NetarchiveSuite 3.18.0 Release Notes

Released 14.12.2011.

Highlights

We have improved the generation of our indices, which primarily should speed up the generation of the index generated for the snapshot harvests. Furthermore, we have made some optimizations to the GUI. Finally, a lot of database cleanup that was handled properly, has now been taken care of.

Upgrade-notes

Be sure to run the `dk.netarkivet.harvester.tools.HarvestdatabaseUpdateApplication` tool after installation of the new release, but before starting the Netarchivesuite system. See the additional tools manual for further details.

Note that the harvesting setting:

```
<sendStatusDelay>60</sendStatusDelay>
```

has been replaced by the two settings

```
<sendReadyInterval>30</sendReadyInterval>
<sendReadyDelay>1000</sendReadyDelay>
```

See HarvesterSettings.java and Harvester settings.xml for details.

Furthermore a number of new archiver settings has been added in relation to the improved indexing-mechanism:

```
settings.archive.indexserver.requestdir
settings.archive.indexserver.maxclients
settings.archive.indexserver.maxthreads
settings.archive.indexserver.checkinterval
settings.archive.indexserver.indexingtimeout
settings.archive.indexserver.listeningcheckinterval
```

And added are also new settings related to the database pool mechanism for the archive database (similar to the one for the Harvest Database):

```
settings.archive.admin.database.pool.minSize
settings.archive.admin.database.pool.maxSize
settings.archive.admin.database.pool.acquireInc
settings.archive.admin.database.pool.maxStm
settings.archive.admin.database.pool.maxStmPerConn
settings.archive.admin.database.pool.idleConnTestPeriod
settings.archive.admin.database.pool.idleConnTestQuery
settings.archive.admin.database.pool.idleConnTestOnCheckin
```

See ArchiveSettings.java and archive settings.xml for details about these new archive settings.

For those deploying the software to Windows servers, note that we have now disabled Console-logging on this category of servers, as it occasionally results in hanging bitarchive applications necessitationg a restart of the applications. See issue NAS-1993.

Note that we have added notifications in the HarvestJobManager, if it thinks it is skipping a scheduling, because the previous scheduling is still running (NAS-1968). A known problem with this is that the harvestJobManager will send out lots of notifications during the scheduling of snapshot-harvests. (NAS-2005).

Known problems and work-arounds for these.

- NAS-2030): If your NAS installation includes two Derby database servers, they currently write to the same database log ($INSTALLDIR/derby.log) and use the same
  properties file ($INSTALLDIR/derby.properties). The fix is to alter the derby.system.home for one of datatases, e.g
  $INSTALLDIR/adminDB as follows
  after having shut down all apps using the database:

```
mkdir $INSTALLDIR/adminserverdir
mv $INSTALLDIR/adminDB $INSTALLDIR/adminserverdir/
add -Dderby.system.home=$INSTALLDIR/adminserverdir  after -Xmx1936m in
the scripts $INSTALLDIR/conf/kill_external_admin_database.sh
and $INSTALLDIR/conf/start_external_admin_database.sh
```

- The adminDB now uses the same pool mechanism as the harvestDB, i.e. c3p0. It seems that the defaults causes access delays in the database access. This can be fixed by overriding two of the default pool settings (idleConnTestOnCheckin, idleConnTestQuery) in the deploy configuration:

```
<archive>
  <admin>
    <database>
      <dir>adminDB</dir>
      <pool>
        <idleConnTestQuery>SELECT COUNT(*) FROM
replica</idleConnTestQuery>
        <idleConnTestOnCheckin>false</idleConnTestOnCheckin>
      </pool>
    </database>
```
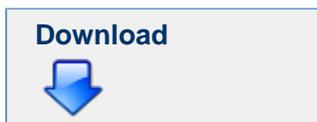
Full list of issues resolved in this release.

T	Key	P	Summary

NAS-1850		IndexArrayOutOfBounds exception in job progress chart generation
NAS-1843		Reformat french translation files
NAS-1409		ArgumentNotValid thrown when executing waitForReportGeneration during cleanup
NAS-1407		SQL truncation error in RunningJobsInfoDBDAO storeFrontierReport()
NAS-1886		FileUtils.RemoveFile throws IOFailure if it cannot delete the file

5 issues

Known issues

T	Key	P	Summary
	NAS-2051		Only one indexing operation allowed at a time
	NAS-2050		The indexes generated by the Indexserver has too big a memory demand
	NAS-1993		BatchJob freezes KB Bitarchive
	NAS-1978		Too much waiting time in indexserver when generating small indices
	NAS-2037		Thread accumulation in IndexServerApplication
	NAS-2043		IDs in messages sent to the Indexserver are not unique!
	NAS-2044		Display of crawl log lines doesn't scale
	NAS-2040		In on of our Bitpreservation classes (DatabaseBasedActiveBitPreservation) stores internally the result of batch-jobs as arrays
	NAS-2042		Change the way the Indexserver names the indices constructed
	NAS-2065		broad crawl job stopped because of timeout is not registered as "aborted because of time limit"
	NAS-2085		UpdateChecksum: java.lang.OutOfMemoryError: GC overhead limit exceeded
	NAS-2031		Harvest Jobs in state submitted can accumulate
	NAS-2026		Arcrepository.store method seems to proceed with a store retry even though upload is completed on all replicas
	NAS-2022		Possible IOFailure exception thrown by FileUtils.copyFile is not caught by method DeployConfiguration.copyNetarchiveSuiteFile
	NAS-2030		Both adminDB and harvestDB if derby-based uses same derby.system.home value
	NAS-2027		The wrong thing is done during upload recovery in arcrepository.startUpload
	NAS-2021		A lot of exceptional return values are ignored in our code
	NAS-2014		Strange top spaces using firefox and chrome
	NAS-1986		Benchmark is not a benchmark
	NAS-1983		Upload failure in TEST11A (checksum mismatch)

Showing 20 out of 29 issues

Download

Manuals

Javadoc