMIN
Get Storage Report List	ROLE_ADMIN
Get Storage Report	ROLE_ADMIN
Get Storage Report Info	ROLE_ADMIN
Start Storage Report	ROLE_ROOT
Cancel Storage Report	ROLE_ROOT
Schedule Storage Report	ROLE_ROOT
Cancel Storage Report Schedule	ROLE_ROOT
Get Deployed Services Report	ROLE_ADMIN
Get Completed Services Report	ROLE_ADMIN
Get Completed Services Report List	ROLE_ADMIN
Get Services Report	ROLE_ADMIN
Get Executor Status	ROLE_ADMIN
Get Supported Executor Actions	ROLE_ADMIN
Perform an Executor Action	ROLE_ROOT
Shutdown Executor	ROLE_ROOT
Create Initial Audit Log	ROLE_ROOT
Get Audit Logs	ROLE_ADMIN
Shutdown Auditor	ROLE_ROOT
Get Content Manifest	ROLE_ADMIN

Roles

The fixed set of users/roles listed below are provided in DuraCloud. Each role in the list below represents a super set of the privileges of those above it.

1. ROLE_ANONYMOUS
 - no username/password
2. ROLE_USER
 - user created by DuraCloud-account admin
3. ROLE_ADMIN
 - administrator of DuraCloud-account
4. ROLE_ROOT
 - DuraSpace personnel

User Management and Access Control

- Users are managed via the [DuraCloud Management Console](#). In the Management Console, an account administrator has the ability to:
 1. Add and remove users to the DuraCloud account
 2. Create Groups and add users to groups in order to simplify access control
- Access Control is managed at the space level
 - Within DuraCloud (via the UI or the REST API), an account administrator has the ability to define which users and groups have access to a space, as well as the type of access (read or write) that is available.

Service Development Guide

Introduction

This document is intended to provide guidelines for developing your own custom services that can be deployed into the DuraCloud services framework.

The framework has been designed to allow any Java code that can be deployed as an OSGi bundle and implements the [ComputeService](#) interface to be dynamically installable into a running DuraCloud application instance.

Overview

In order to understand the flow of services within the services framework, there are several entities with which to become familiar:

1. **DuraService**
 - web application that exposes a REST-API for end-user management of services
 - a. deploy service
 - b. undeploy service
 - c. configure service
 - d. get service listing
 - e. get service properties
 - responsible for retrieving a service to be deployed from the **Service Registry** and streaming it to the **ServicesAdmin**
 - responsible for mediating subsequent end-user management of deployed services to **ServicesAdmin**
 - responsible for passing service configuration to the OSGi container when a service is being deployed.
 - service configuration is defined in the service-repo.xml which is built based on the [ServiceInfo](#) class for each service
2. **Service Registry**
 - persistent holding area of deployable services
 - contains **Service Packages** and a single services-configuration descriptor XML object

- implemented as one or more DuraCloud "space(s)"
 - the properties file to initialize DuraCloud contains duraservice.service-storage properties which allow configuration of the host, port, context and space

3. **Service Package**

- for simple services, this can be a single OSGi bundle, jar file
- for more involved services, this would be a zip file containing
 - a. OSGi bundle jar that implements the [ComputeService](#) interface
 - b. OSGi bundle jars that are both dependencies of the ComputeService bundle (above) and not among the base bundles installed in the **OSGi Container** by default
 - c. any support files (see 'Example Services' below)

4. **ServicesAdmin**

- management component resident in the OSGi container
- exposes an internal REST-API through which communication from **DuraService** takes place
- responsible for actually communicating with services via their common [ComputeService](#) interfaces
- responsible for keeping track of currently deployed services and dependency-link counters

5. **OSGi Container**

- production implementation currently in use is Equinox
- bundles installed by default are listed as dependencies in [the pom-run.xml file in the ServicesAdmin project](#)

Example Services

All of the below services are currently functional and available in the DuraCloud source baseline. They can all be found within the [services](#) directory

Pure Java Services

1. Hello Service
2. Replication Service
3. WebAppUtil Service
4. ImageConversion Service
5. MediaStreaming Service

Web Application Services

1. HelloWebAppWrapper Service
2. J2K Service

System Services

1. Script Service
2. ImageMagick Service

Implementation Guidelines

 This portion is still under construction.

OSGi Compatibility

- Although there are several available OSGi containers, and the DuraCloud test harnesses employ Equinox, Felix, and Knoplerfish, the production container into which services need to be able to deploy is **Equinox**.

ComputeService Interface

Dependency Packaging

Service Descriptor