Bridge Intake

The Bridge Intake Service provides data pulled down from the Duracloud Bridge App a path into Chronopolis. It currently handles bagging of snapshots, which includes validating files from the manifest given by Duracloud.

Links

- Gitlab: Bridge-Intake
- Github:Bridge-Restore-2

Installation

Prereqs

- Running Duracloud Bridge App to connect to
- Staging area for creating bags

RPM

Installation can be done through the provided rpm. This offers builds for both RHEL6 and RHEL7.

Installation Notes

Installed Files For RHEL6

```
/etc/init.d/bridge-intake
/usr/local/chronopolis/intake/dc
/usr/local/chronopolis/intake/dc/application.yml
/usr/local/chronopolis/intake/dc/bridge-intake.jar
```

Installed Files For RHEL7

```
/usr/lib/systemd/system/bridge-intake.service
/usr/local/chronopolis/intake/dc
/usr/local/chronopolis/intake/dc/application.yml
/usr/local/chronopolis/intake/dc/bridge-intake.jar
/usr/local/chronopolis/intake/dc/bridge-prepare
```

When running, the service will check for the following directories and create/apply permissions if they do not match:

- System Logging: /var/log/chronopolis/
- Journal data: /var/lib/chronopolis/data/

Configuration

application.yml

```
# Cron timer for how often the bridge is polled
bridge:
  poll: 0 0 0 * * *

# General chron configuration
## chron.node: obsolete;
## chron.workDirectory: directory for storing information used for ongoing operations, e.g. file csv
## chron.stage.bags.posix.id: the id of the bag staging area for this service
## chron.stage.bags.posix.path: the path on disk to the bag staging area for this service
chron:
  node: chron
  workDirectory: /tmp/chronopolis
  stage.bags:
    posix.id: 4
```

```
      posix.path: /scratch0/bags

# Chronopolis Ingest API Configuration
## ingest.api.endpoint: the http endpoint of the Ingest Server
## ingest.api.username: the username to connect to the Ingest Server with
## ingest.api.password: the password to connect to the Ingest Server with
ingest.api:
  endpoint: http://localhost:8000/
  username: admin
  password: admin

# ACE IMS Configuration
## ace.ims.endpoint: the fqdn of the ims
## ace.ims.port: the port to connect to the ims on; one of: 80, 443
##
ace.ims:
  endpoint: localhost
  port: 80
  waitTime: 6000


# Duracloud Bridge Configuration
# The storage areas the bridge app writes in to
# Connection information to query the bridge
## duracloud.bridge.name: an identifier for the bridge
## duracloud.bridge.snapshots: the path on disk to the snapshots directory
## duracloud.bridge.restores: the path on disk to the restores directory
## duracloud.bridge.username: the username to use when querying the bridge
## duracloud.bridge.password: the password to use when querying the bridge
## duracloud.bridge.password: the http endpoint of the bridge
duracloud:
  bridge:
    - name: bridge-0
      snapshots: /bridge-0/snapshots/
      restores: /bridge-0/restore/
      username: replace-me
      password: replace-me
      endpoint: http://localhost:8080/
    - name: bridge-1
      snapshots: /bridge-1/snapshots/
      restores: /bridge-1/restore/
      username: replace-me
      password: replace-me
      endpoint: http://localhost:8081/

# Logging configuration
logging:
  file: /var/log/bridgeintake/intake.log
  level:
    org.springframework: ERROR
    org.hibernate: ERROR
    org.chronopolis: debug
    org.chronopolis.intake.duracloud.config: trace

# Extraneous settings
#   disable SNI on https connections - false is the default and recommended value
disableSNI: false
```

Release Notes

Release 3.0.1

09 April, 2019

- **Bug Fix:** Resolve error with startup failing to configure our beans

Release 3.0.0

21 March, 2019

- Remove deprecated DPN workflow
- Update .gitlab-ci.yml to cache maven artifacts

Release 2.4.3

21 March, 2019

- Update Chronopolis and chron-test dependencies to latest

Release 2.4.2

23 January, 2019

- Bump Chronopolis dependency version
- **Bug Fix:** Fix csv upload to create the MultiPart data correctly

Release 2.4.1

22 January, 2019

- Bump Chronopolis dependency version

Release 2.4.0

11 January, 2019

- Bridge consolidation work to bring multiple bridge querying to a single service
- Update notification of missing Depositors to fire only once for the lifetime of the bridge
- Fix Bug resulting in successive snapshot runs not being processed if a check against the Depositor fails
- Update ChronopolisIngest for 3.0.0 workflow

Release 2.3.2

09 August, 2018

- Fixed bug which was preventing deserialization of certain classes
- Temporarily disabled HttpFilter because of bug in Ingest Server

Release 2.3.1

07 August, 2018

- Updated logic for appending a prefix to a depositor so that it is used in every operation

Release 2.3.0

10 July, 2018

- Code licensed under BSD-3
- Use separate thread pools when executing Ingestion tasks
    - Long IO operations - bagging, cleaning, etc
    - Short IO operations - HTTP requests
- Bring tokenization process up to date with Chronopolis Core code
- Test that a depositor exists before attempting to push a Bag to Chronopolis or DPN

Release 2.2.1

18 May, 2018

- Fix bug where constraint for bag size was checked incorrectly
- Fix bug where tokens were not being created because a dependency was missing

Release 2.2.0

14 May, 2018

- Fix bug where we could potentially create the wrong number of replications
- Add constraint satisfaction when creating DPN Replications
- Refactor code for Chronopolis and DPN Ingest to better facilitate working on snapshots concurrently
    - Break up DPN Ingestion steps into multiple classes

Release 2.1.0

8 February, 2018

- Actively clean storage upon completion of data replicating
    - For chronopolis: Ingest reports PRESERVED
    - For DPN: All replications completed
- Mark staging as inactive in Ingest for completed replications in Chronopolis

Release 2.0.3

29 November, 2017

- Query on Bags we created when running Tokenization

Release 2.0.2

9 November, 2017

- Remove creation of /var/log/chronopolis from rpm and instead create it on application startup

Release 2.0.1

8 November, 2017

- Use correct properties when creating Ingest Requests to Chronopolis

Release 2.0.0

30 October, 2017

- Integrate Chronopolis 2.0.0 Changes
    - Import tokenization module in order to create ACE Tokens for incoming data
    - Use new properties classes for configuration of storage and api communication
    - Capture basic metrics when creating bags
    - Update ingest flow to create bag and fixity
- Updates to build and init scripts
    - Support EL6 and EL7
    - Quality of life changes to the sysv init script based on rpmlint

Release 1.5.1

30 June, 2017

- Fix bug that caused replications to be created for each node in DPN

Release 1.5.0

26 June, 2017

- Add a prefix for chronopolis bound collections
- When creating replications for dpn, choose replicating nodes in a deterministic manner per snapshot so that all bags in a snapshot end up at the same DPN nodes.