

DSpace 7 Performance Testing

This page describes an effort to measure key aspects of DSpace 7's performance compared to previous versions.

The goal is to establish a feedback loop while DSpace 7 is under development, in order to identify areas worth investigation/improvement.

Areas of interest

For a variety of commonly-accessed pages in a pre-populated DSpace instance, the following will be measured:

- Server memory use
- Server CPU use
- Page load time
- Browser memory use

Where it makes sense, measurements will be taken while simulated load is put on the server.

Test Repository

A test repository has been generated with the following size and structure:

- 288 collections:
 - 4 top-level communities
 - 8 subcommunities in each top-level community
 - 9 collections in each subcommunity
- 69120 items with 705024 bitstreams:
 - 36864 "Tiny" items distributed evenly among 72 collections, each with 1 bitstream:
 - 1 single-byte bitstream in the OTHER bundle
 - 18432 "Small" items distributed evenly among 72 collections, each with 4 bitstreams:
 - 1 single-byte bitstream in the OTHER bundle
 - 1 single-page text PDF in the ORIGINAL bundle
 - 1 derived txt bitstream in the TEXT bundle
 - 1 derived jpg bitstream in the THUMBNAIL bundle
 - 9216 "Medium" items distributed evenly among 72 collections, each with 13 bitstreams:
 - 10 single-byte bitstreams in the OTHER bundle
 - 1 ten-page text PDF in the ORIGINAL bundle
 - 1 derived txt bitstream in the TEXT bundle
 - 1 derived jpg bitstream in the THUMBNAIL bundle
 - 4608 "Big" items distributed evenly among 72 collections, each with 103 bitstreams:
 - 100 single-byte bitstreams in the OTHER bundle
 - 1 hundred-page text PDF in the ORIGINAL bundle
 - 1 derived txt bitstream in the TEXT bundle
 - 1 derived jpg bitstream in the THUMBNAIL bundle

To download and install this repository into a fresh install of DSpace 5, 6, or 7:

- Visit https://drive.google.com/open?id=1MK3drQsa3KtZCOBtrKBXnE4oMo_C2Ub
- Download the test-x.sql.gz file corresponding to your major version of DSpace and restore into postgres via:
 - gunzip test-7.sql.gz
 - dropdb dspace
 - createdb dspace
 - psql dspace 'create extension pgcrypto'
 - psql dspace < test-7.sql
- Download the assetstore, containing the PDFs, text, and other bitstreams, then install via:
 - mv assetstore.tar.gz /dspace/
 - cd /dspace
 - tar -xvf assetstore.tar.gz
 - rm assetstore.tar.gz
- Once the database and assetstore are installed, you can start DSpace and create the discovery index:
 - /dspace/bin/dspace index-discovery
 - The search index will consume about 5G when built
- Note: The database has a built-in admin user: dspace@test, with password test.

To generate your own test data, here's a script that can be used to create test content:

- DSpaceObject Generator: <https://github.com/cwilper/dsogen>

Environment

(TBD)

Methodology

Measurements for each of the following were taken as described below.

Server memory use

- java: jmap -histo:live (total heap)
- node: pm2 (mem column)
- system: free -m (used column, minus buffers/cache)

Server cpu use

- java and node: ps -o %cpu -p
- system: uptime (1-minute cpu load average)

Page load time

- **(Undecided)**

Browser memory use

- Chrome Devtools' Memory profiler

Test Results

(TBD - goal is to test periodically prior to final 7.0 release)