■ MENU

C

G 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 22 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 22

Topic #: 1

[All Professional Data Engineer Questions]

Your company has hired a new data scientist who wants to perform complicated analyses across very large datasets stored in Google Cloud Storage and in a

Cassandra cluster on Google Compute Engine. The scientist primarily wants to create labelled data sets for machine learning projects, along with some visualization tasks. She reports that her laptop is not powerful enough to perform her tasks and it is slowing her down. You want to help her perform her tasks.

What should you do?

- A. Run a local version of Jupiter on the laptop.
- B. Grant the user access to Google Cloud Shell.
- C. Host a visualization tool on a VM on Google Compute Engine.
- D. Deploy Google Cloud Datalab to a virtual machine (VM) on Google Compute Engine.

Show Suggested Answer

by A Rajokkiyam at March 15, 2020, 3:36 a.m.

Comments

Type your comment...

Submit

Rajokkiyam Highly Voted 4 years, 7 months ago Answer should be D. upvoted 48 times ☐ ♣ [Removed] Highly Voted 🖈 4 years, 7 months ago Description: Datalab provides Jupyter for this kind of work upvoted 14 times ■ Abizi Most Recent ② 2 months ago Selected Answer: D D is the right answer upvoted 1 times VictorBa 5 months, 1 week ago Selected Answer: D Google Cloud Datalab is a powerful interactive tool for data exploration, analysis, and machine learning. upvoted 1 times 🗖 🏜 trashbox 5 months, 3 weeks ago Selected Answer: D My answer is Google Cloud Datalab, but since that service has already been discontinued, I question whether a problem like this would actually be asked on the actual exam. upvoted 4 times GCanteiro 9 months, 3 weeks ago Selected Answer: D D sounds good for me upvoted 1 times TVH_Data_Engineer 10 months, 3 weeks ago Selected Answer: A Hash Value for Deduplication: By computing a hash value for each data entry, you create a unique identifier based on the content of the data. This allows you to efficiently identify duplicates, as entries with identical content will have the same hash value. Storing Hash Value and Metadata: Maintaining a database table that includes the hash value and other relevant metadata (like the timestamp of transmission) allows for quick lookups and comparisons. This way, when new data is received, you can check if an entry with the same hash value already exists. Assign global unique identifiers (GUID) to each data entry: While GUIDs are unique, they do not inherently identify duplicate content. Two transmissions of the same data would have different GUIDs. upvoted 1 times 🖃 🏜 simpa17 2 months, 2 weeks ago You mistakenly answered the question above haha upvoted 1 times axantroff 11 months, 2 weeks ago **Selected Answer: D** D sounds good for me upvoted 1 times □ ♣ RT_G 12 months ago Selected Answer: D Agree with D upvoted 1 times

Selected Answer: D

🖃 🏜 rtcpost 1 year ago

D. Deploy Google Cloud Datalab to a virtual machine (VM) on Google Compute Engine.

Google Cloud Datalab is a powerful interactive tool for data exploration, analysis, and machine learning. By deploying it to a VM on Google Compute Engine, you can provide her with a robust and scalable environment where she can work with large datasets, create labeled datasets, and perform data analyses efficiently.

Option A (running a local version of Jupyter on her laptop) might not be sufficient for very large datasets, and her laptop's limited power could still be a bottleneck.

Option B (granting access to Google Cloud Shell) is useful for running command-line tools but may not provide the interactive and visual capabilities she needs.

Option C (hosting a visualization tool on a VM on Google Compute Engine) is beneficial for visualization tasks but does not cover the full spectrum of data analysis and machine learning tasks that Google Cloud Datalab offers.

upvoted 3 times

🗖 🚨 qudquy1a 1 year, 2 months ago

Selected Answer: D

D - as it is a FULL set up, not a shell that is needed...

upvoted 1 times

😑 🏜 sergiomujica 1 year, 2 months ago

Nowadays it should be similar to D, deploy a Vertex workbench

upvoted 2 times

ago 🖹 🏜 yash12 1 year, 2 months ago

As per Options , Correct Answer should be D. ie Datalab

However Datalab is no longer used in GCP (Deprecated in Sep2022), It is Vertex AI or Deep Learning VM Images

upvoted 1 times

🖃 🏜 HeoMaTo 1 year, 2 months ago

Selected Answer: D

I think.

Answer is D

upvoted 1 times

🖃 🚨 Acocado 1 year, 3 months ago

Datalab is deprecated. This question should appear in the exam.

upvoted 2 times

■ Acocado 1 year, 3 months ago

typo- should NOT appear in the exam

upvoted 6 times

😑 🏜 axantroff 1 year ago

Good point - https://cloud.google.com/datalab/deprecation-notice. Google recommends using Vertex AI Workbench instead

upvoted 1 times

dgteixeira 1 year, 4 months ago

Selected Answer: D

Should be D, because Cloud shell alone does not provide access to what they need.

Nowadays is Vertex AI, but still, correct answer is D

upvoted 3 times

■ Maurilio_Cardoso 1 year, 5 months ago

Selected Answer: D

Google Cloud Datalab is now Vertex AI. So, letter D make more sense.

upvoted 3 times

Load full discussion...



