[GraphQL] Convert monolithic query into a pattern of small queries #15974

         Edit   Open with ▾

⑂ **Merged**

ncksllvn merged 52 commits into `master` from `nps-individual-queries` 🗐 on Feb 11

💬 Conversation 21   ⦿ Commits 52   ☑ Checks 4   ± Fil **+1,689 −174** ■■■■□

---

ncksllvn commented on Feb 8 • edited ▾   Member

Description

This PR conducts an experiment to find out what the GraphQL performance looks like when each node fragment (page type) has its own self-contained `nodeQuery`, instead of combining all of them into the notorious million-line monolithic GraphQL query that we see sometimes taking close to 20 minutes.

Note that these changes are backwards-compatible, meaning that the preview server, the original monolithic GraphQL query, transformers, etc. all continue to function the same way.

Ultimately, what I did here was -

1. Go through `GetAllPages.graphql.js` to find each type of top-level node fragment we are using.

2. Open each file representing a node fragment. Change its module to export the original plain fragment for backwards-compatibility, and a self-contained

Reviewers ⚙

🤖 va-vfs-bot   💬

🧑 mpelzsherman   ✓

🧑 brangi   ✓

🧑 theodur   ✓

🧑 kelsonic   ✓

Assignees ⚙

No one—assign yourself

Labels ⚙

None yet

Projects ⚙

None yet

Milestone ⚙

No milestone

`nodeQuery` for this experiment. Each node fragment relied on a set of global paragraph fragments that normally the node fragment has no direct reference to, so I had to also import the paragraph fragments file and add those into the self-contained `nodeQuery` as well.

3. Write a new method onto `build/drupal/api.js` that does the "individualized" version of the GraphQL query, where it executes in parallel all of the `nodeQuery`s.

4. Call this new method from `metalsmith-drupal` when build type isn't prod.

Testing done

Sometimes the Drupal GraphQL API (both monolithic and individualized) is "hot" and will serve the entire response in seconds. Seriously, sometimes like 3 seconds. This is likely when Drupal's cache is ready and the content unchanged. The case we are worried about though is when the cache is cold, which was kinda tough to replicate. Here's an example though, where it appears to shave off **12 minutes** .

The output is kind of verbose, but obviously we can wrap it with the `verbose` build arg. It's useful to see which specific queries are the most expensive. For instance, the `nodeQuery` to get all VA Forms has a lot of nodes (almost 500) and takes almost 3 minutes on its own, which means it is sometimes the bottleneck. We might be able to make this more efficient by splitting it into two queries, with one taking advantage of Drupal's GraphQL cache.

These result are taken from http://jenkins.vfs.va.gov/blue/organizations/jenkins/testing%2Fvets-website/detail/nps-individual-queries/20/pipeline/100.

The "Multiple queries" ran on the vagovstaging build whereas the monolithic query ran on the vagovprod build. Both use Drupal's prod environment.

▼ Multiple queries - finishes at 3 minutes, 2 seconds

```
[2021-02-08T14:59:44.610Z] Step 3 of 29 start: Get
Drupal content
[2021-02-08T14:59:44.610Z] Drupal content
unavailable in local cache:
```

```
/application/.cache/vagovstaging/drupal/pages.json
[2021-02-08T14:59:44.610Z] Attempting to load
Drupal content from API at http://internal-dsva-
vagov-prod-cms-2000800896.us-gov-west-
1.elb.amazonaws.com
[2021-02-08T14:59:44.610Z] Pulling from Drupal via
GraphQL...
[2021-02-08T14:59:44.610Z] 2 page nodes loaded
from GetNodeHealthServicesListingPages
[2021-02-08T14:59:44.610Z]
GetNodeHealthServicesListingPages done: 1.146s
[2021-02-08T14:59:44.610Z] 13 page nodes loaded
from GetNodeHealthCareLocalFacilityPages
[2021-02-08T14:59:44.610Z]
GetNodeHealthCareLocalFacilityPages done: 1.154s
[2021-02-08T14:59:44.610Z] 158 page nodes loaded
from GetNodeEventPages
[2021-02-08T14:59:44.610Z] GetNodeEventPages done:
1.157s
[2021-02-08T14:59:44.610Z] 1 page nodes loaded
from GetNodeOffices
[2021-02-08T14:59:44.610Z] GetNodeOffices done:
1.386s
[2021-02-08T14:59:44.610Z] 2 page nodes loaded
from GetNodeHealthCareRegionPages
[2021-02-08T14:59:44.610Z]
GetNodeHealthCareRegionPages done: 1.481s
[2021-02-08T14:59:44.610Z] 12 page nodes loaded
from GetNodeLandingPages
[2021-02-08T14:59:44.611Z] GetNodeLandingPages
done: 1.510s
[2021-02-08T14:59:44.611Z] 29 page nodes loaded
from GetNodePressReleasePages
[2021-02-08T14:59:44.611Z]
GetNodePressReleasePages done: 1.512s
[2021-02-08T14:59:44.611Z] 3 page nodes loaded
from GetNodeEventListingPage
[2021-02-08T14:59:44.611Z] GetNodeEventListingPage
done: 2.896s
[2021-02-08T14:59:44.611Z] 108 page nodes loaded
from GetNodePersonProfiles
[2021-02-08T14:59:44.611Z] GetNodePersonProfiles
done: 3.303s
[2021-02-08T14:59:44.611Z] 2 page nodes loaded
from GetNodeLeadershipListingPages
[2021-02-08T14:59:44.611Z]
GetNodeLeadershipListingPages done: 3.623s
[2021-02-08T14:59:44.611Z] 1 page nodes loaded
from GetNodePublicationListingPages
[2021-02-08T14:59:44.611Z]
GetNodePublicationListingPages done: 3.704s
[2021-02-08T14:59:44.611Z] 76 page nodes loaded
from GetNodeQa
```

```
[2021-02-08T14:59:44.611Z] GetNodeQa done: 3.717s
[2021-02-08T14:59:44.611Z] 42 page nodes loaded
from GetNodeNewsStoryPages
[2021-02-08T14:59:44.611Z] GetNodeNewsStoryPages
done: 3.851s
[2021-02-08T14:59:44.611Z] 274 page nodes loaded
from GetNodePages
[2021-02-08T14:59:44.611Z] GetNodePages done:
4.041s
[2021-02-08T14:59:44.611Z] 1 page nodes loaded
from GetNodeBasicLandingPage
[2021-02-08T14:59:44.611Z] GetNodeBasicLandingPage
done: 4.631s
[2021-02-08T14:59:44.611Z] 2 page nodes loaded
from GetNodeLocationsListingPages
[2021-02-08T14:59:44.611Z]
GetNodeLocationsListingPages done: 4.846s
[2021-02-08T14:59:44.611Z] 365 page nodes loaded
from GetArchivedNodeEventPages
[2021-02-08T14:59:44.611Z]
GetArchivedNodeEventPages done: 4.875s
[2021-02-08T14:59:44.611Z] 13 page nodes loaded
from GetNodeMultipleQaPages
[2021-02-08T14:59:44.611Z] GetNodeMultipleQaPages
done: 4.878s
[2021-02-08T14:59:44.611Z] 0 page nodes loaded
from GetNodeMediaListImages
[2021-02-08T14:59:44.611Z] GetNodeMediaListImages
done: 5.020s
[2021-02-08T14:59:44.611Z] 2 page nodes loaded
from GetNodeStoryListingPages
[2021-02-08T14:59:44.611Z]
GetNodeStoryListingPages done: 5.064s
[2021-02-08T14:59:44.611Z] 2 page nodes loaded
from GetNodeVamcOperatingStatusAndAlerts
[2021-02-08T14:59:44.611Z]
GetNodeVamcOperatingStatusAndAlerts done: 5.130s
[2021-02-08T14:59:44.611Z] 0 page nodes loaded
from GetNodeMediaListVideos
[2021-02-08T14:59:44.611Z] GetNodeMediaListVideos
done: 5.883s
[2021-02-08T14:59:44.611Z] 10 page nodes loaded
from GetNodeSupportResourcesDetailPage
[2021-02-08T14:59:44.611Z]
GetNodeSupportResourcesDetailPage done: 5.917s
[2021-02-08T14:59:44.611Z] 0 page nodes loaded
from GetCampaignLandingPages
[2021-02-08T14:59:44.611Z] GetCampaignLandingPages
done: 5.988s
[2021-02-08T14:59:44.611Z] 6 page nodes loaded
from GetNodeStepByStep
[2021-02-08T14:59:44.611Z] GetNodeStepByStep done:
6.005s
```

```
[2021-02-08T14:59:44.611Z] 85 page nodes loaded
from GetNodeHealthCareRegionDetailPage
[2021-02-08T14:59:44.611Z]
GetNodeHealthCareRegionDetailPage done: 6.082s
[2021-02-08T14:59:44.611Z] 0 page nodes loaded
from GetNodeChecklist
[2021-02-08T14:59:44.611Z] GetNodeChecklist done:
6.087s
[2021-02-08T14:59:44.611Z] 2 page nodes loaded
from GetNodePressReleaseListingPages
[2021-02-08T14:59:44.611Z]
GetNodePressReleaseListingPages done: 6.329s
[2021-02-08T15:00:16.599Z]
GetNonNodeWebsiteComponents done: 37.508s
[2021-02-08T15:02:52.973Z] 438 page nodes loaded
from GetNodeVaForms
[2021-02-08T15:02:52.973Z] GetNodeVaForms done:
3:01.635 (m:ss.mmm)
[2021-02-08T15:02:52.973Z] Finished all queries -
1649 pages
[2021-02-08T15:02:52.973Z] http://internal-dsva-
vagov-prod-cms-2000800896.us-gov-west-
1.elb.amazonaws.com response time: : 3:01.642
(m:ss.mmm)
[2021-02-08T15:02:52.973Z] Total time to load
content from GraphQL: 3:02.001 (m:ss.mmm)
[2021-02-08T15:02:52.973Z] Drupal successfully
loaded!
[2021-02-08T15:02:52.973Z] Successfully piped
Drupal content into Metalsmith!
[2021-02-08T15:02:52.973Z] Step 3 end [191751ms]:
Get Drupal content
```

▼ Monolithic query: 14 minutes, 48 seconds

```
[2021-02-08T14:59:44.590Z] Step 3 of 29 start: Get
Drupal content
[2021-02-08T14:59:44.590Z] Drupal content
unavailable in local cache:
/application/.cache/vagovprod/drupal/pages.json
[2021-02-08T14:59:44.590Z] Attempting to load
Drupal content from API at http://internal-dsva-
vagov-prod-cms-2000800896.us-gov-west-
1.elb.amazonaws.com
[2021-02-08T15:14:36.188Z] http://internal-dsva-
vagov-prod-cms-2000800896.us-gov-west-
1.elb.amazonaws.com response time: : 14:48.555
(m:ss.mmm)
[2021-02-08T15:14:36.188Z] Total time to load
content from GraphQL: 14:48.902 (m:ss.mmm)
[2021-02-08T15:14:36.188Z] Drupal successfully
loaded!
```

```
[2021-02-08T15:14:36.188Z] Successfully piped
Drupal content into Metalsmith!
[2021-02-08T15:14:36.188Z] Step 3 end [898942ms]:
Get Drupal content
```

Test this locally but pulling the branch and running `yarn build:content --pull-drupal` .

Acceptance criteria

☑ Peak GraphQL response times are drastically reduced

Definition of done

☑ Events are logged appropriately

☑ Documentation has been updated, if applicable

☑ A link has been provided to the originating GitHub issue
   (or connected to it via ZenHub)

☑ No sensitive information (i.e. PII/credentials/internal
   URLs/etc.) is captured in logging, hardcoded, or specs

🎉 1   ❤️ 1   👀 1

ncksllvn added 31 commits on Feb 5

○ Begin pattern for separate nodeQuery   28bd8d7

○ Add landing page, health care region   5e9a4e1
   page

○ Add bio page; fix health care region   ✕ 53c3c7c
   page

○ parallelize the requests   ✕ 3e4da4c

○ add office   1354da3

○ Add health care local facility   5408aea

○ add services listing   ✕ 7dddef6

○ add news stories   1297990
```

| | | |
|---|---|---|
| ○ | add press release page | 23f0e63 |
| ○ | add press release listing | a193a05 |
| ○ | add event listing pages | 15045fb |
| ○ | add story listing, event | ✕ 1e3c592 |
| ○ | add location lists and leader lists | ✕ dbbe040 |
| ○ | add GetNodeVamcOperatingStatusAndAlerts | ✕ 5212582 |
| ○ | add publication GetNodePublicationListingPages, GetNodeHealthCareRegi… ··· | ✕ 45a9760 |
| ○ | Add the non-node queries | 6de7ae7 |
| ○ | fail on graphql errors | 516880f |
| ○ | Add NodeQA and fix components | 74e9e46 |
| ○ | Add multiple faq | ✕ fbc3b9c |
| ○ | add step by step | ✕ 0b430f9 |
| ○ | Add GetNodeMediaListImages | 6463ddb |
| ○ | Add GetNodeChecklist | d50306e |
| ○ | add GetNodeMediaListVideos | 051d591 |
| ○ | add GetNodeSupportResourcesDetailPage | 08f91f2 |
| ○ | Add GetNodeBasicLandingPage | b572e36 |
| ○ | Add GetCampaignLandingPages | abc01a1 |
| ○ | update health_care_region_page filter | b3c7541 |
| ○ | add to rest of build | ✕ c388c0a |
| ○ | remove types | d447ce7 |
| ○ | Merge remote-tracking branch 'origin/master' into nps-individual-queries | ✕ ac26f5f |
| ○ | remove the prod flag | ✕ 46d2e30 |

🚀 **va-vfs-bot** temporarily deployed to master/nps-individual-queries on Feb 9  `Inactive`

👁 **ncksllvn** requested review from **mpelzsherman** and **department-of-veterans-affairs/vsa-public-websites-frontend** on Feb 9

✓ **kelsonic** approved these changes on behalf of **department-of-veterans-affairs/vsa-public-websites-frontend** on Feb 9

[ View changes ]

**kelsonic** left a comment • edited ▾     `Member`

LGTM! 🎉 Thank you SO much for doing this!! 🙏 Will this affect how we write our graphql fragments in the future?

🚀 1

**ncksllvn** commented on Feb 9     `Member`  `Author`

**@kelsonic** it will but there is already a pattern set up so that we can follow pretty easily. `src/site/stages/build/drupal/graphql/faqMultipleQa.graphql.js` is a good example of a simple query that also includes some shared paragraph fragments

🚀 1

🚀 **va-vfs-bot** temporarily deployed to master/nps-individual-queries on Feb 9  `Inactive`

**mpelzsherman** reviewed on Feb 9

**View changes**

`src/site/stages/build/drupal/graphql/paragraph-fragments/staffProfile.paragraph.graphql.js`

| ... | ... | @@ -35,7 +35,7 @@ module.exports = ` |
|-----|-----|---|
| 35 | 35 | alt |
| 36 | 36 | title |
| 37 | 37 | url |
| 38 |  | −         derivative(style: _1 |
|  | 38 | +         derivative(style: _1 |

**mpelzsherman** on Feb 9 · Member

Curious why this change was needed?

**ncksllvn** on Feb 9 · Author · Member

Using `_1_1_SQUARE_MEDIUM_THUMBNAIL` was reported as a GraphQL error and was relying on `src/site/utilities/stringHelpers.js` to turn it into `_11SQUAREMEDIUMTHUMBNAIL`. Instead of using the `stringHelper`, I just changed the value, because I can't find any context for why it was like that. Asking the CMS folks if they have any idea,
https://dsva.slack.com/archives/CDHBKAL9W/p1612907370100500

**ncksllvn** on Feb 9 · Author · Member

Found it...it's safe to just make this change. The string-helpers that did the replacement before is just leftover complexity from a CMS feature flag, `7b54b49` #diff-78d2ba7081dd26ac2a8a81b3bc6c0c80461f62412580c370dff8564362e33ee4R71.

**ncksllvn** on Feb 11 · Author · Member

I wrote a ticket to remove this old code, department-of-veterans-affairs/va.gov-team#19722

Reply...

Resolve conversation

**mpelzsherman** approved these changes on Feb 9

View changes

**brangi** self-requested a review on Feb 9

**brangi** approved these changes on Feb 9

View changes

**ncksllvn** added 3 commits on Feb 10

Issue a count query to generate
pagination dynamically                                    a1cdc87

start index at 1                                          08a8676

fix some bugs with pagination                        ✕ 24674e9

**va-vfs-bot** temporarily deployed to master/nps-
individual-queries on Feb 10  Inactive

**ncksllvn** changed the title ~~[GraphQL] Experiment with
using multiple node queries~~ [GraphQL] Break apart
monolithic GraphQL query into a pattern of small
GraphQL queries on Feb 10

**ncksllvn** changed the title ~~[GraphQL] Break apart
monolithic GraphQL query into a pattern of small
GraphQL queries~~ [GraphQL] Break apart monolithic
query into a pattern of small queries on Feb 10

**ncksllvn** changed the title ~~[GraphQL] Break apart monolithic query into a pattern of small queries~~ **[GraphQL] Convert monolithic query into a pattern of small queries** on Feb 10

**kelsonic** commented on Feb 11                    Member

@department-of-veterans-affairs/frontend-review-group could we get a PR review when you have time please? 🙂

👀 1

✓ **theodur** approved these changes on behalf of **department-of-veterans-affairs/frontend-review-group** on Feb 11

[View changes]

**theodur** left a comment                    Member

LGTM

**ncksllvn** added 2 commits on Feb 11

Merge remote-tracking branch 'origin/master' into nps-individual-queries                    ✗ e8a0211

strengthen the megamenu safety check                    ✗ f713f36

**ncksllvn** reviewed on Feb 11                    [View changes]

```
src/site/stages/build/drupal/menus.js
```

269	269		// eslint-disable-next-line no-co
270		−	console.warn('formatHeaderData ha
271		−	return null;
	270	+	throw new Error('formatHeaderData

ncksllvn on Feb 11 [Author] [Member]

I made this check stronger because I just can't imagine a case where it is acceptable for a build to report success when the MegaMenu is actually failing to build.

Reply...

Resolve conversation

va-vfs-bot deployed to **master/nps-individual-queries** on Feb 11

View deployment

ncksllvn merged commit **ef0f068** into `master` on Feb 11
4 of 5 checks passed

View details   Revert

ncksllvn deleted the `nps-individual-queries` branch on Feb 11

Restore branch

ncksllvn added a commit that referenced this pull request on Feb 11

Revert "[GraphQL] Convert monolithic   ✕ 7939583 query into a pattern of small qu… ···

ncksllvn mentioned this pull request on Feb 11

Revert "[GraphQL] Convert monolithic query into a pattern of small qu...   ⟨↕⟩ Closed

#16027

☑ 0 of 4 tasks complete ▬▬▬▬▬▬

k80bowman mentioned this pull request 25 days ago

Discovery on Improving CMS Export Build Times department-of-veterans-affairs/va.gov-team#19612   ⟨⊙⟩ Closed