Question #39Topic 1

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 visualizations for operations teams with the following requirements:

✑ The report must include telemetry data from all 50,000 installations for the most resent 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.

Which approach meets the requirements?

A. Load the data into Google Sheets, use formulas to calculate a metric, and use filters/sorting to show only suboptimal links in a table.

B. Load the data into Google BigQuery tables, write Google Apps Script that queries the data, calculates the metric, and shows only suboptimal rows in a table in Google Sheets.

C. Load the data into Google Cloud Datastore tables, write a Google App Engine Application that queries all rows, applies a function to derive the metric, and then renders results in a table using the Google charts and visualization API.

D. Load the data into Google BigQuery tables, write a Google Data Studio 360 report that connects to your data, calculates a metric, and then uses a filter expression to show only suboptimal rows in a table. Most Voted

Correct Answer: C 🗳️

Community vote distribution

D (80%)

C (20%)