⊙ 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 80 DISCUSSION**

Actual exam question from Google's Professional Data Engineer
Question #: 80
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

...roles 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.
MJTelco is building a custom interface to share data. They have these requirements:
1. They need to do aggregations over their petabyte-scale datasets.
2. They need to scan specific time range rows with a very fast response time (milliseconds).
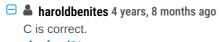Which combination of Google Cloud Platform products should you recommend?

A. Cloud Datastore and Cloud Bigtable

B. Cloud Bigtable and Cloud SQL

C. BigQuery and Cloud Bigtable

D. BigQuery and Cloud Storage

Show Suggested Answer

by [deleted] at *March 21, 2020, 6:19 p.m.*

Comments

Type your comment

Submit

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

correct: C

👍 ↩ 🏳 upvoted 19 times

⊟ 👤 **atnafu2020** `Highly Voted 👍` 4 years, 9 months ago

C
Bigquery and Big table =PB storage capacity
Bigtable=to read scan rows Big query select row to read

👍 ↩ 🏳 upvoted 9 times

⊟ 👤 **09878d5** `Most Recent ⊘` 6 months ago

Why not A? Can someone please explain

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **baht** 1 year, 10 months ago

`Selected Answer: C`

Response C => Bigquery and bigtable

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **ga8our** 2 years ago

Why not A? If we're already using Bigtable, what's the use of another, slower analytic solution, like BigQuery? Wouldn't Datastore be more useful to store our data than BigQuery?

👍 ↩ 🏳 upvoted 5 times

⊟ 👤 **dconesoko** 2 years, 3 months ago

`Selected Answer: C`

bigquery and bigtable

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **sandipk91** 3 years, 9 months ago

C is correct, no doubt

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **sumanshu** 3 years, 10 months ago

Vote for C

👍 ↩ 🏳 upvoted 2 times

⊟ 👤 **daghayeghi** 4 years, 1 month ago

C:
They need to do aggregations over their petabyte-scale datasets: Bigquery
They need to scan specific time range rows with a very fast response time (milliseconds): Bigtable

👍 ↩ 🏳 upvoted 6 times

⊟ 👤 **Insane7** 4 years, 6 months ago

Why not D? Biqquery and GCS.
Also Big Table is no sql, where as BQ is SQL

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **Gcpyspark** 4 years, 4 months ago

With GCS you can only scan the rows from BigQuery using External federated Datasources, with that millisecond latency is not possible. Also "scan specific time range rows with a very fast response time" is a natural fit use case for Cloud Bigtable.

👍 ↩ 🏳 upvoted 5 times

⊟ 👤 **haroldbenites** 4 years, 8 months ago

C is correct.

👍 ↩ 🏳 upvoted 4 times

Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses