EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 146 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer
Question #: 146
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work on the data science team at a manufacturing company. You are reviewing the company's historical sales data, which has hundreds of millions of records. For your exploratory data analysis, you need to calculate descriptive statistics such as mean, median, and mode; conduct complex statistical tests for hypothesis testing; and plot variations of the features over time. You want to use as much of the sales data as possible in your analyses while minimizing computational resources. What should you do?

A. Visualize the time plots in Google Data Studio. Import the dataset into Vertex AI Workbench user-managed notebooks. Use this data to calculate the descriptive statistics and run the statistical analyses.

B. Spin up a Vertex AI Workbench user-managed notebooks instance and import the dataset. Use this data to create statistical and visual analyses.

C. Use BigQuery to calculate the descriptive statistics. Use Vertex AI Workbench user-managed notebooks to visualize the time plots and run the statistical analyses.

D. Use BigQuery to calculate the descriptive statistics, and use Google Data Studio to visualize the time plots. Use Vertex AI Workbench user-managed notebooks to run the statistical analyses.

Show Suggested Answer

by imamapri at *Feb. 4, 2023, 5:06 a.m.*

Comments

Type your comment...

Submit

gscharly 6 months, 3 weeks ago

Selected Answer: C

went with c

upvoted 1 times

Aastha_Vashist 7 months, 2 weeks ago

Selected Answer: C

went with c

upvoted 2 times

pico 11 months, 4 weeks ago

Selected Answer: C

Option D is not as efficient because using Google Data Studio for time plots may not be as well-suited for handling large datasets, and it's more focused on data visualization. Option A involves importing data into Vertex AI Workbench first, which may not be the most efficient way to leverage BigQuery for handling large-scale data computations.

upvoted 3 times

andresvelasco 1 year, 1 month ago

Selected Answer: D

I would go with D, thinking that Bigquery + Datastudio avoid having to load 100s of MILLIONS of records in memory for the most basic tasks, as required by the Notebook.

upvoted 3 times

friedi 1 year, 4 months ago

Selected Answer: D

D minimizes resources the most, since it minimizes the usage of Vertex AI notebooks, which basically require provisioning a VM in the background for the entire duration of development.

upvoted 3 times

bechir141bf 1 year, 4 months ago

Why not D ?

upvoted 4 times

tavva_prudhvi 1 year, 4 months ago

Using BigQuery to calculate the descriptive statistics is a good choice, but using Google Data Studio for visualizations may not be as flexible as using Vertex AI Workbench user-managed notebooks. Google Data Studio is a great tool for creating dashboards and reports, but it may not allow for the level of customization that is required for detailed exploratory data analysis.

upvoted 1 times

M25 1 year, 5 months ago

Selected Answer: C

https://cloud.google.com/architecture/data-science-with-r-on-gcp-eda#ai_platform_notebooks
https://cloud.google.com/vertex-ai-workbench#section-5

upvoted 1 times

[Removed] 1 year, 6 months ago

Selected Answer: C

I think the key here is that it says the dataset would be imported into the notebook for B, therefore no longer utilising BigQuery for calculating the descriptive stats, otherwise I would pick B. Therefore I think C is better. Can anyone find any documentation where Google gives best practice on this? It seems quite subjective

upvoted 1 times

TNT87 1 year, 8 months ago

Selected Answer: C

C. Use BigQuery to calculate the descriptive statistics. Use Vertex AI Workbench user-managed notebooks to visualize the time plots and run the statistical analyses.

BigQuery is a powerful data analysis tool that can handle massive datasets, making it an ideal solution for calculating descriptive statistics for hundreds of millions of records. It can also perform complex statistical tests for hypothesis testing. For time series analysis, using Vertex AI Workbench user-managed notebooks would be the best solution as it provides a

flexible environment for data exploration, visualization, and statistical analysis. By using the two tools together, the data science team can efficiently analyze the sales data while minimizing computational resources.
Its C not B

upvoted 3 times

Gudwin 1 year, 6 months ago

What is the point of actually giving chatGPT answers, some of which are incorrect?

upvoted 7 times

TNT87 1 year, 8 months ago

Answer B
https://cloud.google.com/vertex-ai-workbench

upvoted 1 times

TNT87 1 year, 8 months ago

Answer C not B

upvoted 1 times

imamapri 1 year, 9 months ago

Selected Answer: B

Vote B. You can do all of the task in vertex AI workbench while minimizing computational resources.

upvoted 3 times

tavva_prudhvi 1 year, 4 months ago

Option B is also a viable solution, but it has some drawbacks compared to option C. While it is true that you can spin up a Vertex AI Workbench user-managed notebooks instance and import the dataset, this option requires more computational resources and may not be as cost-effective as using BigQuery to calculate the descriptive statistics. Additionally, while you can create statistical and visual analyses within the Vertex AI Workbench user-managed notebooks, it may not be as easy to create custom visualizations as it is with Google Data Studio.

Therefore, while option B is a valid solution, option C is likely to be more efficient and cost-effective, as it takes advantage of the strengths of each tool.

upvoted 1 times

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.