

[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 101 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 101

Topic #: 1

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

You need to copy millions of sensitive patient records from a relational database to BigQuery. The total size of the database is 10 TB. You need to design a solution that is secure and time-efficient. What should you do?

- A. Export the records from the database as an Avro file. Upload the file to GCS using gsutil, and then load the Avro file into BigQuery using the BigQuery web UI in the GCP Console.
- B. Export the records from the database as an Avro file. Copy the file onto a Transfer Appliance and send it to Google, and then load the Avro file into BigQuery using the BigQuery web UI in the GCP Console.
- C. Export the records from the database into a CSV file. Create a public URL for the CSV file, and then use Storage Transfer Service to move the file to Cloud Storage. Load the CSV file into BigQuery using the BigQuery web UI in the GCP Console.
- D. Export the records from the database as an Avro file. Create a public URL for the Avro file, and then use Storage Transfer Service to move the file to Cloud Storage. Load the Avro file into BigQuery using the BigQuery web UI in the GCP Console.



[Show Suggested Answer](#)

by [Rajokkiyam](#) at March 22, 2020, 6:35 a.m.

Comments

Type your comment...

[Submit](#)

 **Ganshank**  5 years ago

You are transferring sensitive patient information, so C & D are ruled out. Choice comes down to A & B. Here it gets tricky. How to choose Transfer Appliance: (<https://cloud.google.com/transfer-appliance/docs/2.0/overview>) Without knowing the bandwidth, it is not possible to determine whether the upload can be completed within 7 days, as recommended by Google. So the safest and most performant way is to use Transfer Appliance. Therefore my choice is B.

   upvoted 63 times

 **grshankar9** 3 months, 2 weeks ago

With reasonable network connectivity (for example, 1 Gbps), transferring 100 TB of data online takes over 10 days to complete. If this rate is acceptable, an online transfer is likely a good solution for your needs. If you only have a 100 Mbps connection (or worse from a remote location), the same transfer takes over 100 days. At this point, it's worth considering an offline-transfer option such as Transfer Appliance.

   upvoted 1 times

 **tprashanth** 4 years, 9 months ago

<https://cloud.google.com/solutions/migration-to-google-cloud-transferring-your-large-datasets>

The table shows for 1Gbps, it takes 30 hrs for 10 TB. Generally, corporate internet speeds are over 1Gbps. I'm inclined to pick A

   upvoted 5 times

 **BigQuery** 3 years, 5 months ago

SAY MY NAME!


You need to Transfer Sensitive Patient information, over public ISP you shouldn't do that.

   upvoted 3 times

 **grshankar9** 3 months, 2 weeks ago

Security is not a concern, the data is encrypted at rest as well as in transit

   upvoted 1 times

 **forepick** 1 year, 11 months ago

If you transfer 10TBs over the wire, your network will be blocked for the entire transfer time. This isn't something a company would be happy to swallow.

   upvoted 2 times

 **TNT87** 4 years, 7 months ago

Answer is B,gsutil has a limit of 1TBaccording to Google documentation,if data is morethan 1TBthen we have to use Transfer Appliance.

   upvoted 19 times

 **grshankar9** 3 months, 2 weeks ago

While gsutil itself doesn't have a strict data limit, the underlying Google Cloud Storage service does, allowing for individual object sizes up to 5 Terabytes (TiB). This means that when using gsutil to transfer data, the maximum file size you can upload or download is 5 TiB

   upvoted 1 times

 **grshankar9** 3 months, 2 weeks ago

According to Google documentation, for files of size > 1TB STS is to be used

   upvoted 1 times

 **Yiouk** 3 years, 9 months ago

The answer is clearly seen here: <https://cloud.google.com/architecture/migration-to-google-cloud-transferring-your-large-datasets#transfer-options>

   upvoted 9 times

 **AzureDP900** 2 years, 4 months ago

B is right answer

   upvoted 4 times

 **SSV**  4 years, 10 months ago

Answer should be B: A is also correct but it has its own limit. It allows only 5TB data upload at a time to cloud storage.

<https://cloud.google.com/storage/quotas>


I will go with B

   upvoted 8 times

 **VASI** 4 years, 3 months ago


5Tb "for individual objects". Create smaller AVRO files.

   upvoted 2 times

  **VASI** 4 years, 3 months ago

AVRO compression can reduce file size to a tenth

   upvoted 3 times

  **grshankar9** Most Recent 3 months, 2 weeks ago

Selected Answer: A

C and D are ruled out owing to limitations of Cloud Storage with regards to max file size of 5 TB (Refer - <https://cloud.google.com/storage-transfer/docs/known-limitations-transfer#:~:text=performance%20uploads%3A%206GiB-,Scaling%20limitations,of%20Gbps%20in%20transfer%20speed>). The answer should be between A & B. if we were to assume bandwidth of 1 GBPS, it would only take about 1 day. Even if the bandwidth were a tenth of 1 GBPS, it would take about 10 days to transfer 10 TB. With a transfer appliance, it would take minimum of 25 days. I would go with A.



   upvoted 1 times

  **clouditis** 4 months, 3 weeks ago

Selected Answer: B

B - only because it says sensitive data, you can never be sure uploading over internet!

   upvoted 1 times

  **Preetmehta1234** 7 months, 1 week ago

Selected Answer: B



Also,

For your scenario with 10 TB of data in Cloud SQL, if you export to Avro without specifying compression, you can expect the resulting Avro file to be around the same size, potentially slightly smaller depending on the data characteristics. Here in this question, there is no mentioning about compression.

So let's not assume that the data being used in Avro format will get compressed.



If Google cloud storage itself, can't handle an object of size greater than 5 TB, there is no point of using gsutil

   upvoted 1 times

  **grshankar9** 3 months, 2 weeks ago

gsutil is deprecated and has been replaced by gcloud storage. Gcloud storage is faster and requires less manual optimization for uploads and downloads

   upvoted 1 times

  **Preetmehta1234** 7 months, 1 week ago

Selected Answer: B

<https://cloud.google.com/storage-transfer/docs/known-limitations-transfer>

Cloud Storage 5TiB object size limit

Cloud Storage supports a maximum single-object size up to 5 terabytes. If you have objects larger than 5TiB, the object transfer fails for those objects for either Cloud Storage or Storage Transfer Service.

   upvoted 1 times

  **hussain.sain** 10 months, 1 week ago

Selected Answer: D

while Option A is feasible and could work depending on specific requirements and security measures implemented, Option D (exporting as Avro, using Storage Transfer Service, and then loading into BigQuery) generally offers a more secure, efficient, and managed approach for transferring sensitive patient records into BigQuery from a relational database. Avro files uploaded to GCS will need to be secured. While GCS itself offers security features like IAM policies and access controls, using a public URL (as suggested in Option A) introduces additional security concerns.

   upvoted 1 times

  **Naresh_4u** 1 year ago

Selected Answer: B

to securely transfer data and looking at the size of data B is the correct option.

   upvoted 1 times

  **GCanteiro** 1 year, 3 months ago

Selected Answer: A

IMO "A" is the most suitable option since the transfer appliance could take 25 days to get the appliance and then 25 days to ship it back and have the data available.

<https://cloud.google.com/storage-transfer/docs/4.0/overview#transfer-speeds>

   upvoted 1 times

  **TVH_Data_Engineer** 1 year, 4 months ago

Selected Answer: B

Given the sensitivity of the patient records and the large size of the data, using Google's Transfer Appliance is a secure and efficient method. The Transfer Appliance is a hardware solution provided by Google for transferring large amounts of data. It

efficient method. The Transfer Appliance is a hardware solution provided by Google for transferring large amounts of data. It enables you to securely transfer data without exposing it over the internet.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **rocky48** 1 year, 5 months ago

Selected Answer: B

Option B combines security, efficiency, and ease of use, making it a suitable choice for transferring sensitive patient records to BigQuery.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **spicebits** 1 year, 5 months ago

Selected Answer: A

10 TB is nothing. With a single 10 GB interconnect you could transfer the data in 3 hours or even with a 1 GB speeds without interconnect you could transfer it in one weekend. The transfer appliance will take 25 days to get the appliance and then 25 days while you wait for the data to be available that is not "time-efficient" at all. I go with A instead of B.

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **spicebits** 1 year, 5 months ago

I got the 25 days + 25 days from here: <https://cloud.google.com/transfer-appliance/docs/4.0/overview#transfer-speeds>

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **A_Nasser** 1 year, 7 months ago

Selected Answer: A

transfer appliance will take time more than gsutil. and we did not mention yet if the location of the organization has google data centre

👍 ↩ 🚩 upvoted 3 times

🗄️ 👤 **DineshVarma** 1 year, 8 months ago

Selected Answer: D

As per Google recommendation above 1TB of transfer from onprem or from Google cloud or other cloud storage like s3 etc we need to use storage transfer service.

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **arien_chen** 1 year, 8 months ago

Transfer Appliance would take 20 days for expected turnaround time. <https://cloud.google.com/architecture/migration-to-google-cloud-transferring-your-large-datasets#expected%20turnaround~:text=The%20expected%20turnaround%20time%20for%20a%20network%20appliance%20to%20be%20shipped%2C%20loaded%20with%20your%20data%2C%20shipped%20back%2C%20and%20rehydrated%20on%20Google%20Cloud%20is%2020%20days.>

The best answer would be A.

If gsutil consume/leverage 100MB it would take 12 days and more time-efficient than B.

This is a reasonable assumption.

<https://cloud.google.com/static/architecture/images/big-data-transfer-how-to-get-started-transfer-size-and-speed.png>

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **Colourseun** 1 year, 8 months ago

I will go with "A" because of the transition time to take transfer appliance to Google and that also depends in the organisation location. gsutil works anywhere internet is available.

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **aewis** 1 year, 9 months ago

Selected Answer: B

A will take crazy time if the organization didn't have a dedicated link

👍 ↩ 🚩 upvoted 1 times

[Load full discussion...](#)

> Home

> All Exams

> Examtopics PRO

> Training Courses

