Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 44 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 44
Topic #: 1
[All Professional Data Engineer Questions]

You are deploying a new storage system for your mobile application, which is a media streaming service. You decide the best fit is Google Cloud Datastore. You have entities with multiple properties, some of which can take on multiple values. For example, in the entity 'Movie' the property 'actors' and the property
'tags' have multiple values but the property 'date released' does not. A typical query would ask for all movies with actor=<actorname> ordered by date_released or all movies with tag=Comedy ordered by date_released. How should you avoid a combinatorial explosion in the number of indexes?

A. Manually configure the index in your index config as follows:

```
Indexes:
-kind: Movie
    Properties:
    -name: actors
    name: date_released
-kind: Movie
    Properties:
    -name: tags
    name: date_released
```

B. Manually configure the index in your index config as follows:

```
Indexes:
    -kind: Movie
        Properties:
        -name: actors
        -name: tags
```

-name: date_published

C. Set the following in your entity options: exclude_from_indexes = 'actors, tags'

D. Set the following in your entity options: exclude_from_indexes = 'date_published'

Show Suggested Answer

by AWSandeep at *Sept. 3, 2022, 1:26 p.m.*

Comments

Type your comment...

Submit

Wasss123 Highly Voted 1 year, 1 month ago

Selected Answer: A

Correct answer is A
Read in reference : https://cloud.google.com/datastore/docs/concepts/indexes#index_limits
n this case, you can circumvent the exploding index by manually configuring an index in your index configuration file:
indexes:
- kind: Task
properties:
- name: tags
- name: created
- kind: Task
properties:
- name: collaborators
- name: created
This reduces the number of entries needed to only (|tags| * |created| + |collaborators| * |created|), or 6 entries instead of 9

upvoted 7 times

jkhong Most Recent 10 months, 2 weeks ago

Selected Answer: A

you can circumvent the exploding index by manually configuring an index in your index configuration file:

https://cloud.google.com/datastore/docs/concepts/indexes#index_limits

upvoted 1 times

Krish6488 10 months, 2 weeks ago

Selected Answer: D

Tempted to go with D as the syntax in Option A seems incorrect. D is still a possible answer because one of the ways to get rid of index errors is to remove the entities that are causing the index to explode. In this case its date_released and hence D appears right to me

upvoted 3 times

DGames 10 months, 3 weeks ago

Selected Answer: A

Option B & D reject because mention date_publised in question date_released is column
Option C also not correct, I would go with option A.

upvoted 3 times

Ender_H 1 year, 1 month ago

Selected Answer: D

Correct Answer D:

This is the way the DB is typically queried:
- movies with actor=<actorname> ordered by date_released
- movies with tag=Comedy ordered by date_released

so it seems that we need indices in actor,tag and date_released for sorting.

A: this would be the correct answer, however, the format is incorrect, the correct format would be '-

name: date_released' correctly indented.

B: This seems to be unnecessary, since typically actor and tag are not queried together. also, there is a clear indentation issue

C: We don't want to ignore actor and tag, we need those indices.

D: If we leave datastore to automatically create the indices and if we specify that the 'date_released' property needs to be excluded from indices, then we would have less indices (but maybe slower queries when ordering them, but hey, how many 'comedies' there could be in the world)

upvoted 3 times

Ender_H 1 year, 1 month ago

And here is the correct way to configure indices:
https://cloud.google.com/datastore/docs/tools/indexconfig

so this would be the best answer:
indexes:
- kind: Movie
properties:
- name: actors
- name: date_released
direction: asc. <This could be left out, it defaults to direction: asc if excluded>

- kind: Movie
properties:
- name: tag
- name: date_released
direction: asc. <This could be left out, it defaults to direction: asc if excluded>

upvoted 3 times

Ender_H 1 year, 1 month ago

Findings for this answer:
Indices, if not defined, will be automatically created:
"By default, a Datastore mode database automatically predefines an index for each property of each entity kind. These single property indexes are suitable for simple types of queries."
source: https://cloud.google.com/datastore/docs/concepts/indexes

In the index limits section we see this:
"a Datastore mode database creates an entry in a predefined index for every property of every entity except those you have explicitly declared as excluded from your indexes."
source: https://cloud.google.com/datastore/docs/concepts/indexes#index_limits

upvoted 1 times

Hm92730 1 year, 1 month ago

What do people think about C? The question is asking how to avoid a combinatorial explosion in the number of indexes. It says "You have entities with multiple properties, some of which can take on multiple values". Put this with the below text from the documentation for Datastore indexes, it seems they're looking for "exclude the properties that will cause combinatorial explosion" which would be C.

"The situation becomes worse in the case of entities with multiple properties, each of which can take on multiple values. To accommodate such an entity, the index must include an entry for every possible combination of property values. Custom indexes that refer to multiple properties, each with multiple values, can "explode" combinatorially, requiring large numbers of entries for an entity with only a relatively small number of possible property values."[1]
[1] https://cloud.google.com/datastore/docs/concepts/indexes#index_limits

upvoted 1 times

soichirokawa 1 year, 2 months ago

B. is correct
To avoid combinatoric explosion of indexes.
"Two queries of the same form but with different filter values use the same index."
https://cloud.google.com/datastore/docs/concepts/indexes

upvoted 1 times

Wasss123 1 year, 1 month ago

Correct answer is A
In the same reference you provided
In this case, you can circumvent the exploding index by manually configuring an index in your index configuration file:
indexes:
- kind: Task
properties:
- name: tags

- name: created
- kind: Task
properties:
- name: collaborators
- name: created
This reduces the number of entries needed to only (|tags| * |created| + |collaborators| * |created|), or 6 entries instead of 9

upvoted 1 times

AWSandeep 1 year, 2 months ago

Selected Answer: A

A. Manually configure the index in your index config as follows:

upvoted 1 times