Performance Testing & Content Access Control

Seth Shaw
University of Nevada, Las Vegas

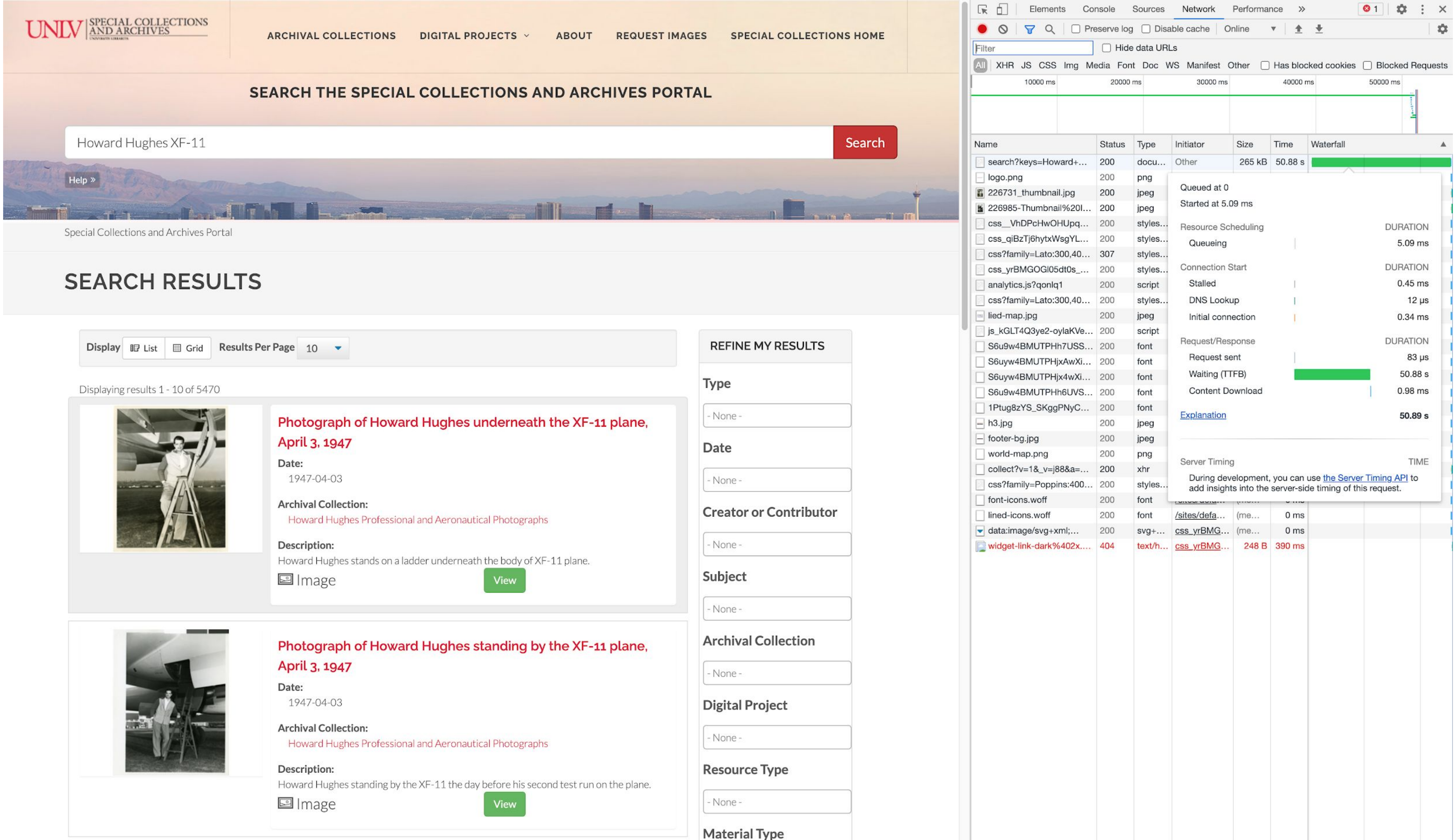
Current Node Counts by Type

Object Type	**Count**
Archival Collection-Level Description	3,955
Archival Component-Level Description	229,304
Archival Container Record	17,379
Digital Object Record	103,487
Everything Else	*803*
Total	**354,928**

With *a lot* more digital objects on the way...

“The maximum page load time across the entire site should always remain below eight seconds. [Digital Object and search] pages should have a maximum page load time of five seconds. The front page of the site should have a maximum page load time of three seconds.”

Sheltren, Jeff, et al. *High Performance Drupal: Fast and Scalable Designs*. 1st ed., Sebastopol, CA, O’Reilly Media, 2013.



SEARCH THE SPECIAL COLLECTIONS AND ARCHIVES PORTAL

Howard Hughes XF-11

Special Collections and Archives Portal

SEARCH RESULTS

Displaying results 1 - 10 of 5470

Photograph of Howard Hughes underneath the XF-11 plane, April 3, 1947

Date:
1947-04-03

Archival Collection:
Howard Hughes Professional and Aeronautical Photographs

Description:
Howard Hughes stands on a ladder underneath the body of XF-11 plane.

Image

Photograph of Howard Hughes standing by the XF-11 plane, April 3, 1947

Date:
1947-04-03

Archival Collection:
Howard Hughes Professional and Aeronautical Photographs

Description:
Howard Hughes standing by the XF-11 the day before his second test run on the plane.

Image

REFINE MY RESULTS

Type
- None -

Date
- None -

Creator or Contributor
- None -

Subject
- None -

Archival Collection
- None -

Digital Project
- None -

Resource Type
- None -

Material Type

Name	Status	Type	Initiator	Size	Time
search?keys=Howard+…	200	docu…	Other	265 kB	50.88 s
logo.png	200	png			
226731_thumbnail.jpg	200	jpeg			
226985-Thumbnail%20I…	200	jpeg			
css__VhDPcHwOHUpq…	200	styles…			
css_qiBzTj6hytxWsgYL…	200	styles…			
css?family=Lato:300,40…	307	styles…			
css_yrBMGOGl05dt0s_…	200	styles…			
analytics.js?qonlq1	200	script			
css?family=Lato:300,40…	200	styles…			
lied-map.jpg	200	jpeg			
js_kGLT4Q3ye2-oylaKVe…	200	script			
S6u9w4BMUTPHh7USS…	200	font			
S6uyw4BMUTPHjxAwXi…	200	font			
S6uyw4BMUTPHjx4wXi…	200	font			
S6u9w4BMUTPHh6UVS…	200	font			
1Ptug8zYS_SKggPNyC…	200	font			
h3.jpg	200	jpeg			
footer-bg.jpg	200	jpeg			
world-map.png	200	png			
collect?v=1&_v=j88&a=…	200	xhr			
css?family=Poppins:400…	200	styles…			
font-icons.woff	200	font			
lined-icons.woff	200	font	/sites/defa…	(me…	0 ms
data:image/svg+xml;…	200	svg+…	css_yrBMG…	(me…	0 ms
widget-link-dark%402x.…	404	text/h…	css_yrBMG…	248 B	390 ms

Queued at 0
Started at 5.09 ms

Resource Scheduling	DURATION
Queueing	5.09 ms
Connection Start	**DURATION**
Stalled	0.45 ms
DNS Lookup	12 µs
Initial connection	0.34 ms
Request/Response	**DURATION**
Request sent	83 µs
Waiting (TTFB)	50.88 s
Content Download	0.98 ms
Explanation	50.89 s

Server Timing TIME

During development, you can use the Server Timing API to add insights into the server-side timing of this request.

Enter WebProfiler

Available in devel.
Requires installing additional JS libraries.

The view was taking *forever* to complete but SOLR logs showed it was returning results in milliseconds.

Islandora Performance Analysis Report

Public version: https://seth-shaw-unlv.github.io/2021-02-18/performance_analysis

Was it our custom theme? 🤔

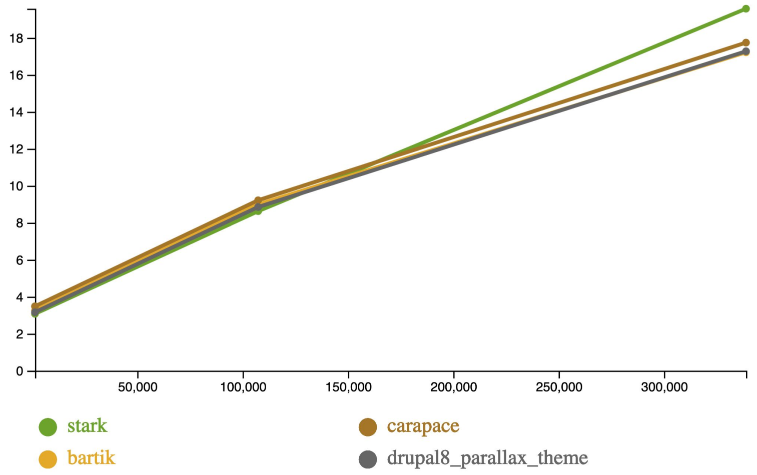


Fig. 1—Average TTFB for a digital object page for each theme at increasing load levels. Note that bartik is obscured by drupal8_parallax_theme because they are so close.

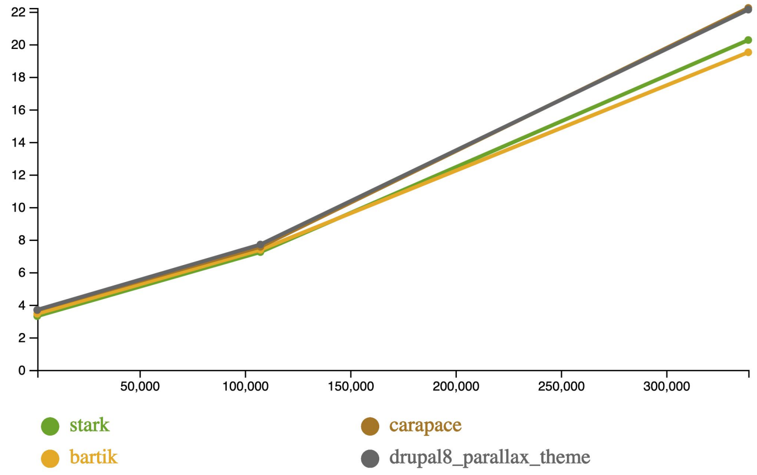


Fig. 2—Average TTFB for a search page at increasing load levels. Each color is a separate theme. Note that carapace is obscured by drupal8_parallax_theme because they are so close.

No, our theme performs on par with others.

Enter XHProf/Tideways...

A pain to install, but worth the effort...

It quickly showed that most of the request time is consumed by database queries coming from permissions_by_term.

Performance with/without Permissions by Term

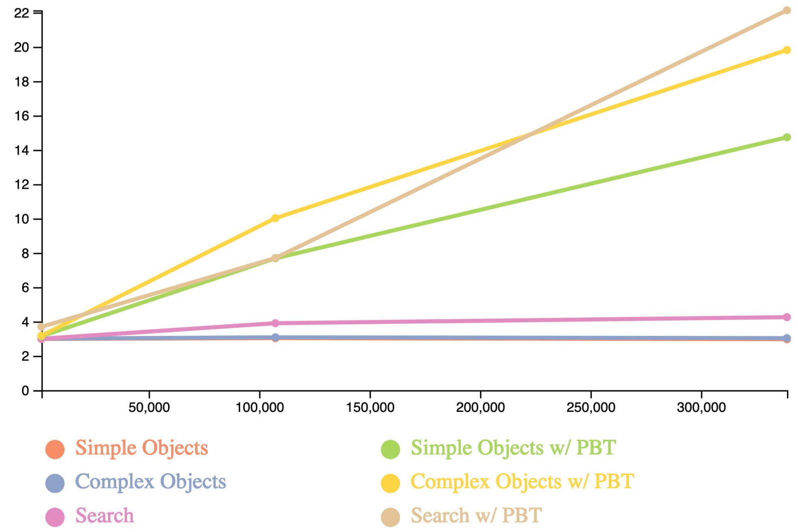


Fig. 3—Shows the average time to first byte for the different types of pages (search, simple objects, and complex objects), using the Special theme, with and without Permissions by Term (PBT) enabled as load increases. Note that the without permissions is obscured behind complex objects because they are so close.

Disabling Permissions by Term brought all the tests under 4 seconds!
But what to do about it?

Content Access Control Solutions Investigation

Public version: https://seth-shaw-unlv.github.io/2021-02-19/content_access_control

Why is Permissions by Term enabled, anyway?

- We have digitized photographs that require special permissions to access.

- We are also loading born-digital materials that are temporarily restricted but still need preservation.

Restricted items currently only represent 2.2% of our digital object records, but the volume it is expected to grow rapidly as we begin adding more born-digital content.

Why not fix Permissions by Term?

- Others have identified performance issues but the ticket was mostly ignored.
- I found one fix that reduced the performance hit by nearly 50%, but more would be necessary.
- I'm not optimistic about my patches being accepts as a previous one was ignored.
- I would rather not keep re-rolling local patches.

Content Access Control Options

- Access by entity: needs a maintainer and an unknown number of fixes to make it stable again.
- Advanced Access / Role Access Control: maintenance is TBD and media with role references are not being restricted as I would expect
- Content Access: will not scale for our repository.
- Embargoes: embargoed nodes and media still appear in Views.
- Group: no simple way to move content from one group to another.
- Taxonomy Access Control Lite: among other issues, this module's community is dead.

Top Contenders

Advanced Access / Role Access Control:

Pro

- Relatively simple migration
- Mimics Drupal grants for Media

Con

- Uncertain future
- Media logic needs tweaking
- No IP-based restrictions (yet)

Embargoes:

Pro

- Built for Islandora
- Time and IP-based restrictions
- Community engagement likely to grow

Con

- Needs heavy refactoring
- Adds more entities
- Items still appearing in Views

Moving Forward: *Attempt* to fix Advanced Access...

If it works by the end of February:

- Code an update hook to transform the existing term references into the corresponding role references.

If it doesn't work by the deadline:

- Purge existing items with restricted-access from the system
- Work on implementing Embargoes and loading restricted items ***after public launch.***