

Google Discussions



Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 126 DISCUSSI...

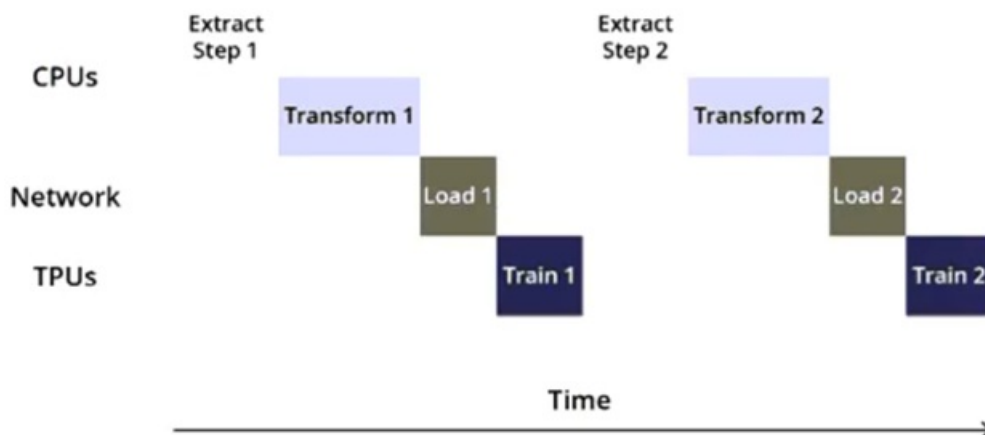
Actual exam question from Google's Professional Machine Learning Engineer

Question #: 126

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You are training an object detection model using a Cloud TPU v2. Training time is taking longer than expected. Based on this simplified trace obtained with a Cloud TPU profile, what action should you take to decrease training time in a cost-efficient way?



- A. Move from Cloud TPU v2 to Cloud TPU v3 and increase batch size.
- B. Move from Cloud TPU v2 to 8 NVIDIA V100 GPUs and increase batch size.
- C. Rewrite your input function to resize and reshape the input images.
- D. Rewrite your input function using parallel reads, parallel processing, and prefetch.



Show Suggested Answer

by  mymy9418 at Dec. 18, 2022, 3:20 a.m.

Comments

Type your comment...

Submit

  **pshemol** Highly Voted 1 year, 10 months ago

Selected Answer: D

parallel reads, parallel processing, and prefetch is needed here



   upvoted 6 times

  **fitri001** Most Recent 6 months, 2 weeks ago

Selected Answer: D

Optimizing the data pipeline with parallel reads, processing, and prefetching can significantly improve training speed on TPUs by reducing I/O wait times. This approach utilizes the TPU's capabilities more effectively and avoids extra costs associated with hardware upgrades.



   upvoted 3 times

  **fitri001** 6 months, 2 weeks ago

A. Moving to a different TPU version (v3) and increasing the batch size might improve training speed, but it's an expensive solution without a guarantee of the most efficient outcome.

B. Switching to GPUs (V100) also increases costs and may not be optimized for your specific workload.

   upvoted 1 times

  **fitri001** 6 months, 2 weeks ago

(C) can be part of the preprocessing step, but it likely won't address the core issue if the bottleneck is related to how data is being fed into the training process.

   upvoted 1 times

  **M25** 1 year, 6 months ago

Selected Answer: D

Went with D



   upvoted 1 times

  **TNT87** 1 year, 8 months ago

Selected Answer: D

Based on the profile, it appears that the Compute time is relatively low compared to the HostToDevice and DeviceToHost time. This suggests that the data transfer between the host (CPU) and the TPU device is a bottleneck. Therefore, the best action to decrease training time in a cost-efficient way would be to reduce the amount of data transferred between the host and the device.

   upvoted 1 times

  **hiromi** 1 year, 10 months ago

Selected Answer: D

D

- https://www.tensorflow.org/guide/data_performance

   upvoted 4 times

  **mymy9418** 1 year, 10 months ago

Selected Answer: D

i didn't see v3 has any benefit than v2

https://cloud.google.com/tpu/docs/system-architecture-tpu-vm#performance_benefits_of_tpu_v3_over_v2

   upvoted 1 times

Start Learning for free



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)



We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.