Google Discussions

Exam Professional Data Engineer All Questions
View all questions & answers for the Professional Data Engineer exam

Go to Exam

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 96 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 96
Topic #: 1
[All Professional Data Engineer Questions]

You want to analyze hundreds of thousands of social media posts daily at the lowest cost and with the fewest steps.
You have the following requirements:
- You will batch-load the posts once per day and run them through the Cloud Natural Language API.
- You will extract topics and sentiment from the posts.
- You must store the raw posts for archiving and reprocessing.
- You will create dashboards to be shared with people both inside and outside your organization.

You need to store both the data extracted from the API to perform analysis as well as the raw social media posts for historical archiving. What should you do?

A. Store the social media posts and the data extracted from the API in BigQuery.

B. Store the social media posts and the data extracted from the API in Cloud SQL.

C. Store the raw social media posts in Cloud Storage, and write the data extracted from the API into BigQuery.

D. Feed to social media posts into the API directly from the source, and write the extracted data from the API into BigQuery.

Show Suggested Answer

by rickywck at *March 17, 2020, 9:51 a.m.*

Comments

Type your comment...

Submit

[Removed] Highly Voted 3 years, 7 months ago

Answer: C
Description: Social media posts can images/videos which cannot be stored in bigquery

upvoted 46 times

- **Shawvin** 2 years ago
  Yes, the raw data needs to be archived too
  upvoted 1 times
- **Devx198912233** 3 years, 3 months ago
  but the posts are fed into cloud natural language api. which means we have to consider the posts to be text only
  upvoted 4 times
- **asksathvik** 2 years, 2 months ago
  Also to run batch queries data needs to be in Cloud Storage, so why not just store it there?
  upvoted 1 times

psu Highly Voted 3 years, 6 months ago

Answer should be C, becose they ask you to save a copy of the raw posts for archival, which may not be possible if you directly feed the posts to the API.

upvoted 17 times

itz_me_sudhir Most Recent 8 months ago

can any one help me with the rest of question from 101 to 209 as i dont have a contributor access

upvoted 2 times

zellck 11 months ago

Selected Answer: C

C is the answer.

upvoted 2 times

sedado77 1 year, 1 month ago

Selected Answer: C

I got this question on sept 2022. Answer is C

upvoted 5 times

Erso 1 year, 1 month ago

Selected Answer: C

C is the correct one

upvoted 1 times

medeis_jar 1 year, 10 months ago

Selected Answer: C

Only C make sense.

upvoted 2 times

MaxNRG 1 year, 10 months ago

Selected Answer: C

You must store the raw posts for archiving and reprocessing, Store the raw social media posts in Cloud Storage.
B is expensive
D is not valid since you have to store the raw posts for archiving
Between A and C I's say C, since we're going to make dashboards and Data Studio will connect well with big query. and besides A would probably be more expensive.

upvoted 3 times

BigQuery 1 year, 11 months ago

SAY MY NAME!

upvoted 4 times

StefanoG 1 year, 11 months ago

Selected Answer: C

Analysis BQ
Storage GCS

upvoted 2 times

fire558787 2 years, 2 months ago

I believe the API accesses data only from GCS Buckets not BigQuery (but I'm not entirely sure)

upvoted 1 times

sumanshu 2 years, 4 months ago

Vote for C

upvoted 2 times

DPonly 2 years, 9 months ago

Answer should be C because we need to consider storage archival

upvoted 2 times

arghya13 2 years, 11 months ago

I'll go with option C

upvoted 2 times

Alasmindas 2 years, 12 months ago

I will go with Option C, because of the following reasons:-
a) Social media posts are "raw" - which means - it can be of any format (blob/object storage) is preferred.
b) The output from the application (assuming the application is Cloud NLP) is to be future stored for archival purpose - and hence again Google Cloud storage is the best option - so option C
Option A &C - Incorrect, although Option D fulfils the requirement of "fewest step" but storing data in big query for archival purpose is not a google recommended approach
Option B : Cloud SQL rules out as it does not solve either for archival storage or for analytics purpose.

upvoted 3 times

singhkrishna 3 years, 1 month ago

cost of long term storing is almost same in GCS and BQ, so answer D makes sense from that angle..

upvoted 1 times

Tanmoyk 3 years, 1 month ago

The job is supposed to run in batch process once in a day , so there is no requirement of stream data. The best economical and less complex steps is answer C

upvoted 2 times

Load full discussion...

EXAMTOPICS

Platform

- Home
- All Exams
- Examtopics PRO
- Training Courses

© 2024 ExamTopics