EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 253 DISCUSSION

Actual exam question from Google's Professional Data Engineer
Question #: 253
Topic #: 1
[All Professional Data Engineer Questions]

You are deploying a batch pipeline in Dataflow. This pipeline reads data from Cloud Storage, transforms the data, and then writes the data into BigQuery. The security team has enabled an organizational constraint in Google Cloud, requiring all Compute Engine instances to use only internal IP addresses and no external IP addresses. What should you do?

A. Ensure that your workers have network tags to access Cloud Storage and BigQuery. Use Dataflow with only internal IP addresses.

B. Ensure that the firewall rules allow access to Cloud Storage and BigQuery. Use Dataflow with only internal IPs.

C. Create a VPC Service Controls perimeter that contains the VPC network and add Dataflow, Cloud Storage, and BigQuery as allowed services in the perimeter. Use Dataflow with only internal IP addresses.

D. Ensure that Private Google Access is enabled in the subnetwork. Use Dataflow with only internal IP addresses.

Show Suggested Answer

by scaenruy at *Jan. 3, 2024, 4:27 p.m.*

Comments

Type your comment...

Submit

raaad Highly Voted 10 months ago

Selected Answer: D

- Private Google Access for services allows VM instances with only internal IP addresses in a VPC network or on-premises networks (via Cloud VPN or Cloud Interconnect) to reach Google APIs and services.
- When you launch a Dataflow job, you can specify that it should use worker instances without external IP addresses if Private Google Access is enabled on the subnetwork where these instances are launched.
- This way, your Dataflow workers will be able to access Cloud Storage and BigQuery without violating the organizational constraint of no external IPs.

upvoted 5 times

Jordan18 10 months ago

why not C?

upvoted 3 times

BIGQUERY_ALT_ALT 9 months, 4 weeks ago

VPC Service Controls are typically used to define and enforce security perimeters around APIs and services, restricting their access to a specified set of Google Cloud projects. In this scenario, the security constraint is focused on Compute Engine instances used by Dataflow, and VPC Service Controls might be considered a bit heavy-handed for just addressing the internal IP address requirement.

upvoted 3 times

GCP001 10 months ago

Even if you create VPC service control, your dataflow worker will run on google compute engine instances with private ips only after policy enforcement.
Without external IP addresses, you can still perform administrative and monitoring tasks.
You can access your workers by using SSH through the options listed in the preceding list. However, the pipeline cannot access the internet, and internet hosts cannot access your Dataflow workers.

upvoted 5 times

GCP001 10 months ago

ref - https://cloud.google.com/dataflow/docs/guides/routes-firewall

upvoted 4 times

Lestrang Most Recent 4 months, 4 weeks ago

Selected Answer: D

No way it is C.
Like the use case for Google VPC Service Controls perimeter is not to establish secure connectivity on its own but rather to control connectivity, like allowing vms within x premise to access, and blocking vms outside premise even if in same VPC from access.

D on the other hand is completely sensical.

upvoted 1 times

Moss2011 8 months ago

Selected Answer: C

According to this documentation: https://cloud.google.com/vpc-service-controls/docs/overview I think the correct answer is C. Take into account the phrase "organizational constraint" and the VPC Service Control allow you to do that.

upvoted 1 times

Tryolabs 8 months, 1 week ago

Selected Answer: D

https://cloud.google.com/vpc/docs/private-google-access

"VM instances that only have internal IP addresses (no external IP addresses) can use Private Google Access. They can reach the external IP addresses of Google APIs and services."

upvoted 1 times

pandeyspecial 9 months, 1 week ago

Selected Answer: C

It should be C

upvoted 1 times

Matt_108 9 months, 3 weeks ago

Selected Answer: C

Option D, as GCP001 said

upvoted 1 times

Matt_108 9 months, 3 weeks ago

Missclicked the answer <.<

upvoted 2 times

GCP001 10 months ago

Selected Answer: D

https://cloud.google.com/dataflow/docs/guides/routes-firewall

upvoted 4 times

scaenruy 10 months ago

Selected Answer: C

C. Create a VPC Service Controls perimeter that contains the VPC network and add Dataflow, Cloud Storage, and BigQuery as allowed services in the perimeter. Use Dataflow with only internal IP addresses.

upvoted 1 times

BIGQUERY_ALT_ALT 9 months, 4 weeks ago

C is wrong. Option D is simple and straight forward. VPC Service Controls are typically used to define and enforce security perimeters around APIs and services, restricting their access to a specified set of Google Cloud projects. In this scenario, the security constraint is focused on Compute Engine instances used by Dataflow, and VPC Service Controls might be considered a bit heavy-handed for just addressing the internal IP address requirement.

upvoted 2 times