⟵ **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 216 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 216

Topic #: 1

[All Professional Data Engineer Questions]

You are developing an Apache Beam pipeline to extract data from a Cloud SQL instance by using JdbcIO. You have two projects running in Google Cloud. The pipeline will be deployed and executed on Dataflow in Project A. The Cloud SQL. instance is running in Project B and does not have a public IP address. After deploying the pipeline, you noticed that the pipeline failed to extract data from the Cloud SQL instance due to connection failure. You verified that VPC Service Controls and shared VPC are not in use in these projects. You want to resolve this error while ensuring that the data does not go through the public internet. What should you do?

    A. Set up VPC Network Peering between Project A and Project B. Add a firewall rule to allow the peered subnet range to access all instances on the network.

    B. Turn off the external IP addresses on the Dataflow worker. Enable Cloud NAT in Project A.

    C. Add the external IP addresses of the Dataflow worker as authorized networks in the Cloud SQL instance.

    D. Set up VPC Network Peering between Project A and Project B. Create a Compute Engine instance without external IP address in Project B on the peered subnet to serve as a proxy server to the Cloud SQL database.

Show Suggested Answer

by 👤 **e70ea9e** at *Dec. 30, 2023, 9:41 a.m.*

Comments

Type your comment...

⊟ 👤 **aoifneofi_ef** `Highly Voted 👍` 8 months, 2 weeks ago

`Selected Answer: D`

It is a tie between A and D.
Option A will definitely provide necessary connectivity but is less secure as access is enabled to "all instances". Which i feel is unnecessary considering industry best practices.

Option D provides the necessary connectivity but brings in the unnecessary overhead of managing an extra VM and introduces a bit of extra complexity.

Since the question emphasises on data not going through public internet(which is satisfied in both options), i would prioritise security over simplicity and choose option D in this case.

👍 ↩ ⚑ upvoted 6 times

⊟ 👤 **clouditis** `Highly Voted 👍` 4 months, 3 weeks ago

`Selected Answer: A`

A looks to be the best out of the 4, D is complicated involving Compute Engine which is unnecessary making it cumbersome to address the problem

👍 ↩ ⚑ upvoted 5 times

⊟ 👤 **71083a7** `Most Recent ⊙` 3 months, 1 week ago

`Selected Answer: D`

"all instances" freaks me out

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **julydev82** 4 months, 3 weeks ago

`Selected Answer: A`

vpc create a subnet beetwen sql Cloud and the dataflow.
https://cloud.google.com/sql/docs/mysql/private-ip#multiple_vpc_connectivity

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **fadlkhafdofpew** 7 months, 3 weeks ago

`Selected Answer: A`

The answer is A. While D might work, it adds unnecessary complexity. Setting up a proxy is an extra layer of infrastructure that isn't required

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **Saaaurabh** 8 months, 3 weeks ago

`Selected Answer: A`

If properly implemented with the right routing and firewall rules, Option A can be the correct and most straightforward solution, as it leverages VPC Peering to maintain internal traffic without going through the public internet.

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **meh_33** 8 months, 4 weeks ago

A is correct

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **kk1211** 9 months, 4 weeks ago

still confused between A and D

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **Lenifia** 10 months ago

`Selected Answer: A`

A is correct

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **kajitsu** 10 months, 1 week ago

`Selected Answer: A`

no proxy needed

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **Lestrang** 11 months ago

`Selected Answer: A`

People referencing "VPC Network Peering does not provide transitive routing. For example, if VPC networks net-a and net-b are connected using VPC Network Peering, and VPC networks net-a and net-c are also connected using VPC Network

Peering, VPC Network Peering does not provide connectivity between net-b and net-c."

the question states that cloud sql is running in project B.
Which means the instance is already part of the VPC in project B, so with Network Peering workers from A can definitely access data in B. No proxy is needed.

👍 ↩ ⚑ upvoted 1 times

⊟ 👤 **fabiogoma** 11 months, 2 weeks ago

Selected Answer: A

Why so many people are voting for D? There's no need for a proxy, the peering is enough to allow network traffic between subnets.

👍 ↩ ⚑ upvoted 2 times

　⊟ 👤 **fabiogoma** 11 months, 2 weeks ago

　Now I see why, I put this on ChatGPT and it thinks the right answer is D. I'm pretty sure that's a hallucination.

　👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **ccpmad** 11 months, 2 weeks ago

Proxy? no, it is not necessary..

A

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **josech** 11 months, 3 weeks ago

Selected Answer: D

https://cloud.google.com/sql/docs/mysql/connect-multiple-vpcs

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **chrissamharris** 1 year, 2 months ago

Selected Answer: A

A - The requirement for a proxy is un-necessary:
https://cloud.google.com/sql/docs/mysql/private-ip#multiple_vpc_connectivity

👍 ↩ ⚑ upvoted 4 times

⊟ 👤 **ML6** 1 year, 2 months ago

Option D. Source: https://cloud.google.com/sql/docs/mysql/private-ip#multiple_vpc_connectivity

👍 ↩ ⚑ upvoted 2 times

⊟ 👤 **JyoGCP** 1 year, 2 months ago

Selected Answer: D

Option D

👍 ↩ ⚑ upvoted 1 times

Load full discussion...