Google Discussions

EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 95 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 95
Topic #: 1
[All Professional Data Engineer Questions]

You have a data pipeline that writes data to Cloud Bigtable using well-designed row keys. You want to monitor your pipeline to determine when to increase the size of your Cloud Bigtable cluster. Which two actions can you take to accomplish this? (Choose two.)

A. Review Key Visualizer metrics. Increase the size of the Cloud Bigtable cluster when the Read pressure index is above 100.

B. Review Key Visualizer metrics. Increase the size of the Cloud Bigtable cluster when the Write pressure index is above 100.

C. Monitor the latency of write operations. Increase the size of the Cloud Bigtable cluster when there is a sustained increase in write latency.

D. Monitor storage utilization. Increase the size of the Cloud Bigtable cluster when utilization increases above 70% of max capacity.

E. Monitor latency of read operations. Increase the size of the Cloud Bigtable cluster of read operations take longer than 100 ms.

Show Suggested Answer

by rickywck at *March 17, 2020, 9:48 a.m.*

Comments

Type your comment...

Submit

jvg637 Highly Voted 4 years, 7 months ago

Answer is C & D.
C –> Adding more nodes to a cluster (not replication) can improve the write performance
https://cloud.google.com/bigtable/docs/performance
D –> since Google recommends adding nodes when storage utilization is > 70%
https://cloud.google.com/bigtable/docs/modifying-instance#nodes

upvoted 53 times

> **dabrat** 4 years ago
>
> Storage utilization (% max)
> The percentage of the cluster's storage capacity that is being used. The capacity is based on the number of nodes in your cluster.
>
> In general, do not use more than 70% of the hard limit on total storage, so you have room to add more data. If you do not plan to add significant amounts of data to your instance, you can use up to 100% of the hard limit.
>
> Important: If any cluster in an instance exceeds the hard limit on the amount of storage per node, writes to all clusters in that instance will fail until you add nodes to each cluster that is over the limit. Also, if you try to remove nodes from a cluster, and the change would cause the cluster to exceed the hard limit on storage, Cloud Bigtable will deny the request. If you are using more than the recommended percentage of the storage limit, add nodes to the cluster. You can also delete existing data, but deleted data takes up more space, not less, until a compaction occurs.
>
> upvoted 4 times
>
> > **dabrat** 4 years ago
> >
> > https://cloud.google.com/bigtable/docs/monitoring-instance
> >
> > upvoted 3 times
>
> **sergio6** 3 years, 1 month ago
>
> Adding nodes to the cluster In Bigtable scales linearly the performances both read and write
> https://cloud.google.com/bigtable/docs/performance#typical-workloads
>
> upvoted 1 times

Barniyah Highly Voted 4 years, 6 months ago

Key visualizer is bigtable metric , So A and B incorrect
storage utilization also bigtable metric , So D incorrect
The question want you to monitor pipeline metrics (which is dataflow metrics) , in our case we can only monitor latency .
The answer will be : C & E

upvoted 10 times

> **ch3n6** 4 years, 4 months ago
>
> No. it is C, D. "You have a data pipeline that writes data to Cloud Bigtable using well-designed row keys."
> why are you monitoring read anyway? you are just writing.
>
> upvoted 14 times

Parandhaman_Margan Most Recent 1 month, 3 weeks ago

Selected Answer: BC

Correct answers are **B** (Write pressure) and **C** (latency)

upvoted 1 times

TVH_Data_Engineer 5 months, 2 weeks ago

Selected Answer: BC

The question focus is on writing. BC is correct. when the writing pressure is above 100, it is time to increase. same logic with C

upvoted 1 times

musumusu 1 year, 8 months ago

why not B ?

upvoted 1 times

> **musumusu** 1 year, 8 months ago
>
> i am feeling to go with B and D. In option C, when latency is low, latency can be low for write operation for other reason. but in option B, its showing clearly when write pressure more than 100. But why no one is talking about B
>
> upvoted 2 times

RoshanAshraf 1 year, 9 months ago

Selected Answer: CD

Key visualizer is Metrics for Performance issues. Ruled out
Storage and Write Operations ; C and D

upvoted 3 times

zellck 1 year, 11 months ago

Selected Answer: CD

CD is the answer.

https://cloud.google.com/bigtable/docs/monitoring-instance#disk
Storage utilization (% max)
- The percentage of the cluster's storage capacity that is being used. The capacity is based on the number of nodes in your cluster.
In general, do not use more than 70% of the hard limit on total storage, so you have room to add more data.

upvoted 3 times

John_Pongthorn 2 years, 1 month ago

Selected Answer: CD

Well-designed row key : A B are not nessary
Write : CD both are involved in the question the most.

upvoted 2 times

Fezo 2 years, 4 months ago

Answer: CD
https://cloud.google.com/bigtable/docs/scaling

upvoted 2 times

medeis_jar 2 years, 10 months ago

Selected Answer: CD

as explained by MaxNRG

upvoted 2 times

MaxNRG 2 years, 10 months ago

Selected Answer: CD

D: In general, do not use more than 70% of the hard limit on total storage, so you have room to add more data. If you do not plan to add significant amounts of data to your instance, you can use up to 100% of the hard limit
C: If this value is frequently at 100%, you might experience increased latency. Add nodes to the cluster to reduce the disk load percentage.
The key visualizer metrics options, suggest other things other than increase the cluster size.
https://cloud.google.com/bigtable/docs/monitoring-instance

upvoted 3 times

hendrixlives 2 years, 10 months ago

Selected Answer: CD

CD.

I agree with jvg637

upvoted 1 times

StefanoG 2 years, 11 months ago

Selected Answer: AD

from https://cloud.google.com/bigtable/docs/monitoring-instance
Disk load - If this value is frequently at 100%, you might experience increased latency. Add nodes to the cluster to reduce the disk load percentage.
Storage utilization (% max) - In general, do not use more than 70% of the hard limit on total storage, so you have room to add more data. If you do not plan to add significant amounts of data to your instance, you can use up to 100% of the hard limit.

upvoted 2 times

KokkiKumar 2 years, 11 months ago

I am Voting for CD

upvoted 2 times

u_t_s 3 years ago

Answer should be D & E

upvoted 1 times

tavva_prudhvi 2 years, 7 months ago

Why are you monitoring read operations, when youre supposed to write? why E?

upvoted 1 times

sergio6 3 years, 1 month ago

D--> 70% is the the recommended percentage of the cluster's storage capacity that is being used, If you are using more than 70% storage limit, add nodes to the cluster
https://cloud.google.com/bigtable/quotas#storage-per-node
https://cloud.google.com/bigtable/docs/monitoring-instance#disk
E--> 100 ms is an order of magnitude lower latency than Google claimed (<10ms)
https://cloud.google.com/bigtable/docs/performance#typical-workloads

upvoted 2 times

hauhau 3 years, 2 months ago

BC
D:you can just add node, not cluster
The percentage of the cluster's storage capacity that is being used. The capacity is based on the number of nodes in your cluster.(https://cloud.google.com/bigtable/docs/monitoring-instance)
After you create a Cloud Bigtable instance, you can update any of the following settings without any downtime:

(The number of nodes in each cluster)
https://cloud.google.com/bigtable/docs/modifying-instance

upvoted 1 times

Load full discussion...