

Digital Preservation in Production

...

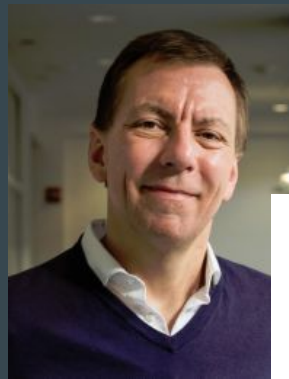
CNI Spring 2017

Who are we?



Bill Branan
Services Technical Director
DuraSpace

bbranan@duraspace.org



Dave Pcolar
Chief Technology Officer
DPN

dave@dpn.org

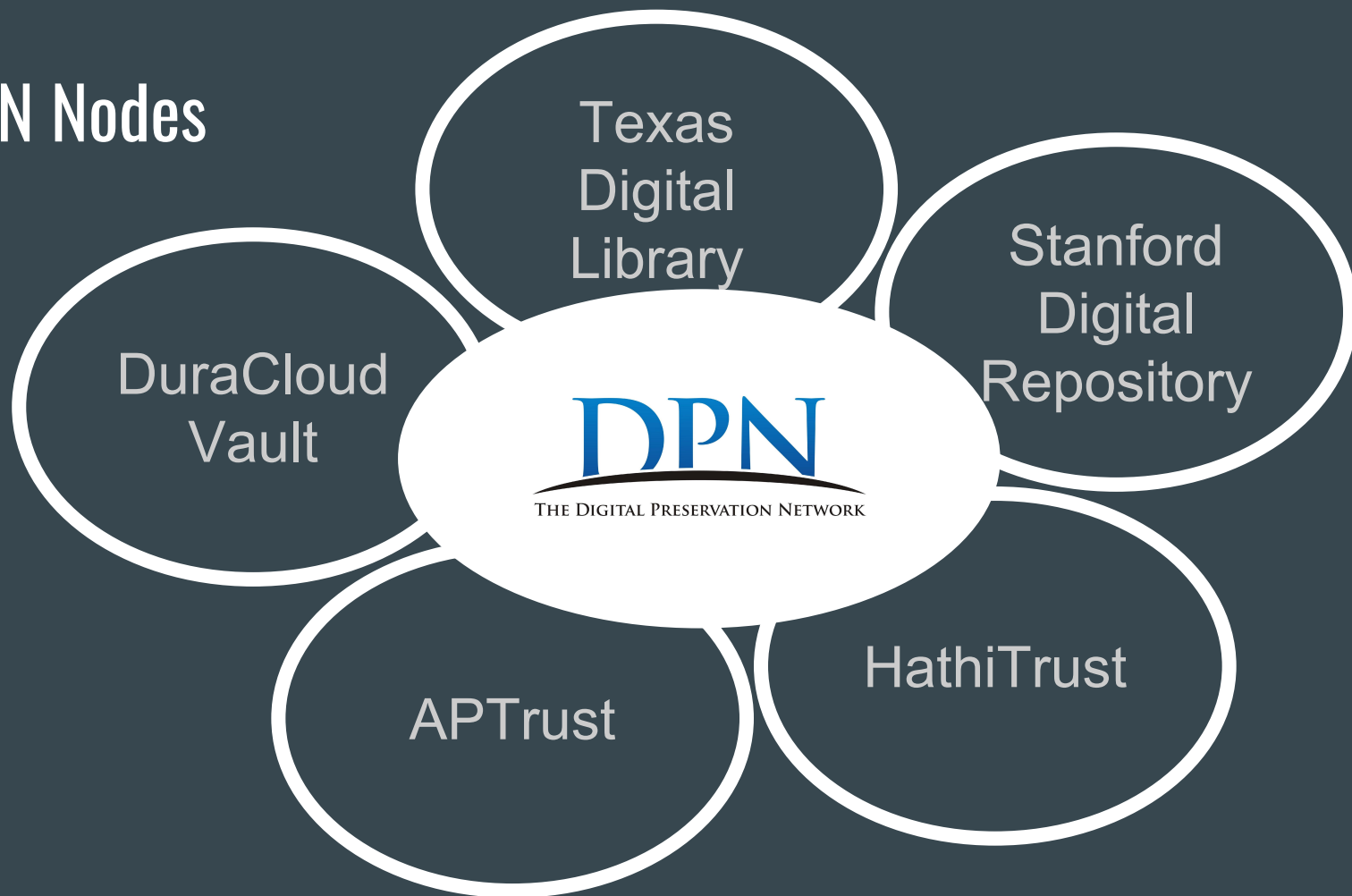
What is DPN?

- Network of distributed preservation systems (nodes)
- Built for truly long-term preservation
- Both technical and legal



Mission: Create a system that ensures the survival of the digital scholarly record for the next generation

DPN Nodes

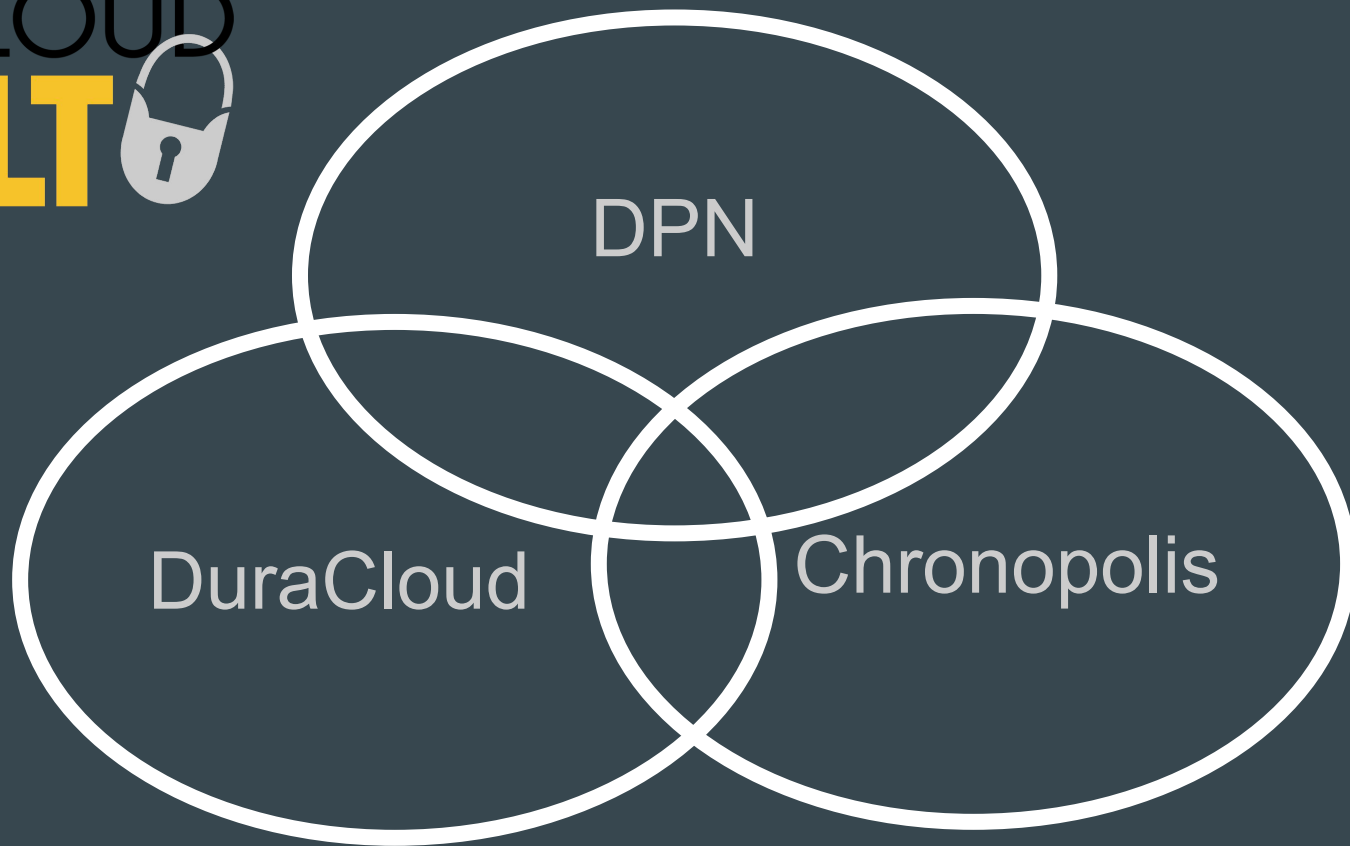


Moving into Production

- Legal
 - Member Deposit Agreement
 - Node Service Agreement
- Technical
 - Node software - version 1 and 2
 - DuraCloud Vault deposit workflow
- Operational
 - DPN-Node demarcation and ingest

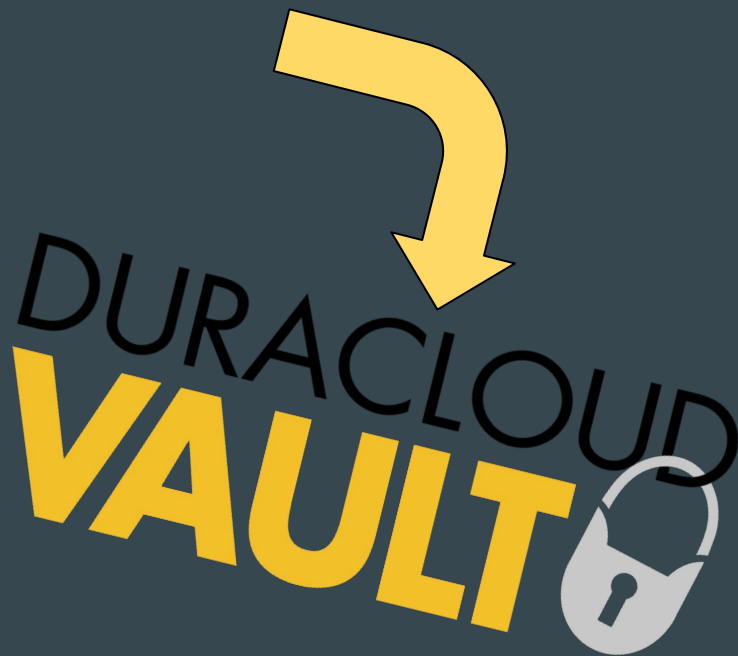


DURACLOUD
VAULT



Deposit in 7 easy steps

1. Become a DPN member
2. Sign Deposit Agreement
3. Create DuraCloud Vault account
 - a. No additional fees
 - b. No additional agreements
4. Attend DuraCloud Vault orientation
5. Select content to deposit
 - a. 5 TB per year is included in membership
6. Transfer content into DuraCloud Vault space
7. Submit Snapshot



DuraCloud :: Spaces

←

→

↺

🏠

Secure

https://vault.duracloud.org/duradmin/spaces/sm/1063/sample-collection-1

☆

⋮

DURACLOUD™

Spaces

DPN

DATA PRESERVATION NETWORK

Spaces

+ Add Space

filter

Spaces

☒ sample-collection-1

☐ testspace

☐ x-duracloud-admin

☐ x-snapshot-metadata

Content Items

Refresh

Upload

type prefix

Space Detail

sample-collection-1

Delete

Audit Log

Manifest

Make publicly readable

Snapshot

Create Snapshot

Invoking this feature will render this space read only.

Permissions

Edit

Add

User/Group	Read	Write
bbranan	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Details

Items0

Recount

Created

2017-03-29T20:31:17

History

There is no history available for this space.

Duracloud Administrator Release v4.2.6 rev:20fac | ©2017 [DuraCloud](#) | [DuraSpace](#) | [Help Center](#) | [Contact Us](#)

DURASPACE™

DuraCloud Sync

localhost:8888/sync/status

DURACLOUD SYNC TOOL

StatusConfiguration

Overview

StopStart

Running

Sync Started

Wed Mar 29 16:37:42 EDT 2017

Queue Size

121

Error Count

0

Active Syncs (5)

Download History

0% of 0 MBsEncodeUtil.java

0% of 0 MBsInitUtil.class

0% of 0 MBsorg.duracloud.common.util.ChecksumUtilTest.txt

0% of 0 MBsLineParsingIterator\$Parseable.class

Recent Activity

File (hover for full path)	Action	Size	Duration	Completed
Securable.class	In sync	210 B	2 secs	03/29/17 04:38 PM
RestHttpHelper\$Method\$1.class	In sync	972 B	2 secs	03/29/17 04:38 PM
ContentItem.class	In sync	902 B	2 secs	03/29/17 04:38 PM

Queued for Synchronization (121)Errors (0)

File Path	Size	Last Modified Date
C:\work\code\duracloud\common\target\classes\org\duracloud\common\util\SerializationUtil.class	2.7 KB	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\util\SystemPropertiesReader.class	2.3 KB	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\util\TableSpec.class	1.1 KB	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\util\TagUtil.class	3.1 KB	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\util\UserUtil.class	254 B	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\web\EncodeUtil.class	1.2 KB	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\web\NetworkUtil.class	2.7 KB	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\web\RestHttpHelper\$1.class	240 B	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\web\RestHttpHelper\$Method\$5.class	973 B	01/06/2017 14:34 PM EST
C:\work\code\duracloud\common\target\classes\org\duracloud\common\web\RestHttpHelper\$Method\$6.class	975 B	01/06/2017 14:34 PM EST

DuraCloud Sync v4.2.0 rev 086cc | DuraCloud | DuraSpace

DuraCloud :: Spaces

←

→

↺

🏠

🔒 Secure

https://vault.duracloud.org/duradmin/spaces/sm/1063/sample-collection-1

☆

DURACLOUD™

Spaces

DPN

With the power of DuraCloud

Getting Started

Help

bbranan

Logout

DURACLOUD

VAULT

Spaces

+ Add Space

filter

Spaces

☒ sample-collection-1

☐ testspace

☐ x-duracloud-admin

☐ x-snapshot-metadata

Content Items

Showing 1 - 200 of 250 show more>

type prefix

☐ collection1/pom.xml

☐ collection1/src/main/assembly/dep.xml

☐ collection1/src/main/java/org/duracloud/common/constant/Constants.java

☐ collection1/src/main/java/org/duracloud/common/constant/ManifestFormat.java

☐ collection1/src/main/java/org/duracloud/common/error/DuraCloudCheckedException.java

☐ collection1/src/main/java/org/duracloud/common/error/DuraCloudExceptionCore.java

☐ collection1/src/main/java/org/duracloud/common/error/DuraCloudRuntimeException.java

☐ collection1/src/main/java/org/duracloud/common/error/ExceptionMessages.java

☐ collection1/src/main/java/org/duracloud/common/error/InvalidUsernameException.java

☐ collection1/src/main/java/org/duracloud/common/error/ManifestVerifyException.java

☐ collection1/src/main/java/org/duracloud/common/error/MessageFormattable.java

☐ collection1/src/main/java/org/duracloud/common/error/NoUserLoggedInException.java

☐ collection1/src/main/java/org/duracloud/common/error/RetryFlaggableException.java

☐ collection1/src/main/java/org/duracloud/common/model/AclType.java

☐ collection1/src/main/java/org/duracloud/common/model/ContentItem.java

☐ collection1/src/main/java/org/duracloud/common/model/Credential.java

☐ collection1/src/main/java/org/duracloud/common/model/DuraCloudUserType.java

☐ collection1/src/main/java/org/duracloud/common/model/Securable.java

Refresh

Upload

Space Detail

sample-collection-1

Delete

Audit Log

Manifest

Make publicly readable

Snapshot

Create Snapshot

Invoking this feature will render this space read only.

Permissions

Edit

Add

User/Group	Read	Write
bbranan	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Details

Items250

Recount

Created2017-03-29T20:31:17

History

There is no history available for this space.

Duracloud Administrator Release v4.2.6 rev:20fac | ©2017 [DuraCloud](#) | [DuraSpace](#) | [Help Center](#) | [Contact Us](#)

DURASPACE™

DuraCloud :: Spaces

←

→

↻

🏠

Secure | https://vault.duracloud.org/duradmin/spaces/sm/1063/vault_1063_sample-collection-1_2017-03-29-20-42-59?snapshot=true

☆

⋮

DURACLOUD

Spaces

DPN

DATA PROTECTION NETWORK

Spaces

+

Add Space

filter

Spaces

☐ sample-collection-1

☐ testspace

☐ x-duracloud-admin

☐ x-snapshot-metadata

Snapshots

vault_1063_sample-collection-1_2017-03-29-20-42-59

Snapshot Content Items

Refresh

Showing 1 - 200 of 251 show more>

type prefix

.collection-snapshot.properties

collection1/pom.xml

collection1/src/main/assembly/dep.xml

collection1/src/main/java/org/duracloud/common/constant/Constants.java

collection1/src/main/java/org/duracloud/common/constant/ManifestFormat.java

collection1/src/main/java/org/duracloud/common/error/DuraCloudCheckedException.java

collection1/src/main/java/org/duracloud/common/error/DuraCloudExceptionCore.java

collection1/src/main/java/org/duracloud/common/error/DuraCloudRuntimeException.java

collection1/src/main/java/org/duracloud/common/error/ExceptionMessages.java

collection1/src/main/java/org/duracloud/common/error/InvalidUsernameException.java

collection1/src/main/java/org/duracloud/common/error/ManifestVerifyException.java

collection1/src/main/java/org/duracloud/common/error/MessageFormattable.java

collection1/src/main/java/org/duracloud/common/error/NoUserLoggedInException.java

collection1/src/main/java/org/duracloud/common/error/RetryFlaggableException.java

collection1/src/main/java/org/duracloud/common/model/AclType.java

collection1/src/main/java/org/duracloud/common/model/ContentItem.java

collection1/src/main/java/org/duracloud/common/model/Credential.java

collection1/src/main/java/org/duracloud/common/model/DuraCloudUserType.java

collection1/src/main/java/org/duracloud/common/model/Securable.java

collection1/src/main/java/org/duracloud/common/model/SimpleCredential.java

collection1/src/main/java/org/duracloud/common/retry/ExceptionHandler.java

collection1/src/main/java/org/duracloud/common/retry/Retriable.java

collection1/src/main/java/org/duracloud/common/retry/Retrier.java

collection1/src/main/java/org/duracloud/common/servlet/DefaultFilter.java

collection1/src/main/java/org/duracloud/common/stream/ChecksumInputStream.java

Snapshot Detail

vault_1063_sample-collection-1_2017-03-29-20-42-59

Details

Description

Sample Snapshot 1

Alternate Ids

Snapshot Date

Wed Mar 29 2017 16:42:59 GMT-0400 (Eastern Daylight Time)

Source Host

vault.duracloud.org

Source Store

1063

Source Space

sample-collection-1

Status

WAITING_FOR_DPN

Preservation Network Member ID

e3fda0b7-5935-4c5a-a936-f05cc414222b

Content Item Count

251

Total Size

1.1 MB (1113526 bytes)

Snapshot History

Download

Show 10 entries

Search:

Date

Property

Wed Mar 29 2017 16:54:10 GMT-0400 (Eastern Daylight Time)

Name

Value

snapshot-action

SNAPSHOT_STAGED

snapshot-id

vault_1063_sample-collection-1_2017-03-29-20-42-59

Wed Mar 29 2017 16:53:53 GMT-0400 (Eastern Daylight Time)

Name

Value

snapshot-action

SNAPSHOT_INITIATED

initiating-user

bbranan@duraspace.org

snapshot-id

vault_1063_sample-collection-1_2017-03-29-20-42-59

Showing 1 to 2 of 2 entries

Previous

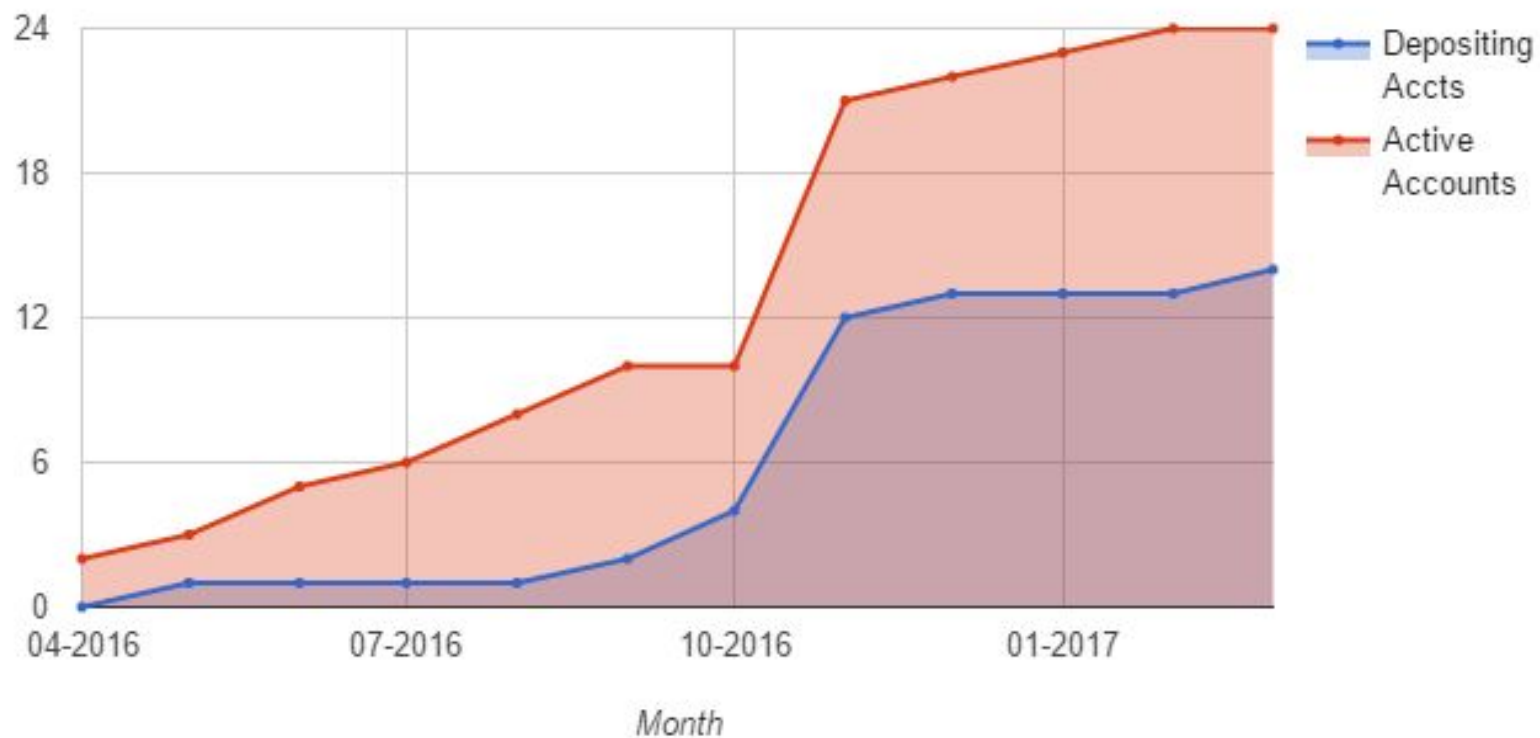
1

Next

Duracloud Administrator Release v4.2.6 rev:20fac | ©2017 [DuraCloud](#) | [DuraSpace](#) | [Help Center](#) | [Contact Us](#)

DURASPACE

Depositing Accounts and Active Accounts



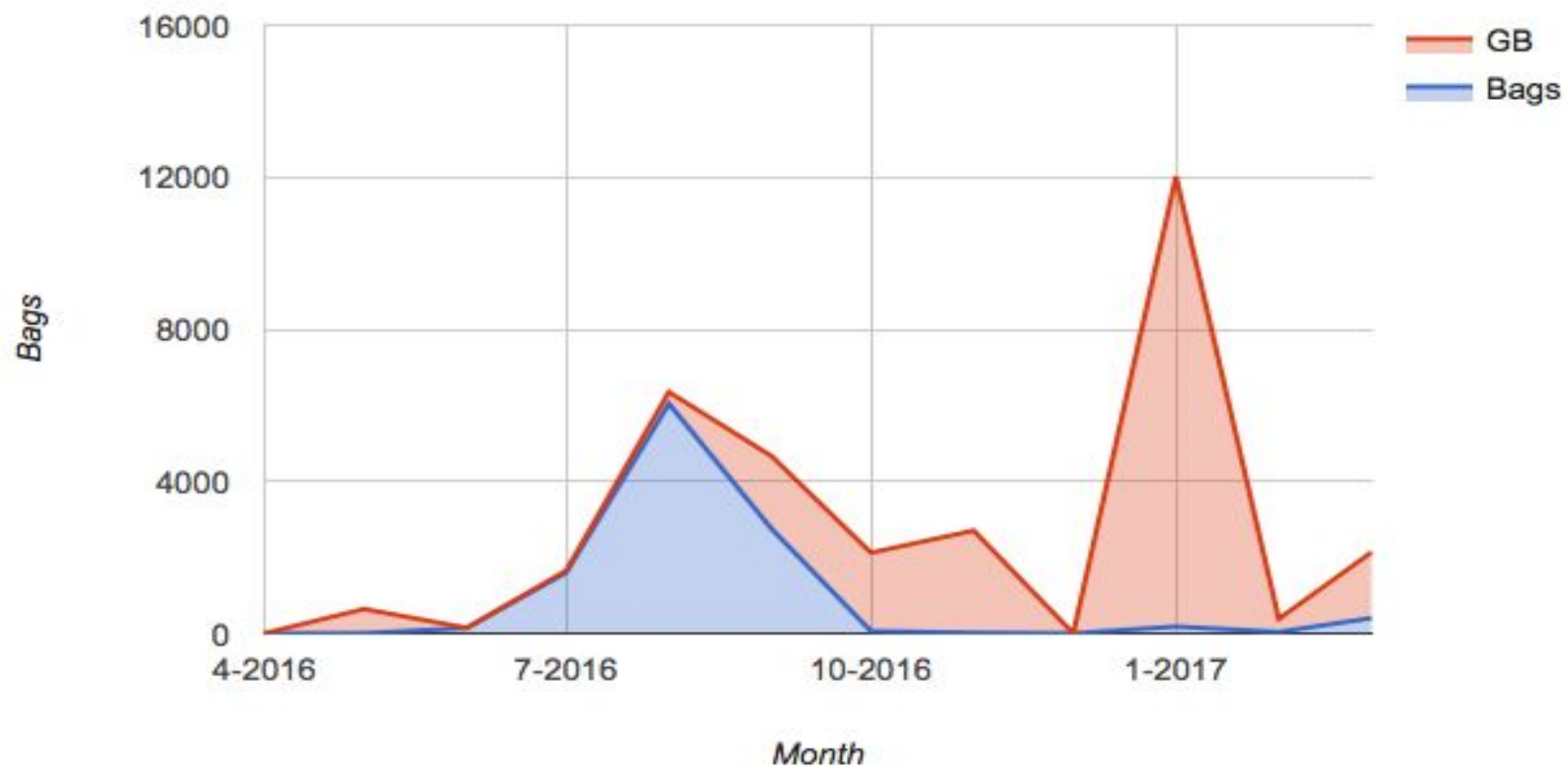
Deposits per month



Total Deposits over Time



Deposits per month



Active DPN Members using DuraCloud Vault

- Indiana University - Bloomington
- University of Iowa
- University of New Mexico
- Figshare
- Iowa State University
- University of Tennessee
- Indiana University Purdue University Indianapolis
- University of Kentucky
- California Institute of Technology
- University of Washington
- Brigham Young University
- Kansas State University
- Pennsylvania State University
- State University of New York at Buffalo
- Tufts University
- University of California Los Angeles
- University of California San Diego
- University of Illinois at Chicago
- University of Nebraska
- University of Wisconsin
- Utah State University
- University of Utah
- Vanderbilt University
- University of Kansas

Deposit Best Practices

1. Start early
2. Plan content deposit strategy
3. Prefer stable content sets
4. Group files into collections
5. Avoid very small snapshots (< 50 GB) where possible
6. Avoid very large snapshots (> 1 TB) where possible
 - a. Snapshot size is not limited by DPN bag size
7. Use archive files (tar or zip) to group very small files, but otherwise use sparingly
 - a. Archive files over the bag size limit must be split

Collections

- University of Iowa - Van Allen Belt Discovery
 - Original data tapes from the Explorer Missions
- University of Kentucky - Civil Rights History
 - Robert Penn Warren - oral history interviews
- Indiana University Purdue University Indianapolis (IUPUI)
 - Indianapolis Motor Speedway Collection

Challenges and Lessons Learned

1. Legal and structural framework is just as important to preservation as technology
2. Selecting the highest priority content is difficult
3. Heterogeneous systems: preservation benefits, coordination challenges
4. Data transfer takes time
5. Large data sets
6. System upgrades and software releases
7. Metadata creation and workflows

What is next?

- Expanding Membership
- Expanding Ingests
- Member/Vendor workflow automation
- Large Transfer Optimization
- Pilots
 - Bags over 250GB
 - Restoration Workflows
 - Peer to Peer Transfers - Heavy Users
 - Transaction Audits and Reporting

Questions?

Bill Branan

Services Technical Director
DuraSpace

bbranan@duraspace.org

Dave Pcolar

Chief Technology Officer
DPN

dave@dpn.org