RK-J - Task #1722

Split content kind into content and contentSupplement2 in order to speed up getDocumentAPI

10/08/2021 12:02 PM - Ram Kordale

Status:	Closed	**Start date:**	10/11/2021
Priority:	Normal	**Due date:**	10/13/2021
Assignee:	Ram Kordale	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	1-RK-J-1	**Spent time:**	0.00 hour

Description

Made around 30 calls to getDocumentAPI for a short (2 line) and long (about 70 lines) document and got the following results:

	Total calls	Average	Max	60th	80th	95th	98th
longer	30	58.16666667	151	60	70	151	151
shorter	35	47.69444444	144	45	55	128	144

This gives us sufficient data to ensure that splitting content such that unnecessary fields such as cnh are not retrieved during getDocumentAPI call should make it more efficient.

Going forward, content will only have the following fields:
allSectionNames
allSectionNamesAndUrls
baseUrl
br2ir2
breadcrumbStructuredData
data
documentId
documentProcessingId
docViewCount
heading
indexName
lUrl
nextDocument
nextSectionName
oUrl
previousDocument
prevSectionName
subSections

And, contentSupplement2 will contain the following fields:
Idvalues
url
updatedBy
updatedAt
createdBy
createdAt
checksum
links
cnh
s
book
status
rank

JFYI that we were not able to use projections because some of the required fields are >1500 bytes and this causes compositeIndex based projection operations to be not useful.

More notes:
- projection query is not suitable for getDocument Api since allSectionNames and allSectionNamesAndUrls are multi-valued fields . according to the documentation using multi-valued fields in a projection will return a separate entity for each combination.

https://cloud.google.com/datastore/docs/concepts/queries#:~:text=Projecting%20a%20property%20with%20array%20values%20will%20not%20populate%20all%20values%20for%20that%20property.%20Instead%2C%20a%20separate%20entity%20will%20be%20

returned%20for%20each%20unique%20combination%20of%20projected%20values%20matching%20the%20query.

- some fields with data if size > 1500 bytes are excluded from indexing when we save it to datastore. such fields cannot be used in composite indexes used for projection queries according to documentation
--Reference:
https://cloud.google.com/datastore/docs/concepts/indexes#:~:text=Composite%20indexes%20are%20composed%20of%20multiple%20properties%20and%20require%20that%20each%20individual%20property%20must%20not%20be%20excluded%20from%20your%20indexes.
--Reference: https://cloud.google.com/datastore/docs/concepts/storage-size#composite_indexes

History

#1 - 10/20/2021 05:22 AM - Ram Kordale

- Assignee set to Saitej Varri

#2 - 10/20/2021 05:29 AM - Saitej Varri

- Status changed from New to In Progress

#3 - 10/20/2021 05:29 AM - Saitej Varri

- Status changed from In Progress to Resolved

#4 - 10/31/2022 05:41 AM - Ayush Khandelwal

- Status changed from Resolved to Feedback

- Assignee changed from Saitej Varri to Ram Kordale

#5 - 03/29/2023 09:31 AM - Ram Kordale

- Status changed from Feedback to Closed