

[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 36 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 36

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

Flowlogistic Case Study -

Company Overview -

Flowlogistic is a leading logistics and supply chain provider. They help businesses throughout the world manage their resources and transport them to their final destination. The company has grown rapidly, expanding their offerings to include rail, truck, aircraft, and oceanic shipping.

Company Background -

The company started as a regional trucking company, and then expanded into other logistics market. Because they have not updated their infrastructure, managing and tracking orders and shipments has become a bottleneck. To improve operations, Flowlogistic developed proprietary technology for tracking shipments in real time at the parcel level. However, they are unable to deploy it because their technology stack, based on Apache Kafka, cannot support the processing volume. In addition, Flowlogistic wants to further analyze their orders and shipments to determine how best to deploy their resources.

Solution Concept -

Flowlogistic wants to implement two concepts using the cloud:

Use their proprietary technology in a real-time inventory-tracking system that indicates the location of their loads

-
- ⇒ Perform analytics on all their orders and shipment logs, which contain both structured and unstructured data, to determine how best to deploy resources, which markets to expand into. They also want to use predictive analytics to learn earlier when a shipment will be delayed.

Existing Technical Environment -

Flowlogistic architecture resides in a single data center:

⇒ Databases

8 physical servers in 2 clusters

- SQL Server `` user data, inventory, static data

3 physical servers

- Cassandra `` metadata, tracking messages

10 Kafka servers `` tracking message aggregation and batch insert

⇒ Application servers `` customer front end, middleware for order/customs

60 virtual machines across 20 physical servers

- Tomcat `` Java services

- Nginx `` static content

- Batch servers

⇒ Storage appliances

- iSCSI for virtual machine (VM) hosts

- Fibre Channel storage area network (FC SAN) `` SQL server storage

- Network-attached storage (NAS) image storage, logs, backups

⇒ 10 Apache Hadoop /Spark servers

- Core Data Lake

- Data analysis workloads

⇒ 20 miscellaneous servers

- Jenkins, monitoring, bastion hosts,

Business Requirements -

⇒ Build a reliable and reproducible environment with scaled panty of production.

⇒ Aggregate data in a centralized Data Lake for analysis

⇒ Use historical data to perform predictive analytics on future shipments

⇒ Accurately track every shipment worldwide using proprietary technology

⇒ Improve business agility and speed of innovation through rapid provisioning of new resources

⇒ Analyze and optimize architecture for performance in the cloud

⇒ Migrate fully to the cloud if all other requirements are met

Technical Requirements -

Handle both streaming and batch data

▪

⇒ Migrate existing Hadoop workloads

⇒ Ensure architecture is scalable and elastic to meet the changing demands of the company.

⇒ Use managed services whenever possible

⇒ Encrypt data flight and at rest

⇒ Connect a VPN between the production data center and cloud environment

SEO Statement -

We have grown so quickly that our inability to upgrade our infrastructure is really hampering further growth and efficiency. We are efficient at moving shipments around the world, but we are inefficient at moving data around.

We need to organize our information so we can more easily understand where our customers are and what they are shipping.

CTO Statement -

IT has never been a priority for us, so as our data has grown, we have not invested enough in our technology. I have a good staff to manage IT, but they are so busy managing our infrastructure that I cannot get them to do the things that really matter, such as organizing our data, building the analytics, and figuring out how to implement the CFO' s tracking technology.

CFO Statement -

Part of our competitive advantage is that we penalize ourselves for late shipments and deliveries. Knowing where our shipments are at all times has a direct correlation to our bottom line and profitability. Additionally, I don't want to commit capital to building out a server environment.

Flowlogistic's CEO wants to gain rapid insight into their customer base so his sales team can be better informed in the field. This team is not very technical, so they've purchased a visualization tool to simplify the creation of BigQuery reports. However, they've been overwhelmed by all the data in the table, and are spending a lot of money on queries trying to find the data they need. You want to solve their problem in the most cost-effective way. What should you do?

- A. Export the data into a Google Sheet for virtualization.
- B. Create an additional table with only the necessary columns.
- C. Create a view on the table to present to the virtualization tool.
- D. Create identity and access management (IAM) roles on the appropriate columns, so only they appear in a query.


Show Suggested Answer

by [deleted] at March 20, 2020, 4:36 p.m.

Comments



Type your comment...

Submit

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

Answer: C

   upvoted 39 times

  **Radhika7983** Highly Voted 4 years, 5 months ago

Answer is C. A logical view can be created with only the required columns which is required for visualization. B is not the right option as you will create a table and make it static. What happens when the original data is updated. This new table will not have the latest data and hence view is the best possible option here.

   upvoted 25 times

  **jin0** 2 years, 2 months ago

I don't think so because in question they worried about spending money for query but, using view could not make money safe because logical view scan all of the data in the table. so, for saving money for query then Answer B is more suitable

   upvoted 4 times

  **mark1223jkh** 11 months, 3 weeks ago

The point is reducing the number of columns, not caching. Yes, it will query the table, but with only the necessary columns the view has.

   upvoted 2 times

  **cqrm3n** Most Recent 3 months ago

Selected Answer: C

Answer is C:

A BigQuery view solves this problem by:

Simplifying the dataset → Shows only the relevant columns/data.

Improving query efficiency → Reduces query costs by filtering unnecessary data.

Enhancing usability → Sales teams get only the insights they need without dealing with complex

queries.

Reducing costs → Since views don't store data separately, they don't incur additional storage costs.

   upvoted 1 times

  **SamuelTsch** 6 months, 2 weeks ago

Selected Answer: C

Answer C contains both requirements - minimal columns by creating view of the table and visualization tool for report


   upvoted 3 times

  **rocky48** 1 year, 5 months ago

Selected Answer: C

Answer: C

   upvoted 2 times

  **rtcpost** 1 year, 6 months ago

Selected Answer: C

C. Create a view on the table to present to the virtualization tool.

Creating a view in BigQuery allows you to define a virtual table that is a subset of the original data, containing only the necessary columns or filtered data that the sales team requires for their reports. This approach is cost-effective because it doesn't involve exporting data to external tools or creating additional tables, and it ensures that the sales team is working with the specific data they need without running expensive queries on the full dataset. It simplifies the data for non-technical users while keeping the data in BigQuery, which is a powerful and cost-efficient data warehousing solution.

Options A (exporting to Google Sheet) and B (creating an additional table) might introduce data redundancy and maintenance overhead, and they don't provide the same level of control and security as creating a view. Option D (IAM roles) doesn't address the issue of simplifying the data for the sales team; it's more focused on access control.



   upvoted 4 times

  **Mathew106** 1 year, 9 months ago

Selected Answer: C

C. You won't pay for storage for the view, and it will only include the necessary columns. Even if we assume that we don't talk about a materialized view, a logical view query can use the cache as much as a table query. So a new table does not have any benefit over a view, even if the view is logical.

   upvoted 2 times

  **abi01a** 2 years ago

The answer is C

   upvoted 2 times

  **kplam** 2 years ago

Answer is C

   upvoted 3 times

  **lucaluca1982** 2 years, 1 month ago

Selected Answer: B

B it is more cost-effective and efficient approach to handle reports



   upvoted 1 times

  **bha11111** 2 years, 1 month ago

Selected Answer: C

C is correct

   upvoted 3 times

  **Booqq** 2 years, 2 months ago

C — view is better than another table to keep data consistent

   upvoted 3 times

  **JJJJim** 2 years, 2 months ago

Selected Answer: C

Answer is C, creating view tables can easy and flexible to do the most cost-effective way.

   upvoted 2 times

  **PolyMoe** 2 years, 3 months ago

Selected Answer: C




The appropriate solution is C, creating a view on the table, by selecting the relevant columns only (and not by creating another, static, table)

   upvoted 2 times

  **dconesoko** 2 years, 3 months ago

Selected Answer: B

Providing that the question was explicit in B and D about the selection of the appropriate columns it quite intriguing that it did not mention the selection of the appropriate column for the view. We can definitely build a view which might just present the same data or something much complex, thus i vote for B



   upvoted 1 times

   upvoted 1 times

  **GCPpro** 2 years, 3 months ago

C is the correct answer

   upvoted 1 times

  **Isaga** 2 years, 3 months ago

Selected Answer: C

I mean C

   upvoted 1 times

[Load full discussion...](#)



Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)



© 2024 ExamTopics