◌ **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 295 DISCUSSION**

Actual exam question from Google's Professional Data Engineer

Question #: 295

Topic #: 1

[All Professional Data Engineer Questions]

You are designing the architecture to process your data from Cloud Storage to BigQuery by using Dataflow. The network team provided you with the Shared VPC network and subnetwork to be used by your pipelines. You need to enable the deployment of the pipeline on the Shared VPC network. What should you do?

> A. Assign the compute.networkUser role to the Dataflow service agent.
>
> B. Assign the compute.networkUser role to the service account that executes the Dataflow pipeline.
>
> C. Assign the dataflow.admin role to the Dataflow service agent.
>
> D. Assign the dataflow.admin role to the service account that executes the Dataflow pipeline.

Show Suggested Answer

by 👤 **scaenruy** at *Jan. 4, 2024, 11:50 a.m.*

Comments

    Type your comment...

Submit

⊟ 👤 **raaad** 👍Highly Voted 👍 1 year, 3 months ago

    Selected Answer: A

- Dataflow service agent is the one responsible for setting up and managing the network resources that Dataflow requires.
- By granting the compute.networkUser role to this service agent, we are enabling it to provision the necessary network resources within the Shared VPC for your Dataflow job.

👍 ↩ 🚩 upvoted 10 times

👤 **saschak94** `Highly Voted 👍` 1 year, 2 months ago

Selected Answer: A

All projects that have used the resource Dataflow Job have a Dataflow Service Account, also known as the Dataflow service agent.

Make sure the Shared VPC subnetwork is shared with the Dataflow service account and has the Compute Network User role assigned on the specified subnet.

👍 ↩ 🚩 upvoted 5 times

👤 **gabbferreira** `Most Recent ⊘` 1 week, 5 days ago

Selected Answer: A

It's a

👍 ↩ 🚩 upvoted 1 times

👤 **desertlotus1211** 1 month, 1 week ago

Answer is B:

The Dataflow service agent manages Dataflow internals, but does not launch pipeline worker VMs. So Answer A is incorrect

👍 ↩ 🚩 upvoted 1 times

👤 **loki82** 3 months, 1 week ago

Selected Answer: A

https://cloud.google.com/dataflow/docs/concepts/security-and-permissions#df-service-account

👍 ↩ 🚩 upvoted 1 times

> 👤 **desertlotus1211** 1 month, 1 week ago
>
> The Dataflow service agent manages Dataflow internals, but does not launch pipeline worker VMs.
>
> 👍 ↩ 🚩 upvoted 1 times

> 👤 **gord_nat** 1 month, 1 week ago
>
> There is no compute.networkUser role in df agent, only in IAM. Answer is B
>
> 👍 ↩ 🚩 upvoted 1 times

👤 **Pime13** 3 months, 3 weeks ago

Selected Answer: B

https://cloud.google.com/dataflow/docs/guides/specifying-networks
https://cloud.google.com/dataflow/docs/guides/specifying-networks#shared
Make sure the Shared VPC subnetwork is shared with the Dataflow service account and has the Compute Network User role assigned on the specified subnet. The Compute Network User role must be assigned to the Dataflow service account in the host project.

👍 ↩ 🚩 upvoted 1 times

👤 **SamuelTsch** 6 months ago

Selected Answer: A

From https://cloud.google.com/dataflow/docs/guides/specifying-networks, it says "Make sure the Shared VPC subnetwork is shared with the Dataflow service account and has the Compute Network User role assigned on the specified subnet. The Compute Network User role must be assigned to the Dataflow service account in the host project."

👍 ↩ 🚩 upvoted 2 times

> 👤 **ach5** 6 months ago
>
> service account - it's B
>
> 👍 ↩ 🚩 upvoted 4 times

👤 **Preetmehta1234** 7 months, 1 week ago

Selected Answer: B

If you see in the comments, A was answer by people around 8 months ago but recent ones have answered B with the documentation. The GCP documentation evolves with time

👍 ↩ 🚩 upvoted 4 times

👤 **Preetmehta1234** 7 months, 1 week ago

Selected Answer: B

service account that executes the Dataflow pipeline
It's straight forward

👍 ↩ 🚩 upvoted 2 times

Preetmehta1234 7 months, 1 week ago

Selected Answer: B

Assign the compute.networkUser role to the service account that executes the Dataflow pipeline

👍 ↩ 🚩 upvoted 2 times

Jeyaraj 9 months, 2 weeks ago

The correct answer is B. Assign the compute.networkUser role to the service account that executes the Dataflow pipeline.

Here's why:

Shared VPC and Network Access: When using a Shared VPC, you need to grant specific permissions to service accounts in the service project (where your Dataflow pipeline runs) to access resources in the host project's network.
compute.networkUser Role: This role grants the necessary permissions for a service account to use the network resources in the Shared VPC. This includes accessing subnets, creating instances, and communicating with other services within the network.
Service Account for Pipeline Execution: The service account that executes your Dataflow pipeline is the one that needs these network permissions. This is because the Dataflow service uses this account to create and manage worker instances within the Shared VPC network.

👍 ↩ 🚩 upvoted 3 times

extraego 10 months, 4 weeks ago

Selected Answer: B

Dataflow service agent is a role that is assigned to a service account. So is compute.networkUser.
https://cloud.google.com/dataflow/docs/concepts/access-control#example

👍 ↩ 🚩 upvoted 4 times

josech 11 months, 3 weeks ago

Selected Answer: B

Option B https://cloud.google.com/knowledge/kb/dataflow-job-in-shared-vpc-xpn-permissions-000004261

👍 ↩ 🚩 upvoted 4 times

chrissamharris 1 year ago

Selected Answer: B

I believe the answer is B. All authentication documentation points to Service Accounts.
https://cloud.google.com/dataflow/docs/concepts/authentication#on-gcp

Dataflow service agent typically manages general interactions with the Dataflow service but does not execute the actual jobs.

👍 ↩ 🚩 upvoted 3 times

Matt_108 1 year, 3 months ago

Selected Answer: A

Option A, I do agree with Raaad, it's the dataflow service agent that needs the networkUser role, because it's the one that provisions the network resources https://cloud.google.com/dataflow/docs/guides/specifying-networks#shared

👍 ↩ 🚩 upvoted 3 times

tibuenoc 1 year, 2 months ago

But your link it's explain that "Network User role must be assigned to the Dataflow service account"

Make sure the Shared VPC subnetwork is shared with the Dataflow service account and has the Compute Network User role assigned on the specified subnet. The Compute Network User role must be assigned to the Dataflow service account in the host project.

👍 ↩ 🚩 upvoted 1 times

ML6 1 year, 2 months ago

All projects that have used the resource Dataflow Job have a Dataflow Service Account, also known as the Dataflow service agent.
Source: https://cloud.google.com/dataflow/docs/concepts/security-and-permissions#df-service-account

👍 ↩ 🚩 upvoted 2 times

task_7 1 year, 3 months ago

Selected Answer: B

compute.networkUser to the service account that executes the Dataflow pipeline.

👍 ↩ 🚩 upvoted 4 times

BIGQUERY_ALT_ALT 1 year, 3 months ago

Selected Answer: B

Option B is Correct.

Explanation:
You need to give compute networkuser role to service account that is processing the pipeline as it will need to deploy nessesary worker nodes on the shared vpc project.

Option A is incorrect as Dataflow Service Agent is Google MGS service account that will not responsible for running or deoplying workers in shared vpc.

Option C and D is incorrect as dataflow.admin is elevated privlages to create and manage all of dataflow components not deploying resources in shared vpc.

👍 ↩ 🚩 upvoted 2 times

Load full discussion…