⊙ 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 81 DISCUSSION**

Actual exam question from Google's Professional Data Engineer
Question #: 81
Topic #: 1
[All Professional Data Engineer Questions]

MJTelco Case Study -

Company Overview -
MJTelco is a startup that plans to build networks in rapidly growing, underserved markets around the world. The company has patents for innovative optical communications hardware. Based on these patents, they can create many reliable, high-speed backbone links with inexpensive hardware.

Company Background -
Founded by experienced telecom executives, MJTelco uses technologies originally developed to overcome communications challenges in space. Fundamental to their operation, they need to create a distributed data infrastructure that drives real-time analysis and incorporates machine learning to continuously optimize their topologies. Because their hardware is inexpensive, they plan to overdeploy the network allowing them to account for the impact of dynamic regional politics on location availability and cost.
Their management and operations teams are situated all around the globe creating many-to-many relationship between data consumers and provides in their system. After careful consideration, they decided public cloud is the perfect environment to support their needs.

Solution Concept -
MJTelco is running a successful proof-of-concept (PoC) project in its labs. They have two primary needs:
☞ Scale and harden their PoC to support significantly more data flows generated when they ramp to more than 50,000 installations.
Refine their machine-learning cycles to verify and improve the dynamic models they use to control topology definition.
▪

MJTelco will also use three separate operating environments `" development/test, staging, and production `" to meet the needs of running experiments, deploying new features, and serving production customers.

Business Requirements -

☞ Scale up their production environment with minimal cost, instantiating resources when and where needed in an unpredictable, distributed telecom user community.

☞ Ensure security of their proprietary data to protect their leading-edge machine learning and analysis.

☞ Provide reliable and timely access to data for analysis from distributed research workers

☞ Maintain isolated environments that support rapid iteration of their machine-learning models without affecting their customers.

Technical Requirements -

Ensure secure and efficient transport and storage of telemetry data

Rapidly scale instances to support between 10,000 and 100,000 data providers with multiple flows each.

Allow analysis and presentation against data tables tracking up to 2 years of data storing approximately 100m records/day

Support rapid iteration of monitoring infrastructure focused on awareness of data pipeline problems both in telemetry flows and in production learning cycles.

CEO Statement -

Our business model relies on our patents, analytics and dynamic machine learning. Our inexpensive hardware is organized to be highly reliable, which gives us cost advantages. We need to quickly stabilize our large distributed data pipelines to meet our reliability and capacity commitments.

CTO Statement -

Our public cloud services must operate as advertised. We need resources that scale and keep our data secure. We also need environments in which our data scientists can carefully study and quickly adapt our models. Because we rely on automation to process our data, we also need our development and test environments to work as we iterate.

CFO Statement -

The project is too large for us to maintain the hardware and software required for the data and analysis. Also, we cannot afford to staff an operations team to monitor so many data feeds, so we will rely on automation and infrastructure. Google Cloud's machine learning will allow our quantitative researchers to work on our high-value problems instead of problems with our data pipelines.

You need to compose visualization for operations teams with the following requirements:

☞ Telemetry must include data from all 50,000 installations for the most recent 6 weeks (sampling once every minute)

☞ The report must not be more than 3 hours delayed from live data.

☞ The actionable report should only show suboptimal links.

☞ Most suboptimal links should be sorted to the top.

Suboptimal links can be grouped and filtered by regional geography.

▪

☞ User response time to load the report must be <5 seconds.

You create a data source to store the last 6 weeks of data, and create visualizations that allow viewers to see multiple date ranges, distinct geographic regions, and unique installation types. You always show the latest data without any changes to your visualizations. You want to avoid creating and updating new visualizations each month. What should you do?

 A. Look through the current data and compose a series of charts and tables, one for each possible combination of criteria.

 B. Look through the current data and compose a small set of generalized charts and tables bound to criteria filters that allow value selection.

 C. Export the data to a spreadsheet, compose a series of charts and tables, one for each possible combination of criteria, and

spread them across multiple tabs.

D. Load the data into relational database tables, write a Google App Engine application that queries all rows, summarizes the data across each criteria, and then renders results using the Google Charts and visualization API.

Show Suggested Answer

by [deleted] at *March 22, 2020, 6:19 p.m.*

Comments

Type your comment...

Submit

👤 **[Removed]** `Highly Voted 👍` 5 years, 1 month ago

Should be B

👍 ↩ 🚩 upvoted 32 times

👤 **Jarek7** `Highly Voted 👍` 2 years ago

`Selected Answer: D`

First I thought B, as D seems too complex with writing app for AppEngine. But B is too simple - just look through the data doesnt seem right.
It must be very old question. Today you would load the data to BQ, optionally you can use Dataprep for simple data cleaning or a Dataflow job for more complex data processing, and finally use Looker to create tables and charts.

👍 ↩ 🚩 upvoted 7 times

> 👤 **Oleksandr0501** 2 years ago
>
> As per old question - must be. I heard, that the exam will mostly have questions rather from 100 to 205 than form 1 to 100. And smb told me, that the other w/s gave questions, that happened more often in exam, in comparison to questions given here
>
> 👍 ↩ 🚩 upvoted 3 times

👤 **oussama7** `Most Recent ⊙` 1 month, 2 weeks ago

`Selected Answer: B`

Filters allow dynamic interaction: Instead of static charts, filters enable users to select date ranges, regions, and installation types without requiring frequent updates.

👍 ↩ 🚩 upvoted 1 times

👤 **Parandhaman_Margan** 1 month, 3 weeks ago

`Selected Answer: B`

Dynamic Filtering → Instead of creating a fixed set of charts for every combination, filters allow users to explore data interactively without manual updates.
Scalability → Creating a small number of general charts with filters reduces maintenance effort and dashboard complexity.

👍 ↩ 🚩 upvoted 1 times

👤 **Augustax** 3 months, 2 weeks ago

`Selected Answer: B`

Data engineer especially the Front-End developer will pick B.

👍 ↩ 🚩 upvoted 1 times

👤 **cloud_rider** 5 months, 1 week ago

`Selected Answer: B`

D is not the right answer as Chart and Viz API is deprecated now (https://en.wikipedia.org/wiki/Google_Chart_API#:~:text=The%20Google%20Chart%20API%20is,charts%20from%20user%2Ds upplied%20data.)

B is the most logical answer as it talks about creation of general chart wtih filter for value selection (as asked in the requirement),

👍 ↩ 🚩 upvoted 2 times

> 👤 **grshankar9** 3 months, 2 weeks ago
>
> Google Chart API is deprecated but there is a 'Google Charts' API now. And Visualization is part of it

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **Nirca** 1 year, 7 months ago

Selected Answer: B

bound to criteria filters that allow value selection. - Simple and Smart.

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **PolyMoe** 2 years, 3 months ago

Selected Answer: D

D.

everything is fixed except data that is updated regularly in order to keep the last 6 weeks. Then, the pipeline does not change ==> obtaining (same) charts and viz on regularly updated data

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **hauhau** 2 years, 5 months ago

Selected Answer: B

B

But can someone explain the question and selection clearly?

👍 ↩ 🏳 upvoted 4 times

⊟ 👤 **cloudmon** 2 years, 6 months ago

Selected Answer: B

It's B. All the other choices are unreasonable.

👍 ↩ 🏳 upvoted 5 times

⊟ 👤 **edwardlin421** 2 years, 6 months ago

ACD-Design for each possible combination of criteria, so if your team has new requirements, you must design new charts.
So, answer shoud be B.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **ducc** 2 years, 8 months ago

Selected Answer: D

the key is " You want to avoid creating and updating new visualizations each month."
only D work for that phrase

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **wan2three** 2 years, 4 months ago

D you might need to load data from source to table for each month. It stated the source will keep last 6 weeks data, but not in D

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **KundanK973** 2 years, 10 months ago

must be D

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **ealpuche** 2 years, 10 months ago

Selected Answer: D

The answer is B

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **rr4444** 2 years, 10 months ago

This Q feels very disconnected from GCP products.....

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **sw52099** 2 years, 11 months ago

Selected Answer: D

Vote D.
Since B just uses "current data", which means if new data enters, you need to re-run those charts again.

👍 ↩ 🏳 upvoted 4 times

⊟ 👤 **wan2three** 2 years, 4 months ago

But q says the data sources only have latest 6 weeks data, so current data means latest?

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **RRK2021** 3 years, 2 months ago

B is optimal to avoid creating and updating new visualizations each month

👍 ↩ 🏳 upvoted 1 times

Load full discussion...

EXAMTOPICS

Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses