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 239 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 239
Topic #: 1
[All Professional Data Engineer Questions]

The data analyst team at your company uses BigQuery for ad-hoc queries and scheduled SQL pipelines in a Google Cloud project with a slot reservation of 2000 slots. However, with the recent introduction of hundreds of new non time-sensitive SQL pipelines, the team is encountering frequent quota errors. You examine the logs and notice that approximately 1500 queries are being triggered concurrently during peak time. You need to resolve the concurrency issue. What should you do?

A. Increase the slot capacity of the project with baseline as 0 and maximum reservation size as 3000.

B. Update SQL pipelines to run as a batch query, and run ad-hoc queries as interactive query jobs.

C. Increase the slot capacity of the project with baseline as 2000 and maximum reservation size as 3000.

D. Update SQL pipelines and ad-hoc queries to run as interactive query jobs.

Show Suggested Answer

by scaenruy at *Jan. 3, 2024, 1:55 p.m.*

Comments

Type your comment...

Submit

raaad Highly Voted 1 year, 4 months ago

Selected Answer: B

- BigQuery allows you to specify job priority as either BATCH or INTERACTIVE.
- Batch queries are queued and then started when idle resources are available, making them suitable for non-time-sensitive workloads.
- Running ad-hoc queries as interactive ensures they have prompt access to resources.

upvoted 9 times

LP_PDE Most Recent 3 months, 1 week ago

Selected Answer: B

By updating your SQL pipelines to run as batch queries you can reduce concurrency, avoid quota errors, and ensure that your analysts have the resources they need for their interactive queries.

upvoted 1 times

ToiToi 6 months ago

Selected Answer: B

This question has nothing to do with increasing slots, it is just confusing and misleading, therefore A and C do not make sense.
D (All interactive queries): Running all queries as interactive would prioritize speed over cost-efficiency and might not be necessary for your non-time-sensitive SQL pipelines.

upvoted 2 times

josech 11 months, 3 weeks ago

Selected Answer: C

You already have a 2000 slots consumption and sudden peaks, so you should use a baseline of 2000 slots and a maximum of 3000 to tackle the peak concurrent activity.
https://cloud.google.com/bigquery/docs/slots-autoscaling-intro

upvoted 3 times

CGS22 1 year ago

Selected Answer: A

Why A is the best choice:

Addresses Concurrency: Increasing the maximum reservation size to 3000 slots directly addresses the concurrency issue by providing more capacity for simultaneous queries. Since the current peak usage is 1500 queries, this increase ensures sufficient headroom.
Cost Optimization: Setting the baseline to 0 means you only pay for the slots actually used, avoiding unnecessary costs for idle capacity. This is ideal for non-time-sensitive workloads where flexibility is more important than guaranteed instant availability.

upvoted 2 times

JyoGCP 1 year, 2 months ago

Selected Answer: B

Option B

upvoted 2 times

CGS22 1 year ago

B: While batch queries are generally more cost-effective for large, non-interactive workloads, they don't solve the concurrency problem. If multiple batch queries are triggered simultaneously, they would still compete for the same limited slot pool.

upvoted 1 times

scaenruy 1 year, 4 months ago

Selected Answer: B

B.
Update SQL pipelines to run as a batch query, and run ad-hoc queries as interactive query jobs.

upvoted 2 times

© 2024 ExamTopics