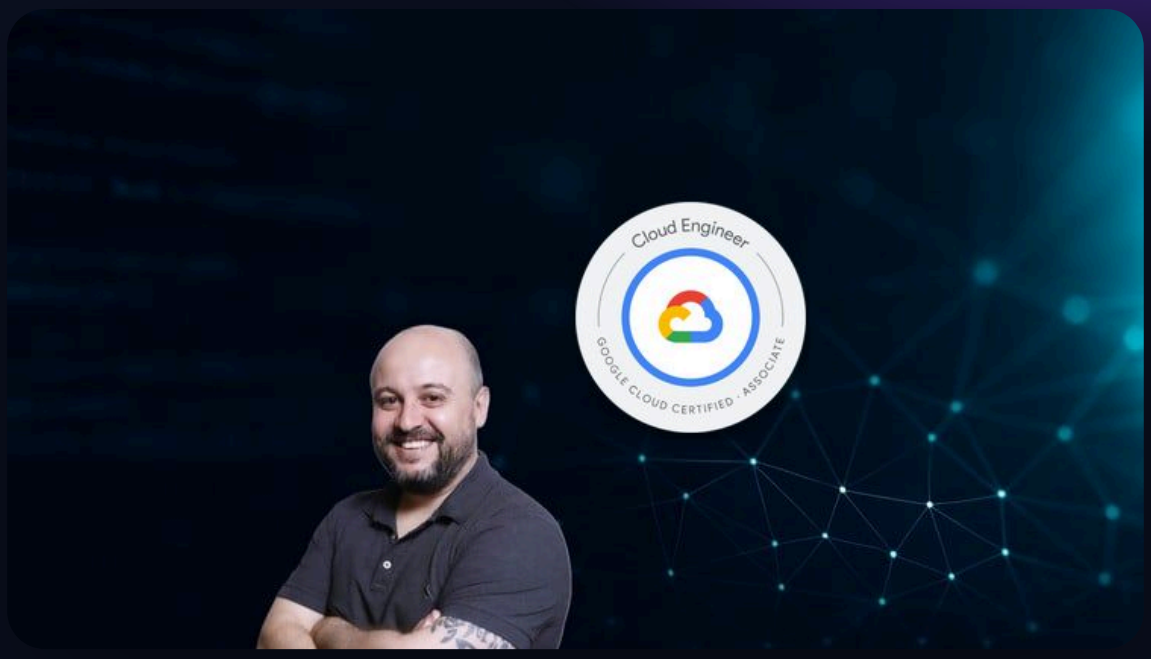
Strecht Simulados · Guia de preparação



Guia de preparação Google Associate Cloud Engineer

Tudo o que a prova cobra, organizado por domínio: conceitos-chave, comparações que confundem, pegadinhas reais, gatilhos do enunciado, questões comentadas e um plano para chegar pronto ao exame.

4	28	70%
capítulos por domínio	questões comentadas	nota de corte

6
provas completas no curso

ANTES DE COMEÇAR

Como usar este guia

Este guia foi feito para ser estudado junto com as 6 provas completas do curso Google Associate Cloud Engineer: o guia explica o que cai e onde estão as armadilhas; as provas treinam o ritmo e mostram o que você ainda não domina.

Quando	O que fazer
Semana 1	Faça o teste diagnóstico (página seguinte) e leia os capítulos 1 e 2. No fim da semana, faça a Prova 1 do curso no tempo do exame.
Semana 2	Leia os demais capítulos, começando pelos domínios em que você errou mais no diagnóstico. Faça as provas 2 e 3.
Semana 3	Releia os alertas de pegadinha e as tabelas de comparação. Faça as provas 4 e 5 e revise cada erro pela explicação.
Semana 4	Estude a folha de cola, faça a Prova 6 e refaça as provas abaixo de 80%. Agende o exame quando estiver acima de 80% em todas (a nota de corte é 70%).

Mapa do exame

Cada prova do curso tem 50 questões e 120 minutos, no formato do exame. Referência de aprovação: 70%.

Domínio (capítulo)	Peso	Neste guia
1. Configurar Acesso e Segurança	20%	18 conceitos · 14 pegadinhas
2. Configurar o Ambiente da Solução em Nuvem	20%	18 conceitos · 14 pegadinhas
3. Garantir a Operação Bem-Sucedida da Solução em Nuvem	30%	21 conceitos · 14 pegadinhas
4. Planejar e Implementar a Solução em Nuvem	30%	20 conceitos · 14 pegadinhas

Como ler os capítulos: os **tópicos-chave** são o que mais cai; a **tabela de memorização** separa o que a prova costuma confundir; os **alertas de prova** mostram as armadilhas das alternativas erradas; os **gatilhos** ligam palavras do enunciado à resposta; e o **checklist** diz se você está pronto para o domínio.

Estude um pouco todo dia, de graça

Instagram @strecht.simulados

Um post de prova por dia: questão comentada, pegadinhas e o "não confunda" de cada certificação.

Seguir no Instagram →

YouTube Strecht Simulados

Shorts diários com pegadinhas de prova e, aos sábados, um simulado comentado.

Inscrever-se no canal →

DIAGNÓSTICO

Você já sabe isso?

Responda às 12 questões sem consultar nada. As respostas, com a explicação e o capítulo para revisar, estão no fim do guia. Comece o estudo pelos capítulos em que errou.

D1 · Configurar Acesso e Segurança

Um administrador quer saber quais permissões fazem parte do papel roles/storage.objectViewer. Qual comando usar?

A. gcloud iam roles describe
B. gcloud auth revoke
C. gcloud iam roles delete
D. gcloud projects list

D2 · Configurar Acesso e Segurança

Um serviço do Cloud Run precisa ler um bucket usando a identidade dedicada sa-leitura, e não a conta de serviço padrão. Qual opção usar na implantação?

A. --labels com o nome da conta
B. --allow-unauthenticated no serviço
C. --service-account com o e-mail da conta
D. --set-env-vars com a chave da conta

D3 · Configurar Acesso e Segurança

A equipe de segurança deve criar, girar e excluir chaves de contas de serviço legadas, sem poder alterar outras configurações das contas nem personificá-las. Qual papel conceder?

A. Papel de administrador de chaves de contas de serviço
B. Papel de editor do projeto inteiro
C. Papel de criador de token da conta de serviço
D. Papel de usuário da conta de serviço

D4 · Configurar o Ambiente da Solução em Nuvem

A empresa quer isolar desenvolvimento e produção, com políticas e acessos diferentes para cada ambiente. Qual organização da hierarquia é recomendada?

A. Rótulos de ambiente em um só projeto
B. Uma conta de faturamento por servidor
C. Projetos separados sob pastas por ambiente
D. Um único projeto com redes separadas

D5 · Configurar o Ambiente da Solução em Nuvem

Ao configurar a federação de identidade de colaboradores para que funcionários autenticados no Okta acessem o Google Cloud, em que nível o pool de identidades é criado?

A. No nível da organização do Google Cloud
B. No nível de cada projeto de aplicação
C. No nível de cada pasta de ambiente
D. No nível da conta de faturamento

D6 · Configurar o Ambiente da Solução em Nuvem

O financeiro quer separar os custos por equipe e por ambiente nos relatórios de faturamento, mesmo quando os recursos ficam no mesmo projeto. O que usar?

A. Uma conta de faturamento para cada recurso
B. Rótulos nos recursos com equipe e ambiente
C. Cotas de uso diferentes para cada equipe
D. Políticas da organização por equipe

D7 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

Os auditores exigem guardar os logs de auditoria por cinco anos para consultas com SQL. Qual é a melhor configuração?

A. Sink de logs para o BigQuery
B. Retenção padrão do Logging
C. Captura de tela dos logs no console
D. Alertas de orçamento mensais

D8 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

Várias equipes compartilham um cluster GKE e precisam de separação lógica de recursos e cotas. O que usar?

A. Rótulos só nos nós
B. Namespaces por equipe
C. Um cluster por pod
D. Uma VPC por contêiner

D9 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

A segurança central precisa bloquear um intervalo de IPs maliciosos em todas as VPCs de todos os projetos da organização, sem que equipes de projeto possam sobrepor a regra. O que usar?

A. Regra de firewall da VPC em cada projeto
B. Política de firewall hierárquica na organização
C. Política de negação do IAM na organização
D. Rota estática para descartar o intervalo

D10 · Planejar e Implementar a Solução em Nuvem

Uma aplicação legada exige módulos específicos do kernel do sistema operacional e acesso direto à máquina. Qual opção de computação escolher?

A. VMs do Compute Engine
B. Cluster GKE Autopilot
C. Serviços do Cloud Run
D. Funções do Cloud Run

D11 · Planejar e Implementar a Solução em Nuvem

Uma equipe vai treinar por semanas um modelo grande dominado por multiplicações de matrizes, com lotes grandes e sem operações personalizadas no laço de treino. Qual hardware é mais indicado?

A. Nós de locatário único com CPU
B. Aceleradores Cloud TPU
C. GPUs NVIDIA em VMs N1
D. VMs de uso geral com CPU

D12 · Planejar e Implementar a Solução em Nuvem

Dados de um SaaS de anúncios devem ser carregados no BigQuery todo dia de forma gerenciada e agendada. O que usar?

A. Serviço de transferência de dados do BigQuery
B. Peering de VPC com o fornecedor do SaaS
C. Script manual executado por um analista
D. Snapshot diário de um disco permanente da VM

01

CAPÍTULO 1 · 20% DA PROVA

Configurar Acesso e Segurança

Aqui o foco é IAM e contas de serviço, com peso de 20%. Preciso saber montar uma vinculação (principal, papel e condição opcional), escolher o tipo de papel certo, entender herança e saber em que nível conceder. A prova adora contas de serviço: qual papel permite usar, personificar, criar ou gerenciar chaves, e como evitar chaves de longa duração com personificação e federação. Onde mais erro: confundir serviceAccountUser com serviceAccountTokenCreator, esquecer que permissões herdadas não somem ao remover o papel no filho e achar que allAuthenticatedUsers é só a empresa.

Neste capítulo

18 tópicos-chave · 14 pegadinhas · 6 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Política e vinculação do IAM

Cada vinculação associa um papel a um ou mais principais, com condição opcional. Principais incluem contas do Google, grupos, contas de serviço e domínios; rótulos, tags de rede e IPs não são principais.

Tipos de papéis

Básicos são Proprietário, Editor e Leitor (Visualizador), amplos demais para produção. Predefinidos são granulares por serviço, como os de instâncias do Compute Engine. Personalizados aplicam o menor privilégio com precisão quando nenhum predefinido serve.

Papéis personalizados

Criados em um projeto só valem naquele projeto; criados na organização valem em toda a hierarquia. Nem toda permissão é suportada em papéis personalizados. roles/iam.roleAdmin cria e mantém papéis sem gerenciar concessões.

Herança e política efetiva

A política efetiva é a união da política do recurso com as herdadas. Papel concedido na pasta continua valendo no projeto mesmo após remover o papel no projeto; é preciso remover na pasta. Papel na organização vale para pastas e projetos atuais e futuros.

Política no nível do recurso

Buckets e contas de serviço, entre outros, aceitam política de IAM própria; conceder no recurso limita o escopo, como dar leitura a um parceiro em um único bucket. Rotas e regras de firewall não têm política de IAM.

Condições e negação

Condições do IAM concedem acesso por atributos, como data de expiração do contrato. Políticas de negação prevalecem sobre concessões, por exemplo impedir exclusão de projetos de produção mesmo por proprietários.

Propagação

Mudanças de IAM costumam valer em cerca de dois minutos, podendo levar mais. Erro de permissão segundos após a concessão indica propagação, não necessidade de aprovação ou recriação da conta.

Conta de serviço: identidade e recurso

Como identidade, recebe papéis (inclusive em outros projetos). Como recurso, tem política que define quem pode usá-la. Desativar a conta bloqueia o uso de forma reversível, sem perder a configuração.

Papéis sobre contas de serviço

Create Service Accounts cria; Service Account Admin gerencia, altera e exclui; Service Account User (serviceAccountUser) anexa e usa a conta em recursos; Service Account Token Creator personifica e gera tokens curtos; Service Account Key Admin gerencia chaves.

Personificação

--impersonate-service-account executa gcloud com permissões da conta de serviço usando a própria identidade, exigindo Token Creator. O método generateAccessToken da Service Account Credentials API emite tokens OAuth 2.0 curtos, com limite padrão de 1 hora que pode chegar a 12 horas por política da organização.

Chaves de contas de serviço

Chaves gerenciadas pelo Google são usadas internamente e giradas automaticamente. Chaves gerenciadas pelo usuário não expiram por padrão e exigem cuidado; gcloud iam service-accounts keys list mostra as existentes. Chave vazada deve ser desativada ou excluída de imediato e o uso investigado.

Federação de identidade da carga de trabalho

Troca tokens de provedores externos (CI, AWS, GitHub Actions) por credenciais de curta duração, sem chaves. Condições de atributo no provedor do pool limitam quem autentica, como um repositório específico. Com personificação, o principal do pool recebe roles/iam.workloadIdentityUser na conta de serviço; quando possível, o Google recomenda dar acesso direto aos recursos.

Federação de identidade para o GKE

Antigo Workload Identity: dá credenciais por pod sem chaves e sem usar a conta dos nós. Papéis podem ser concedidos diretamente ao identificador principal da ServiceAccount do Kubernetes, sem conta de serviço do IAM intermediária.

Contas padrão e agentes de serviço

A conta padrão do Compute Engine tem papel de editor; a boa prática é conta dedicada por carga de trabalho com papéis mínimos. Agentes de serviço são contas gerenciadas pelo Google que permitem aos serviços agir no projeto; remover seus papéis causa falhas.

Acesso ao Cloud Run

roles/run.invoker controla quem invoca um serviço autenticado; para chamada entre serviços, conceda à conta de serviço do serviço chamador, não a allUsers. --allow-unauthenticated concede run.invoker a allUsers. A identidade do próprio serviço é definida com --service-account.

allUsers x allAuthenticatedUsers

allUsers é qualquer pessoa na internet. allAuthenticatedUsers inclui qualquer conta Google autenticada, de dentro ou fora da empresa, e não limita aos funcionários.

Acesso a buckets

Acesso uniforme no nível do bucket desativa ACLs e deixa só o IAM. storage.objectViewer e storage.objectAdmin concedem leitura de objetos. gcloud storage buckets get-iam-policy lista vinculações definidas no próprio bucket.

Ferramentas de análise de acesso

Policy Troubleshooter explica se um principal tem uma permissão em um recurso e por quê. Recomendações de papéis (Policy Intelligence e Active Assist) sugerem papéis menores para permissões nunca usadas. roles/iam.securityReviewer lista recursos e políticas sem alterar; roles/browser navega na hierarquia sem ver recursos.

Macetes para memorizar

💡 Papéis de conta de serviço

Creator cria, Admin administra, User usa, Token Creator vira a conta, Key Admin cuida das chaves.

💡 Vinculação

Quem (principal) + o quê (papel) + quando (condição opcional).

Herança

IAM soma de cima para baixo: o que a pasta dá, o projeto não tira.

Chaves

Chave longa é risco: troque por personificação (dentro do Google) ou federação (fora do Google).

Público

allUsers = mundo; allAuthenticatedUsers = qualquer conta Google, não só a empresa.

Comandos gcloud de IAM

get-iam-policy lê, add-iam-policy-binding concede, roles describe abre o papel, service-accounts create cria a conta.

Tabela de memorização: não confunda

Quando é...	...e quando é
Usuário da conta de serviço (serviceAccountUser) Anexar a conta a VMs ou recursos que o principal cria.	**Criador de token da conta de serviço (Token Creator)** Personificar a conta e gerar credenciais curtas (--impersonate-service-account, generateAccessToken).
Administrador de chaves de contas de serviço Criar, girar e excluir chaves sem alterar contas nem personificá-las.	**Criador de contas de serviço** Criar contas novas sem excluí-las nem gerenciar chaves.
Papel na política do projeto O acesso deve valer para todos os recursos daquele tipo no projeto.	**Papel na política do recurso (bucket, conta de serviço)** Limitar a um bucket ou a uma única conta de serviço, seguindo o menor privilégio.
Federação de identidade da carga de trabalho Workload fora do Google Cloud (CI, AWS, GitHub Actions) sem guardar chaves.	**Personificação de conta de serviço** Pessoa ou script já autenticado no Google agindo temporariamente como a conta.
Revisor de segurança (securityReviewer) Auditar configurações e políticas de IAM de toda a organização sem alterar.	**Navegador (browser)** Navegar por pastas e projetos sem ver o conteúdo dos recursos.
Administrador de papéis (roleAdmin) Criar e manter papéis personalizados sem gerenciar quem os recebe.	**Administrador de IAM do projeto** Também gerencia concessões de papéis.
Política de negação do IAM Bloquear uma permissão mesmo para quem tem papel amplo, como proprietário.	**Remover o papel do usuário** Retirar uma concessão específica no nível em que foi feita.

Números que caem

Item	Valor

Peso do domínio no guia vigente	**20%**
Propagação típica de mudanças de IAM	**cerca de dois minutos**
Validade padrão do token do generateAccessToken	**1 hora, até 12 horas por política da organização**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Remover o papel no projeto e achar que o acesso acabou, embora o papel tenha sido dado na pasta.
Como não cair: A política efetiva é a união com as herdadas; remova na pasta.

⚠ **Alerta de prova:** Conceder serviceAccountUser no projeto para que o desenvolvedor anexe só uma conta.
Como não cair: No projeto vale para todas as contas; conceda na política da própria conta de serviço.

⚠ **Alerta de prova:** Dar Token Creator para quem só precisa gerenciar chaves.
Como não cair: Token Creator permite personificar; o papel certo é Service Account Key Admin.

⚠ **Alerta de prova:** Dar Service Account Admin para quem só deve criar contas.
Como não cair: Admin também altera, exclui e gerencia políticas; use Create Service Accounts.

⚠ **Alerta de prova:** Usar --account=conta-de-servico para testar com permissões da conta.
Como não cair: --account exige a conta já autenticada com credencial própria; use --impersonate-service-account.

⚠ **Alerta de prova:** Guardar chave JSON em bucket privado, segredo do CI ou com rotação anual como solução segura.
Como não cair: Continua sendo chave de longa duração; prefira personificação ou federação.

⚠ **Alerta de prova:** Tornar o repositório privado depois que a chave vazou.
Como não cair: A chave já pode ter sido copiada; desative ou exclua a chave primeiro.

⚠ **Alerta de prova:** Achar que allAuthenticatedUsers limita o acesso aos funcionários.
Como não cair: Inclui qualquer conta Google autenticada, de fora também.

⚠ **Alerta de prova:** Dar papel à conta de serviço dos nós para um pod acessar um bucket.
Como não cair: Daria acesso a todos os pods dos nós; use federação de identidade para o GKE.

⚠ **Alerta de prova:** Remover papéis de contas gerenciadas pelo Google por parecerem desnecessárias.
Como não cair: São agentes de serviço usados pelos serviços; remover causa falhas.

⚠ **Alerta de prova:** Excluir a conta de serviço suspeita de uso indevido.
Como não cair: Perde a configuração; desativar bloqueia de forma reversível.

⚠ **Alerta de prova:** Usar papel personalizado do projeto A no projeto B.
Como não cair: Papel de projeto só vale nele; para reutilizar, crie na organização.

⚠ **Alerta de prova:** Achar que chaves gerenciadas pelo usuário expiram sozinhas ou são giradas pelo Google.
Como não cair: Por padrão não expiram e o usuário as gerencia; só as do Google são giradas automaticamente.

⚠ **Alerta de prova:** Interpretar erro logo após a concessão como papel que precisa de aprovação.
Como não cair: É propagação: mudanças de IAM costumam valer em cerca de dois minutos, podendo levar mais.

Como cai na prova

📝 **Como cai:** Qual papel conceder para uma tarefa restrita (criar contas, gerenciar chaves, auditar, navegar, invocar).
Como resolver: Descarte primeiro proprietário, editor e administrador da organização. Entre os restantes, escolha o que faz exatamente o verbo pedido e nada além.

📝 **Como cai:** Onde conceder o papel (organização, pasta, projeto ou recurso).
Como resolver: Use o menor escopo que cobre o requisito: um bucket = política do bucket; uma conta de serviço = política da conta; todos os projetos Dev = pasta Dev; nunca organização por padrão.

📝 **Como cai:** Como autenticar sem chave de conta de serviço.
Como resolver: Origem fora do Google Cloud (CI, AWS, GitHub) = federação de identidade da carga de trabalho; pod no GKE = federação para o GKE; pessoa ou script no Google = personificação com Token Creator.

📝 **Como cai:** Usuário mantém ou não recebe acesso de forma inesperada.
Como resolver: Mantém acesso após remoção = herança de nível superior. Não recebe logo após concessão = propagação. Quer saber o motivo = Policy Troubleshooter.

📝 **Como cai:** Enunciado de Escolha DUAS sobre boas práticas de IAM e contas de serviço.
Como resolver: Marque as que reduzem escopo e duração: papéis no menor recurso, predefinidos em vez de básicos, só papéis necessários, credenciais de curta duração.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Descobrir qual permissão falta e por quê	**Policy Troubleshooter**
Usuários com permissões que nunca usam	**Recomendações de papéis do IAM**
Acesso só até o fim do contrato	**Condição do IAM com data de expiração**
Ninguém exclui produção, nem proprietário	**Política de negação do IAM**
CI em outro provedor sem chaves	**Federação de identidade da carga de trabalho**
Só um repositório do GitHub pode autenticar	**Condição de atributo no provedor do pool**
Pods com permissões próprias sem a conta dos nós	**Federação de identidade da carga de trabalho para o GKE**
Agir como conta de serviço sem baixar chave	**Personificação com Token Creator**
Token OAuth curto de conta de serviço para script	**generateAccessToken da Service Account Credentials API**
Serviço Cloud Run chamado só por outro serviço	**run.invoker para a conta do chamador**
Controle só pelo IAM, sem ACLs	**Acesso uniforme no nível do bucket**
Auditar IAM da organização sem alterar	**roles/iam.securityReviewer**
Ver principais com papéis no projeto pela CLI	**gcloud projects get-iam-policy**
Conceder papel a grupo pela CLI	**gcloud projects add-iam-policy-binding**

🎯 Se você lembrar de só uma coisa deste capítulo...

- Para criar VM com conta de serviço existente, o desenvolvedor precisa de serviceAccountUser na conta, além dos papéis do Compute Engine.
- Políticas de negação prevalecem sobre concessões.
- Conta de serviço padrão do Compute Engine tem editor; troque por conta dedicada com papéis mínimos.
- Chave vazada: desativar ou excluir primeiro, investigar nos registros depois.

Essenciais para a prova (checklist)

- ☐ Sei os elementos de uma vinculação e quais principais podem receber papéis.
- ☐ Sei diferenciar papéis básicos, predefinidos e personalizados, e o escopo de um papel personalizado.
- ☐ Sei explicar herança e por que remover no projeto não tira acesso dado na pasta.
- ☐ Sei usar condições do IAM e políticas de negação.
- ☐ Sei qual papel de conta de serviço cria, administra, usa, personifica e gerencia chaves.
- ☐ Sei evitar chaves com personificação, generateAccessToken e federação de identidade.
- ☐ Sei configurar a federação de identidade para o GKE e condições de atributo em pools.
- ☐ Sei controlar acesso ao Cloud Run com run.invoker e --allow-unauthenticated.
- ☐ Sei a diferença entre allUsers e allAuthenticatedUsers e o efeito do acesso uniforme no bucket.
- ☐ Sei os comandos gcloud de leitura e concessão de políticas de IAM.

Questões de revisão do capítulo

R1.1 · Configurar Acesso e Segurança

Uma empresa configurou a federação de identidade da carga de trabalho para pipelines do GitHub Actions e quer que só os workflows de um repositório específico consigam autenticar. O que configurar no provedor do pool?

A. Rótulo com o nome do repositório no projeto
B. Chave JSON de conta de serviço por repositório
C. Condição de atributo com o nome do repositório
D. Regra de firewall liberando os IPs do GitHub

R1.2 · Configurar Acesso e Segurança

Um usuário tem o papel de editor concedido em uma pasta. O administrador remove o papel dele no projeto filho, mas o usuário continua editando. Por quê?

A. A remoção leva um mês para valer
B. O papel de editor não pode ser removido
C. O projeto ignora a política da pasta
D. A permissão é herdada da pasta

R1.3 · Configurar Acesso e Segurança

Um desenvolvedor deve anexar às VMs que cria somente a conta de serviço sa-relatorios, e não as demais contas de serviço do projeto. Onde conceder o papel de usuário da conta de serviço?

A. Na política do projeto que contém as contas
B. Na política da própria conta sa-relatorios
C. Na política da pasta que contém o projeto
D. Na política da organização inteira

R1.4 · Configurar Acesso e Segurança

Um administrador encontrou contas de serviço gerenciadas pelo Google com papéis no projeto e quer removê-las por parecerem desnecessárias. O que considerar?

A. Elas não têm efeito nas permissões
B. Elas são contas de usuários externos
C. Elas foram criadas por engano no projeto
D. Elas são agentes usados pelos serviços

Respostas comentadas

R1.1. C) Condição de atributo com o nome do repositório

Por quê: Provedores de pools de identidade da carga de trabalho usam mapeamentos e condições de atributo para limitar quais identidades externas podem obter credenciais.

✗ A) Incorreta. Rótulos organizam recursos, mas não restringem quem autentica.
✗ B) Incorreta. Reintroduz chaves de longa duração, que a federação busca evitar.
✗ D) Incorreta. A autenticação federada não é controlada por regras de firewall.

> **Dica de prova:** Restringir quem usa a federação: condição de atributo.

R1.2. D) A permissão é herdada da pasta

Por quê: A política efetiva é a união da política do recurso com as herdadas; é preciso remover o papel na pasta.

✗ A) Incorreta. Mudanças de IAM valem em minutos.
✗ B) Incorreta. Papéis básicos podem ser removidos.
✗ C) Incorreta. Projetos herdam as políticas da pasta.

> **Dica de prova:** Acesso que persiste: herança da hierarquia.

R1.3. B) Na política da própria conta sa-relatorios

Por quê: Conceder roles/iam.serviceAccountUser na conta de serviço, e não no projeto, segue o menor privilégio ao permitir anexar ou usar só aquela conta.

✗ A) Incorreta. No projeto, o papel valeria para todas as contas de serviço dele.
✗ C) Incorreta. Na pasta, o papel valeria para contas de vários projetos.
✗ D) Incorreta. Na organização, o acesso seria amplo demais.

> **Dica de prova:** Usar só uma conta de serviço: papel na própria conta.

R1.4. D) Elas são agentes usados pelos serviços

Por quê: Agentes de serviço permitem que serviços do Google ajam no projeto; remover papéis deles causa falhas.

✗ A) Incorreta. Têm papéis usados pelos serviços.

✗ B) Incorreta. São contas gerenciadas pelo Google.

✗ C) Incorreta. São criadas pelos próprios serviços do Google.

> **Dica de prova:** Contas gerenciadas pelo Google: agentes de serviço.

02

CAPÍTULO 2 · 20% DA PROVA

Configurar o Ambiente da Solução em Nuvem

Aqui eu preciso dominar a fundação: hierarquia de recursos (organização, pastas, projetos), identidades corporativas (Cloud Identity, GCDS, federação), políticas da organização, faturamento, cotas e APIs. A prova traz cenários de governança e pergunta qual controle central resolve, quase sempre contrastando política da organização com rótulo, firewall ou orçamento. Também cobra o papel exato para cada tarefa de faturamento e organização, com o menor privilégio. Onde mais erro: achar que orçamento bloqueia gasto, confundir rótulo com tag do Resource Manager e esquecer que o ID do projeto é imutável. Domínio com peso de 20% no guia vigente.

Neste capítulo

18 tópicos-chave · 14 pegadinhas · 7 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Hierarquia de recursos

Organização, pastas, projetos e recursos, com herança de IAM e de políticas de cima para baixo. Projetos são a fronteira de isolamento; pastas agrupam projetos que devem herdar as mesmas permissões e políticas, como separar Dev e Prod.

Nó de organização

É associado ao domínio do Google Workspace ou do Cloud Identity; não vem do faturamento, de um projeto raiz nem de habilitar API. Uma organização representa a empresa inteira.

Organização independente (standalone)

Criada automaticamente para novos clientes do teste gratuito sem Cloud Identity. Dispensa domínio verificado, permite vários proprietários e identidades federadas, e já habilita pastas e políticas da organização.

Projeto: nome, ID e número

O nome do projeto pode mudar; o ID e o número são imutáveis, nem trocando faturamento ou pasta. Projeto excluído entra em encerramento e pode ser restaurado por cerca de 30 dias, e o ID continua reservado nesse período.

Mover projeto entre pastas

Projetos podem ser movidos entre pastas e passam a herdar as políticas do novo pai. Não é preciso recriar o projeto nem copiar papéis manualmente.

Política da organização (restrições)

Restringe configurações em toda a hierarquia, de forma preventiva. Pode ser substituída em níveis inferiores, como um único projeto, quando permitido, sem afrouxar o resto. Muitas restrições têm versão gerenciada (managed) e legada.

Restrições mais cobradas

Locais de recursos (gcp.resourceLocations); IP externo em VMs (compute.vmExternalIpAccess); criação de chaves de conta de serviço (iam.disableServiceAccountKeyCreation); domínios permitidos (iam.allowedPolicyMemberDomains); imagens confiáveis (compute.trustedImageProjects); IP público do Cloud SQL (sql.restrictPublicIp); Shielded VM (compute.requireShieldedVm); pular rede default (compute.skipDefaultNetworkCreation).

Cloud Identity e GCDS

Cloud Identity é o provedor de identidade que cria a organização e gerencia usuários e grupos sem precisar de Google Workspace. O Google Cloud Directory Sync sincroniza usuários e grupos do Active Directory ou LDAP para o Cloud Identity.

Federação com Microsoft Entra ID

O provisionamento mantém usuários e grupos criados e removidos automaticamente no Cloud Identity, e o SSO por SAML autentica no Entra ID. Verificação em duas etapas e outras políticas de autenticação ficam no Admin Console.

Federação de identidade de colaboradores

Permite que funcionários de IdPs externos (Okta, Entra ID) acessem o Google Cloud via OIDC ou SAML 2.0 sem contas sincronizadas no Cloud Identity. O pool é criado no nível da organização, não em projeto, pasta ou faturamento.

Rótulos x tags do Resource Manager x tags de rede

Rótulos (labels) são pares chave-valor exportados com o faturamento, para alocar custos; não bloqueiam nada. Tags do Resource Manager são herdadas e usadas em condições do IAM e em políticas da organização. Tags de rede servem a regras de firewall.

Conta de faturamento e papéis

Projetos são vinculados à conta de faturamento de quem paga; departamentos com faturas separadas têm contas próprias. Administrator gerencia a conta e pagamentos; User só vincula projetos; Viewer só vê custos e transações.

Orçamentos

Orçamentos somente de alertas avisam ao atingir limites (por exemplo 50%, 90% e 100%) e não bloqueiam gastos. Podem ser filtrados por projetos, serviços, rótulos e subcontas e publicar notificações no Pub/Sub para acionar automações. Tetos de gastos, em prévia, valem só para alguns serviços, como Cloud Run e Agent Platform.

Exportação do faturamento ao BigQuery

Permite análise detalhada de custos com SQL, cruzando com outros dados. Não traz o histórico completo: datasets multirregionais recebem dados desde o início do mês anterior, e regionais só a partir da ativação; por isso convém ativá-la cedo.

Cotas

Limitam uso de recursos por projeto e região; aumentos são pedidos na página de cotas do console. Erro 429 indica cota de taxa esgotada (reduzir a taxa ou pedir aumento). gcloud compute regions describe mostra limite e uso; o Cloud Monitoring expõe métricas de cota para alertas.

APIs do projeto

Muitos serviços exigem habilitar a API no projeto (gcloud services enable run.googleapis.com). Desativar APIs sem uso (gcloud services disable) reduz a superfície de ataque. Para criar VM, o projeto precisa da API do Compute Engine e de faturamento vinculado.

Cloud Asset Inventory

Mantém o histórico de metadados de recursos e políticas para pesquisar e exportar o inventário da organização. Feeds enviam ao Pub/Sub notificações de mudanças em recursos, políticas de IAM e da organização. O Gemini Cloud Assist precisa da API do Cloud Asset para entender os recursos.

Escopo de métricas e contatos essenciais

Um projeto de escopo do Cloud Monitoring vê métricas de vários projetos adicionados ao escopo de métricas. Contatos essenciais definem destinatários de avisos por categoria, como segurança, faturamento e técnico.

Macetes para memorizar

💡 Hierarquia

O-P-P-R: Organização, Pastas, Projetos, Recursos. Herança desce, nunca sobe.

💡 Papéis de faturamento

Admin administra, User une (vincula), Viewer vê.

Bloquear x avisar

Política da organização proíbe; orçamento só apita.

Três tipos de marcação

Rótulo = fatura; Tag do Resource Manager = IAM e política; Tag de rede = firewall.

Projeto

Nome muda, ID gruda.

Códigos HTTP

403 é proibido (permissão); 429 é demais (cota de taxa).

Identidade de pessoas

Sincronizar contas = GCDS ou provisionamento; sem contas = federação de colaboradores.

Tabela de memorização: não confunda

Quando é...	...e quando é
Política da organização Precisa impedir uma configuração em toda a hierarquia (local, IP externo, chaves, domínios, imagens).	**Rótulo, firewall ou orçamento** Rótulo só classifica custos, firewall só filtra tráfego, orçamento só avisa; nenhum impede criar recursos.
Usuário da conta de faturamento Só vincular projetos à conta, sem alterá-la.	**Administrador da conta de faturamento** Gerenciar a conta, trocar forma de pagamento e permissões.
Cloud Identity com GCDS ou provisionamento A empresa quer as contas criadas e sincronizadas no Cloud Identity (a partir do AD/LDAP ou Entra ID).	**Federação de identidade de colaboradores** A empresa não quer criar contas no Cloud Identity; usuários entram direto pelo IdP via OIDC ou SAML 2.0.
Rótulos (labels) Separar custos por equipe e ambiente nos relatórios de faturamento.	**Tags do Resource Manager** Condições do IAM e políticas da organização baseadas em marcação herdada, como produção.
Orçamento com alerta por e-mail Avisar pessoas quando o gasto atingir os limites.	**Notificação do orçamento no Pub/Sub** Acionar automação programática, como desligar recursos de teste ao atingir 100%.
Relatórios do Cloud Billing Ver sem configurar nada quais serviços, projetos e rótulos mais custaram.	**Exportação do faturamento ao BigQuery** Análise detalhada com SQL cruzando com outros dados.
Criador de projetos na pasta A equipe só pode criar projetos dentro daquela pasta.	**Criador de projetos na organização** Permite criar em qualquer lugar da organização; amplo demais quando o escopo é uma pasta.

Números que caem

Item	Valor
Peso do domínio no guia vigente	**20%**
Período para restaurar projeto excluído	**cerca de 30 dias**
Exportação do faturamento, dataset multirregional	**dados desde o início do mês anterior**
Restrição legada de criação de chaves já aplicada	**organizações criadas a partir de maio de 2024**
Erro de cota de taxa	**429 (falta de permissão é 403)**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Achar que um orçamento de R$ 10 mil impede o projeto de gastar mais que isso ou desliga os recursos.
Como não cair: Orçamento somente de alertas avisa e não bloqueia. Para VMs, o controle exige automação via notificações no Pub/Sub.

⚠ **Alerta de prova:** Usar rótulo de proibição para impedir IP externo, chaves de conta de serviço ou locais fora da UE.
Como não cair: Rótulos não bloqueiam ações. O controle preventivo é restrição de política da organização.

⚠ **Alerta de prova:** Bloquear IP externo com regra de firewall de saída.
Como não cair: Firewall não impede criar o IP externo; use compute.vmExternalIpAccess na política da organização.

⚠ **Alerta de prova:** Remover a política da organização toda ou tirar o projeto da organização para atender uma exceção.
Como não cair: Substitua a política só naquele projeto; o resto continua protegido e herdando governança.

⚠ **Alerta de prova:** Acreditar que o ID do projeto muda ao trocar de pasta ou de faturamento, ou que dá para editá-lo.
Como não cair: Só o nome muda. ID e número são imutáveis.

⚠ **Alerta de prova:** Achar que projeto excluído some na hora ou que dá para recriar com o mesmo ID.
Como não cair: Há período de recuperação de cerca de 30 dias e o ID continua reservado.

⚠ **Alerta de prova:** Separar faturas de dois departamentos com rótulos em uma única conta.
Como não cair: Rótulos geram uma fatura só. Faturas separadas exigem uma conta de faturamento por departamento.

⚠ **Alerta de prova:** Interpretar erro 429 como falta de permissão.
Como não cair: Falta de permissão gera 403. O 429 é cota de taxa esgotada: reduzir a taxa ou pedir aumento.

⚠ **Alerta de prova:** Contornar limite de GPUs criando projeto novo ou mudando de pasta.
Como não cair: Todo projeto tem cotas e a hierarquia não altera cotas regionais. Peça aumento de cota na região.

⚠ **Alerta de prova:** Esperar o histórico do ano passado logo após ativar a exportação do faturamento.
Como não cair: A exportação não recupera histórico antigo; ative-a cedo.

⚠ **Alerta de prova:** Exigir verificação em duas etapas pela política de IAM do projeto.
Como não cair: Políticas de autenticação ficam no Admin Console do Cloud Identity.

⚠ **Alerta de prova:** Dar contas de serviço ou chaves de API para funcionários.
Como não cair: Contas de serviço são para cargas de trabalho e chaves de API não representam usuários. Pessoas usam Cloud Identity ou federação.

⚠ **Alerta de prova:** Continuar usando o superadministrador nas tarefas diárias.
Como não cair: Reserve o superadministrador e conceda administrador da organização a outras contas.

⚠ **Alerta de prova:** Criar pastas dentro de um projeto ou comprar domínio só para ter pastas no teste gratuito.
Como não cair: Pastas ficam sob a organização ou outras pastas; a organização independente dispensa domínio e Cloud Identity.

Como cai na prova

📝 **Como cai:** A segurança quer impedir X em todos os projetos da organização. Qual controle usar?
Como resolver: Procure a alternativa com restrição de política da organização. Descarte rótulo, firewall, orçamento e papel de leitor: nenhum impede a criação.

📝 **Como cai:** Uma pessoa precisa fazer só uma tarefa de faturamento ou organização. Qual papel conceder?
Como resolver: Identifique o verbo: vincular (Billing Account User), gerenciar pagamento (Administrator), ver custos (Viewer), criar projetos na pasta (projectCreator na pasta), políticas da organização (orgpolicy.policyAdmin), IAM da organização (organizationAdmin). Descarte proprietário e editor.

📝 **Como cai:** Cenário de custos: separar, avisar, automatizar ou analisar.
Como resolver: Separar por equipe = rótulos; faturas separadas = contas de faturamento; avisar = orçamento; agir = Pub/Sub; analisar com SQL = exportação ao BigQuery; ver rápido = relatórios do Billing.

📝 **Como cai:** Criação de recurso falha em projeto novo ou em pico.
Como resolver: Serviço não habilitado = gcloud services enable; VM não cria = API do Compute + faturamento vinculado; limite de recursos = aumento de cota na região; 429 = cota de taxa; GPU não aparece = oferta varia por zona.

📝 **Como cai:** Empresa com IdP próprio (AD, Entra ID, Okta) começando no Google Cloud.
Como resolver: AD/LDAP com contas no Google = GCDS; Entra ID com contas no Google = provisionamento + SSO SAML; sem criar contas = federação de identidade de colaboradores com pool na organização.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Impedir em toda a organização, de forma central	**Restrição de política da organização**
Nenhum recurso fora das regiões da UE	**gcp.resourceLocations**
Projetos novos sem rede default	**compute.skipDefaultNetworkCreation**
Só imagens de projetos aprovados	**Restrição de imagens confiáveis (trustedImageProjects)**
Separar custos por equipe no mesmo projeto	**Rótulos nos recursos**
Automação ao atingir 100% do orçamento	**Notificações do orçamento no Pub/Sub**
Alerta só de BigQuery em alguns projetos	**Orçamento filtrado por projeto e serviço**
Custos com SQL, cruzando outros dados	**Exportação do faturamento para o BigQuery**
Usuários do Active Directory local no Cloud Identity	**Google Cloud Directory Sync**

Funcionários no IdP externo, sem criar contas no Google	**Federação de identidade de colaboradores, pool na organização**
Notificação em tempo real de mudança de IAM ou firewall	**Feed do Cloud Asset Inventory para o Pub/Sub**
Avisos de segurança e faturamento chegando à pessoa errada	**Contatos essenciais**
Métricas de dez projetos em um só lugar	**Escopo de métricas do Cloud Monitoring**
Erro 429 em horário de pico	**Cota de taxa da API; pedir aumento ou reduzir taxa**

🎯 **Se você lembrar de só uma coisa deste capítulo...**

- Orçamento somente de alertas nunca bloqueia gasto; desvincular o faturamento é a medida drástica que para a cobrança.
- Rótulo não impõe nada: governança preventiva é política da organização, que pode ser substituída só no projeto que precisa da exceção.
- ID e número do projeto são imutáveis; o nome pode mudar.
- Conceda papéis a grupos, no menor escopo (pasta em vez de organização), e reserve o superadministrador.

Essenciais para a prova (checklist)

- ☐ Sei montar a hierarquia organização, pastas e projetos para isolar Dev e Prod com herança.
- ☐ Sei de onde vem o nó de organização e quando existe organização independente.
- ☐ Sei associar cada restrição de política da organização ao problema que ela resolve.
- ☐ Sei a diferença entre rótulos, tags do Resource Manager e tags de rede.
- ☐ Sei os papéis Admin, User e Viewer da conta de faturamento e quando usar cada um.
- ☐ Sei que orçamentos só alertam e como automatizar a reação pelo Pub/Sub.
- ☐ Sei as limitações de histórico da exportação do faturamento ao BigQuery.
- ☐ Sei diagnosticar API não habilitada, cota regional e erro 429.
- ☐ Sei escolher entre GCDS, provisionamento do Entra ID e federação de identidade de colaboradores.
- ☐ Sei para que servem Cloud Asset Inventory, feeds, escopo de métricas e contatos essenciais.

Questões de revisão do capítulo

R2.1 · Configurar o Ambiente da Solução em Nuvem

Um arquiteto da segurança deve definir restrições de políticas da organização, sem receber acesso aos recursos dos projetos. Qual papel conceder?

A. Administrador de políticas da organização
B. Usuário da conta de faturamento
C. Proprietário dos projetos existentes
D. Editor na raiz da organização inteira

R2.2 · Configurar o Ambiente da Solução em Nuvem

A gerência quer ser avisada por e-mail quando os gastos do mês de um projeto atingirem 50%, 90% e 100% de um valor planejado. O que configurar?

A. Orçamento com limites de alerta
B. Exportação diária para o Cloud Storage
C. Política da organização de custos
D. Cota de CPU reduzida no projeto

R2.3 · Configurar o Ambiente da Solução em Nuvem

Uma empresa usa Google Workspace com o domínio empresa.com.br e vai começar no Google Cloud. De onde vem o nó de organização?

A. É comprado junto com a conta de faturamento
B. É criado a partir do domínio do Workspace
C. É criado por um projeto chamado raiz
D. É criado ao habilitar a API do Compute

R2.4 · Configurar o Ambiente da Solução em Nuvem

Um projeto de laboratório gerou custos inesperados e deve parar de ser cobrado imediatamente, mesmo que os recursos parem. Qual ação atende?

A. Mover o projeto para outra pasta
B. Criar um orçamento com alerta de 100%
C. Desvincular o faturamento do projeto
D. Adicionar rótulos de custo aos recursos

Respostas comentadas

R2.1. A) Administrador de políticas da organização

Por quê: O papel roles/orgpolicy.policyAdmin permite gerenciar políticas da organização.

✗ B) Incorreta. Só vincula projetos à conta.

✗ C) Incorreta. Dá acesso total aos recursos.

✗ D) Incorreta. Dá acesso amplo aos recursos.

> **Dica de prova:** Definir restrições: administrador de políticas da organização.

R2.2. A) Orçamento com limites de alerta

Por quê: Orçamentos somente de alertas avisam ao atingir os limites e não bloqueiam gastos; tetos de gastos existem apenas para alguns serviços.

✗ B) Incorreta. Exporta dados, mas não alerta.

✗ C) Incorreta. Não existe esse controle de gastos.

✗ D) Incorreta. Limita recursos, não avisa sobre gastos.

> **Dica de prova:** Aviso em 50/90/100% do gasto: orçamento com alertas.

R2.3. B) É criado a partir do domínio do Workspace

Por quê: O recurso de organização é associado ao domínio do Google Workspace ou do Cloud Identity; novas contas de teste gratuito também podem receber uma organização independente.

✗ A) Incorreta. O faturamento não cria a organização.

✗ C) Incorreta. Projetos não criam a organização.

✗ D) Incorreta. APIs não criam a organização.

> **Dica de prova:** Origem da organização: domínio do Workspace ou Cloud Identity.

R2.4. C) Desvincular o faturamento do projeto

Por quê: Desativar o faturamento interrompe serviços pagos e pode levar à perda de recursos; é uma medida drástica.

✗ A) Incorreta. Não interrompe gastos.

✗ B) Incorreta. Avisa, mas não interrompe gastos.

✗ D) Incorreta. Não interrompe gastos.

> **Dica de prova:** Parar a cobrança já: desvincular o faturamento.

03

CAPÍTULO 3 · 30% DA PROVA

Garantir a Operação Bem-Sucedida da Solução em Nuvem

Este domínio (30%) é o dia a dia de quem opera: gerenciar VMs, GKE, Cloud Run, Storage e bancos, além de monitorar, registrar e depurar. A prova pergunta o recurso ou comando exato para uma ação operacional (backup, reverter versão, escalar, restaurar a um horário, coletar memória, alertar) e coloca alternativas que parecem próximas, mas fazem outra coisa. Preciso saber de cor os comandos kubectl e gcloud mais comuns e o papel de cada ferramenta de observabilidade. Onde mais erro: confundir HPA, VPA e cluster autoscaler, os tipos de registro de auditoria e os destinos de coletores de registros.

Neste capítulo

21 tópicos-chave · 14 pegadinhas · 7 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Snapshots, imagens e imagens de máquina

Snapshot é backup incremental de disco; programação de snapshots é política de recursos que cria e exclui snapshots automaticamente (ex.: diário com retenção de 14 dias). Imagem personalizada, criada de disco ou snapshot, padroniza o sistema e pode ser compartilhada entre projetos. Imagem de máquina captura propriedades, metadados, permissões e todos os discos da VM.

Alterações em VMs

Disco permanente pode ser aumentado com a VM ligada, expandindo depois o sistema de arquivos. Mudar o tipo de máquina exige parar a VM. Adicionar GPU em VM N1 exige parar a VM, definir manutenção do host como TERMINATE e instalar o driver; VMs com GPU não fazem migração em tempo real.

Custos de VM parada e IPs

VM parada não cobra CPU, memória nem GPU, mas discos e IPs estáticos reservados continuam cobrados. IP externo efêmero pode ser promovido a estático mantendo o endereço; IP interno estático pode ser reservado para não mudar ao recriar a VM.

Escalonamento no GKE

HPA ajusta o número de réplicas (pods) conforme métricas como CPU. VPA analisa uso e recomenda ou aplica requests de CPU e memória; no modo Off só recomenda. Cluster autoscaler adiciona ou remove nós conforme pods pendentes e ociosidade. Reparo automático recria nós com falha.

Pools de nós e namespaces

Pools de nós permitem tipos de máquina diferentes no mesmo cluster, como nós com GPU, sem recriar o cluster. gcloud container clusters resize com --node-pool e --num-nodes muda o número de nós. Namespaces separam equipes com cotas de recursos e RBAC próprios.

Comandos kubectl de operação

rollout undo reverte o deployment; rollout status só mostra andamento; scale --replicas muda réplicas; exec -it abre shell no contêiner; logs (inclusive --previous) e describe diagnosticam pods; get nodes mostra se estão Ready; get services mostra o EXTERNAL-IP do LoadBalancer.

Diagnóstico de pods

ImagePullBackOff indica falha ao baixar a imagem: nome, tag ou permissão incorretos. CrashLoopBackOff pede kubectl logs para ver o erro da aplicação. StatefulSet com volume persistente mantém dados e identidade quando o pod é recriado em outro nó.

GKE Autopilot e requests

O Autopilot usa os requests dos pods para provisionar nós e cobrar. Sem requests, aplica valores padrão; abaixo do mínimo, ajusta; acima do máximo, rejeita a carga.

Revisões e tráfego no Cloud Run

Toda mudança de código ou configuração, inclusive variável de ambiente, cria nova revisão imutável. Divisão de tráfego permite lançamento gradual (ex.: 10%); --no-traffic implanta sem receber tráfego; tag de revisão cria URL próprio para testar sem tráfego de usuários.

Ajustes de instâncias no Cloud Run

Mínimo de instâncias evita partida a frio, com custo de instâncias ociosas. Máximo de instâncias, junto com a simultaneidade, limita a carga em sistemas dependentes (ex.: banco com 100 conexões). Aumentar a simultaneidade ajuda com E/S lenta. Cobrança baseada em instância mantém CPU para trabalho em segundo plano; na baseada em requisições, padrão, a CPU só fica alocada durante requisições.

Proteção de dados no Cloud Storage

Controle de versões guarda versões não atuais ao sobrescrever ou excluir. Ciclo de vida muda classe e apaga por idade, e numNewerVersions limita versões antigas. Política de retenção bloqueada (Bucket Lock) impede excluir ou alterar, nem por administradores. URL assinado dá acesso temporário a um objeto sem conta Google.

Acesso e criptografia de buckets

roles/storage.objectViewer para allUsers torna objetos legíveis publicamente sem permitir escrita. Com CMEK, além da chave padrão do bucket, o agente de serviço do Cloud Storage precisa de roles/cloudkms.cryptoKeyEncrypterDecrypter na chave. Realocação de buckets muda o local mantendo nome e metadados, com Storage Intelligence configurado.

Operação de bancos

Cloud SQL: recuperação a um ponto no tempo usa backups e logs binários; aumento automático de armazenamento evita disco cheio; Query Insights acha consultas lentas. Firestore PITR retém versões por 7 dias com granularidade de um minuto para ler, clonar ou exportar. Database Center centraliza a saúde da frota de bancos.

Operação do BigQuery

Simulação (dry run) informa bytes processados antes de rodar; LIMIT não reduz bytes lidos. Particionamento e clustering reduzem custo de consultas filtradas por data. bq ls -j lista jobs e bq show -j mostra status e erros.

Cloud Monitoring

Ops Agent coleta memória, disco e registros de aplicações nas VMs. Verificação de tempo de atividade testa endpoints de vários locais. Política de alerta usa condições sobre métricas e canais de notificação (e-mail, SMS, Slack, PagerDuty, webhooks, Pub/Sub). Métrica baseada em registros transforma uma mensagem em métrica alertável; métricas personalizadas vêm pela API ou OpenTelemetry; Managed Service for Prometheus aceita PromQL.

Cloud Logging e coletores

Coletores (sinks) roteiam a BigQuery (SQL), Cloud Storage (arquivo barato), Pub/Sub (tempo real) ou buckets de registros. Coletor agregado na organização ou pasta inclui todos os filhos, inclusive projetos futuros. Bucket _Default retém 30 dias; _Required guarda auditoria por 400 dias. Exclusões no roteador e retenção menor reduzem custo; Log Analytics permite SQL sem exportar.

Registros de auditoria

Atividade do administrador e eventos do sistema são sempre gravados e não podem ser desativados. Acesso a dados (quem leu) fica desativado por padrão na maioria dos serviços, como Cloud Storage; o BigQuery é a exceção. VPC Flow Logs e registro de regras de firewall não são registros de auditoria.

Depuração e rede operacional

Cloud Trace mede latência entre serviços; Cloud Profiler mostra CPU e memória por função com baixo overhead; saída da porta serial mostra mensagens de boot. IAP para SSH sem IP externo exige liberar a porta 22 vinda de 35.235.240.0/20. Cloud NAT dá saída sem IP externo; esgotamento pede mais portas por VM ou mais IPs de NAT.

Visões operacionais e otimização

Personalized Service Health mostra incidentes do Google que afetam seus projetos; Cloud Hub reúne alertas, incidentes, manutenções e cotas; investigação do Gemini Cloud Assist gera hipóteses de causa raiz a partir do alerta. Active Assist recomenda parar VMs ociosas; a calculadora de preços estima custos futuros.

Cloud DNS e rotas estáticas

Zona pública do Cloud DNS para domínio registrado em outro provedor exige apontar os servidores de nome no registrador. Zona privada é visível apenas nas redes VPC associadas. Rota estática com uma VM como próximo salto força o tráfego por um appliance; essa VM precisa ter o encaminhamento de IP habilitado.

Artifact Registry e Cloud Workstations

O Artifact Registry guarda imagens de contêiner e pacotes com acesso controlado pelo IAM, usado pelo GKE e pelo Cloud Run. O Cloud Workstations oferece ambientes de desenvolvimento gerenciados e padronizados por configurações, acessados pelo navegador ou pela IDE local, com o código na nuvem.

Macetes para memorizar

💡 Escalonadores do GKE

Horizontal = mais pods; Vertical = pods maiores (Off só recomenda); Cluster = mais nós; Reparo = nó doente recriado.

💡 Auditoria sempre ativa

Admin e Sistema nunca dormem; Dados precisa acordar (exceto BigQuery).

💡 Destinos de coletores

BigQuery pergunta (SQL), Storage guarda (barato), Pub/Sub avisa (tempo real).

💡 Retenção de logs

_Default 30, _Required 400.

Trace x Profiler

Trace segue a requisição entre serviços; Profiler olha dentro da função.

Cloud Run

Mínimo esquenta, máximo protege, simultaneidade economiza, cobrança por instância trabalha em segundo plano.

Mudanças em VM

Disco cresce ligado; tipo de máquina e GPU só com a VM parada.

Tabela de memorização: não confunda

Quando é...	...e quando é
Escalonador horizontal de pods (HPA) Aumentar ou reduzir réplicas conforme CPU.	**Escalonador automático do cluster** Pods pendentes por falta de CPU nos nós: adiciona nós.
Escalonamento vertical de pods (VPA) modo Off Recomendar requests de CPU e memória pelo uso real, sem alterar os pods.	**HPA** Mudar o número de réplicas; não recomenda requests.
Cloud Trace Descobrir qual serviço downstream acrescenta latência na requisição.	**Cloud Profiler** Descobrir quais funções do código gastam mais CPU e memória.
Coletor para BigQuery Guardar registros e consultar com SQL (ex.: auditoria por cinco anos).	**Coletor para Cloud Storage** Arquivar barato por muitos anos sem consultas frequentes.
Snapshot do disco Backup de um disco antes de mudança arriscada.	**Imagem de máquina** Cópia completa da VM com configuração, metadados e todos os discos.
Divisão de tráfego entre revisões Enviar uma fração do tráfego real (ex.: 10%) à nova revisão.	**Tag de tráfego ou --no-traffic** Implantar e testar a revisão sem enviar tráfego de usuários (tag dá URL próprio).
Registros de atividade do administrador Quem alterou configurações; sempre ativos.	**Registros de acesso a dados** Quem leu os dados; precisam ser habilitados na maioria dos serviços.

Números que caem

Item	Valor
Peso do domínio no guia vigente	**30%**

Retenção padrão do bucket _Default	**30 dias**
Retenção dos registros de auditoria no _Required	**400 dias**
Intervalo do encaminhamento TCP do IAP	**35.235.240.0/20 (porta 22 para SSH)**
Retenção da PITR do Firestore	**7 dias, granularidade de um minuto**
Filtro de VMs no Explorador de registros	**resource.type="gce_instance"**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Usar réplica de leitura ou exportação da semana passada para desfazer exclusão acidental no Cloud SQL.
Como não cair: A réplica replica o erro e a exportação perde dados recentes; use recuperação a um ponto no tempo.

⚠ **Alerta de prova:** Achar que VM parada não gera custo algum.
Como não cair: Discos e IPs estáticos reservados continuam cobrados.

⚠ **Alerta de prova:** Usar kubectl scale deployment para aumentar nós do pool.
Como não cair: scale muda réplicas; nós do pool mudam com gcloud container clusters resize.

⚠ **Alerta de prova:** Usar HPA quando os pods estão pendentes por falta de CPU nos nós.
Como não cair: HPA cria mais pods, não nós. Habilite o cluster autoscaler.

⚠ **Alerta de prova:** Habilitar atividade do administrador para saber quem leu objetos.
Como não cair: Ela registra alterações de configuração; leitura exige registros de acesso a dados.

⚠ **Alerta de prova:** Marcar acesso a dados como sempre ativo.
Como não cair: Sempre ativos são só atividade do administrador e eventos do sistema.

⚠ **Alerta de prova:** Colocar LIMIT na consulta do BigQuery para reduzir bytes lidos.
Como não cair: LIMIT não reduz a varredura; use particionamento, clustering e dry run.

⚠ **Alerta de prova:** Proteger por sete anos com regra de ciclo de vida, classe Archive ou versionamento.
Como não cair: Ciclo de vida exclui, Archive não impede exclusão e versões podem ser apagadas; use política de retenção bloqueada.

⚠ **Alerta de prova:** Pausar o versionamento ou mudar para Archive para conter custo de versões antigas.
Como não cair: Crie regra de ciclo de vida com numNewerVersions.

⚠ **Alerta de prova:** Tornar o bucket público por uma hora ou enviar a chave da conta de serviço ao parceiro.
Como não cair: URL assinado com validade curta dá acesso só ao objeto, sem conta Google.

⚠ **Alerta de prova:** Conceder Administrador de objetos a allUsers para um site estático.
Como não cair: Isso permite escrita pública; use Leitor de objetos (objectViewer) para allUsers.

⚠ **Alerta de prova:** Reduzir o máximo de instâncias ou trocar de região para evitar partida a frio no Cloud Run.
Como não cair: Partida a frio se resolve com mínimo de instâncias; máximo serve para limitar carga em dependências.

⚠ **Alerta de prova:** Mínimo de instâncias igual a zero para manter trabalho em segundo plano no Cloud Run.
Como não cair: Use cobrança baseada em instância, que mantém CPU fora das requisições.

⚠ **Alerta de prova:** Criar coletor separado em cada projeto para centralizar auditoria.
Como não cair: Esquece projetos novos; use coletor agregado na organização.

Como cai na prova

📝 **Como cai:** Problema operacional e quatro recursos do mesmo produto (Cloud Run: mínimo, máximo, simultaneidade, cobrança).
Como resolver: Identifique o sintoma: lentidão na primeira requisição (mínimo), banco saturado (máximo), muitas instâncias por E/S lenta (simultaneidade), trabalho após a resposta (cobrança baseada em instância).

📝 **Como cai:** Qual comando kubectl ou gcloud resolve a ação.
Como resolver: Associe verbo a comando: reverter = rollout undo; réplicas = scale; nós = clusters resize; shell = exec -it; erro do app = logs; prontidão = get nodes; IP externo = get services; status de job = bq show -j.

📝 **Como cai:** Qual ferramenta de observabilidade usar.
Como resolver: Métrica de VM ausente = Ops Agent; disponibilidade externa = uptime check; mensagem no log = métrica baseada em registros; latência entre serviços = Trace; CPU por função = Profiler; tráfego = VPC Flow Logs; conexões por regra = registro de regras de firewall.

📝 **Como cai:** Recuperação de dados após erro humano.
Como resolver: Cloud SQL ou Firestore a um horário = PITR; arquivo sobrescrito no bucket = controle de versões; disco antes de mudança = snapshot; VM inteira = imagem de máquina.

📝 **Como cai:** Retenção e custo de registros.
Como resolver: Muito tempo com SQL = sink BigQuery; arquivo barato = sink Storage; centralizar organização = coletor agregado; reduzir custo = exclusões e retenção menor; nunca aumente nível de detalhe ou ative acesso a dados em tudo.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Voltar ao estado exato das 14h05	**Recuperação a um ponto no tempo**
Backup diário automático de discos com retenção	**Programação de snapshots**
Memória da VM não aparece no Monitoring	**Ops Agent**
Site fora do ar visto de vários lugares do mundo	**Verificação de tempo de atividade**
Alerta quando a mensagem X aparecer nos registros	**Métrica baseada em registros com alerta**
SQL sobre registros sem exportar	**Log Analytics em bucket atualizado**
SSH por IAP falhando	**Entrada na porta 22 vinda de 35.235.240.0/20**
Falhas de saída em picos por falta de portas	**Mais portas por VM ou mais IPs no Cloud NAT**
VM não termina o boot e não aceita SSH	**Saída da porta serial**
Nova versão com falhas no GKE	**kubectl rollout undo**
Bloquear IPs em todas as VPCs sem sobreposição	**Política de firewall hierárquica na organização**
Incidente do próprio Google afetando seus projetos	**Personalized Service Health**
VMs ligadas sem uso em vários projetos	**Recomendações de VMs ociosas do Active Assist**
Sub-rede sem IPs livres	**Expandir o intervalo principal da sub-rede**

🎯 Se você lembrar de só uma coisa deste capítulo...

- Mudar variável de ambiente no Cloud Run cria nova revisão imutável.
- CMEK no bucket exige dar ao agente de serviço do Cloud Storage o papel cryptoKeyEncrypterDecrypter na chave.
- Política de firewall hierárquica na organização não é sobreposta por regras de projeto (a ação goto_next é que delega a avaliação aos níveis abaixo); política de negação do IAM controla permissões, não tráfego.
- Adicionar GPU ou mudar tipo de máquina exige VM parada; no GKE, um pool de nós novo com GPU dispensa recriar o cluster.
- Workbench desliga por inatividade, mas o disco continua cobrado.

Essenciais para a prova (checklist)

☐ Sei a diferença entre snapshot, programação de snapshots, imagem personalizada e imagem de máquina.

☐ Sei o que exige parar a VM e o que continua cobrado com ela parada.

☐ Sei quando usar HPA, VPA, cluster autoscaler, reparo automático e pool de nós novo.

☐ Sei os comandos kubectl de reverter, escalar, depurar e consultar nós e serviços.

☐ Sei gerenciar revisões, divisão de tráfego, tags e limites de instâncias no Cloud Run.

☐ Sei aplicar versionamento, ciclo de vida, retenção bloqueada e URL assinado no Cloud Storage.

☐ Sei usar PITR, aumento automático de armazenamento e Query Insights no Cloud SQL.

☐ Sei configurar Ops Agent, uptime checks, políticas de alerta, canais de notificação e métricas baseadas em registros.

☐ Sei escolher o destino do coletor de registros e quando usar coletor agregado.

☐ Sei quais registros de auditoria são sempre ativos e quais precisam ser habilitados.

Questões de revisão do capítulo

R3.1 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

VMs sem IP externo precisam baixar atualizações da internet, sem aceitar conexões de entrada. O que configurar?

A. Balanceador de aplicativo externo
B. IP externo efêmero em cada VM
C. Zona privada do Cloud DNS
D. Cloud NAT na região das VMs

R3.2 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

A segurança precisa saber quantas conexões cada regra de firewall permitiu ou negou. O que habilitar?

A. Exportação do faturamento ao BigQuery
B. Rótulos nas regras
C. Relatórios de cotas
D. Registro das regras de firewall

R3.3 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

Um serviço do Cloud Run tem lentidão na primeira requisição após períodos sem uso. Como reduzir esse atraso de inicialização?

A. Excluir as revisões mais antigas
B. Trocar a região do serviço atual
C. Definir um mínimo de instâncias ativas
D. Reduzir o limite máximo de instâncias

R3.4 · Garantir a Operação Bem-Sucedida da Solução em Nuvem

A equipe quer saber quando o site público da empresa deixa de responder de diferentes lugares do mundo. O que configurar?

A. Relatório do Cloud Billing
B. Verificação de tempo de atividade
C. Coletor de registros do projeto
D. Painel de uso de memória

Respostas comentadas

R3.1. D) Cloud NAT na região das VMs

Por quê: O Cloud NAT permite que recursos sem IP externo iniciem conexões de saída.

✗ A) Incorreta. Serve para tráfego de entrada.
✗ B) Incorreta. Expõe as VMs à internet.
✗ C) Incorreta. Não fornece saída.

> **Dica de prova:** Saída sem IP externo: Cloud NAT.

R3.2. D) Registro das regras de firewall

Por quê: O Firewall Rules Logging registra conexões permitidas ou negadas por regra.

✗ A) Incorreta. Não registra conexões.

✗ B) Incorreta. Não registram conexões.

✗ C) Incorreta. Não registram conexões.

> **Dica de prova:** Auditar regras de firewall: registro das regras.

R3.3. C) Definir um mínimo de instâncias ativas

Por quê: Instâncias mínimas evitam partidas a frio, com custo de manter instâncias ociosas.

✗ A) Incorreta. Não afeta a inicialização.

✗ B) Incorreta. Não evita a partida a frio.

✗ D) Incorreta. Não evita a partida a frio.

> **Dica de prova:** Partida a frio no Cloud Run: instâncias mínimas.

R3.4. B) Verificação de tempo de atividade

Por quê: Verificações de tempo de atividade testam endpoints de vários locais e podem gerar alertas.

✗ A) Incorreta. Mostra custos.

✗ C) Incorreta. Não testa o site.

✗ D) Incorreta. Não testa de fora.

> **Dica de prova:** Site fora do ar visto de fora: verificação de tempo de atividade.

04

CAPÍTULO 4 · 30% DA PROVA

Planejar e Implementar a Solução em Nuvem

Este é um dos domínios mais pesados (30%) e pede que eu escolha o produto certo para cada requisito: computação, armazenamento, banco, rede, balanceador, mensageria e IaC. A prova descreve um cenário com uma palavra-chave decisiva (escala até zero, NFS, consistência forte global, protocolo TCP próprio, pouca banda) e oferece produtos vizinhos como distratores. Também cobra comandos de implantação (gcloud run deploy --source, set-autoscaling, get-credentials, terraform init/import, bq load). Onde mais erro: classes de armazenamento pelo padrão de acesso, tipo de balanceador e a diferença entre Transfer Appliance e Storage Transfer Service.

Neste capítulo

20 tópicos-chave · 14 pegadinhas · 7 macetes · 14 gatilhos · 4 questões comentadas

Tópicos-chave

Escolha de computação

Cloud Run para contêiner sem estado com escala até zero; função do Cloud Run para trecho leve orientado a evento ou HTTP; jobs do Cloud Run para rodar até terminar, agendados pelo Cloud Scheduler. Compute Engine quando há controle total do sistema operacional ou módulos do kernel. Agent Runtime para agente de IA em produção com sessões e Memory Bank.

GKE Autopilot x Standard

No Autopilot o Google gerencia os nós e a cobrança é pelos recursos solicitados pelos pods; restringe cargas privilegiadas. O Standard permite gerenciar pools de nós, DaemonSets privilegiados e ajustes nos nós.

Tipos de máquina

Personalizado para combinações de vCPU e memória fora do predefinido (ex.: 6 vCPUs e 40 GB). E2 de núcleo compartilhado (e2-micro, e2-small) para cargas leves e baratas. Família otimizada para memória (M) para bancos em memória como SAP HANA. Imagens premium incluem licença de Windows e SQL Server na cobrança.

Spot VMs

Desconto alto, mas podem ser interrompidas pelo Compute Engine; as preemptivas são a versão legada. Ideais para lote tolerante a interrupção. No GKE, use pool de nós spot com taints e tolerâncias só para os lotes, sem arriscar a carga crítica.

TPU x GPU

Cloud TPU para modelos grandes dominados por multiplicação de matrizes, treinos de semanas e lotes grandes. GPUs são mais flexíveis quando há operações personalizadas ou não suportadas na TPU.

Grupos gerenciados de instâncias (MIG)

MIG regional distribui VMs entre zonas e sobrevive à falha de uma zona. Recuperação automática usa verificação de integridade da aplicação para recriar VMs travadas. set-autoscaling define mínimo, máximo e meta de CPU; resize define tamanho fixo.

Classes do Cloud Storage

Standard: dados quentes, sem duração mínima nem taxa de recuperação. Nearline: cerca de uma vez por mês, mínimo de 30 dias. Coldline: cerca de uma vez por trimestre, mínimo de 90 dias. Archive: menos de uma vez por ano, mínimo de 365 dias e acesso caro. Autoclass move objetos entre classes quando o acesso é imprevisível.

Discos e arquivos

Hyperdisk Balanced atende à maioria das cargas; Hyperdisk Extreme é para IOPS muito altas; Hyperdisk permite aumentar ou reduzir desempenho com o disco em uso. Disco regional replica de forma síncrona entre duas zonas. SSD local é muito rápido, mas efêmero. Filestore é NFS gerenciado; NetApp Volumes suporta NFS, SMB, iSCSI e NVMe/TCP; Managed Lustre é sistema de arquivos paralelo POSIX para IA.

Escolha de banco

Cloud SQL: relacional regional, escala vertical; HA usa instância em espera em outra zona; réplicas de leitura aliviam leitura. Spanner: relacional com consistência forte global e escala horizontal. AlloyDB: compatível com PostgreSQL de alto desempenho. Bigtable: NoSQL de colunas largas para séries temporais e IoT. Firestore: documentos serverless com tempo real e offline. BigQuery: data warehouse serverless. Memorystore: cache com Valkey, Redis e Memcached.

Pub/Sub e processamento

Assinatura push entrega mensagens a endpoint HTTPS, como Cloud Run; pull exige consumidor que busque. Tópico de mensagens inativas (dead-letter) recebe mensagens após o limite de tentativas. Pub/Sub com Dataflow para transformar fluxos em tempo real. Managed Service for Apache Kafka mantém compatibilidade com código Kafka sem administrar brokers.

Movimentação de dados

Transfer Appliance para grandes volumes locais com pouca banda. Storage Transfer Service para dados de outras nuvens, HTTP e ambientes locais. BigQuery Data Transfer Service agenda cargas de SaaS, outras nuvens e Cloud Storage. bq load carrega do Cloud Storage; tabelas externas e BigLake consultam o bucket sem carregar.

VPC e firewall

VPC em modo personalizado (--subnet-mode=custom) não cria sub-redes automáticas. Regras implícitas negam entrada e permitem saída, com a menor prioridade (65535). Número menor tem prioridade maior; em empate, negar vence. Alvo por conta de serviço é mais rígido que tag de rede; tags seguras são controladas pelo IAM e funcionam em políticas de firewall.

Conectividade

VPC compartilhada: projeto host fornece redes a projetos de serviço com gestão central. Peering: liga VPCs, inclusive de organizações diferentes, por IP interno. Cloud VPN de alta disponibilidade: dois túneis, SLA de 99,99%. Cloud Router usa BGP para rotas dinâmicas. Interconexão dedicada: conexão física privada de alta capacidade, sem internet.

Acesso privado a serviços

Acesso privado do Google, habilitado por sub-rede, deixa VMs sem IP externo alcançarem APIs do Google. Acesso a serviços particulares reserva intervalo para Cloud SQL só com IP privado. Saída de VPC direta ou conector de VPC sem servidor ligam Cloud Run a recursos internos.

Balanceadores

Aplicativo externo global: IP anycast para várias regiões, exige nível Premium. Aplicativo externo regional: processamento em uma região (residência de dados). Aplicativo interno regional: IP privado dentro da VPC. Rede de proxy externo: TCP e SSL com IP global. Cloud Armor adiciona DDoS e WAF; certificado gerenciado pelo Google renova sozinho.

Níveis de serviço de rede

Premium usa a rede global do Google e é exigido por balanceadores globais. Standard usa a internet pública por custo menor e atende só balanceadores regionais.

Terraform e IaC

terraform init baixa provedores e prepara o back-end; terraform import associa recurso existente ao estado sem recriar; back-end gcs guarda o estado em bucket com bloqueio. Config Connector declara recursos do Google Cloud como objetos do Kubernetes; Helm empacota apps; Fabric FAST é landing zone de referência em estágios; Application Design Center publica modelos aprovados e gera Terraform.

Implantação no Cloud Run e eventos

gcloud run deploy --source . usa Cloud Build e buildpacks sem Dockerfile. --add-cloudsql-instances liga ao Cloud SQL pelo conector integrado. Eventarc entrega eventos, como criação de objeto no bucket ou mensagem em tópico, a serviços e funções do Cloud Run.

Acesso e manutenção de VMs e nós

OS Login usa as contas e os papéis do IAM para autorizar SSH, sem distribuir chaves nos metadados. O VM Manager oferece patches, inventário e configuração do sistema operacional em escala, com agendamento e relatórios. No GKE, nós particulares usam apenas IPs internos e acessam a internet por Cloud NAT.

Ferramentas de IA para infraestrutura

A Gemini CLI é um agente de IA de código aberto no terminal, disponível no Cloud Shell, que usa ferramentas integradas e servidores MCP para gerar e corrigir configurações como as do Terraform. O Google Antigravity é a plataforma centrada em agentes, em que agentes autônomos planejam, escrevem, executam e testam código na IDE com supervisão humana.

Macetes para memorizar

💡 Classes de armazenamento

30-90-365: Nearline mês, Coldline trimestre, Archive ano. Standard não tem mínimo.

💡 Bancos

SQL regional, Spanner global, AlloyDB PostgreSQL turbo, Bigtable chave rápida, Firestore celular, BigQuery análise, Memorystore cache.

💡 Prioridade de firewall

Menor número manda; empate, nega.

💡 Transferência

Banda pouca, caixa na porta (Appliance); outra nuvem, Storage Transfer; SaaS no BigQuery, Data Transfer Service.

💡 Balanceador global

Global só com Premium; Standard é regional.

💡 Terraform

init prepara, import adota, gcs compartilha o estado com bloqueio.

💡 Arquivos compartilhados

NFS = Filestore; SMB = NetApp Volumes; paralelo de IA = Managed Lustre; temporário veloz = SSD local.

Tabela de memorização: não confunda

Quando é...	...e quando é
Transfer Appliance Muitos TB no datacenter local com conexão lenta (ex.: 500 TB).	**Storage Transfer Service** Dados em outra nuvem, HTTP ou local com banda suficiente (ex.: 200 TB de outro provedor).
Cloud SQL Relacional regional que escala verticalmente, com HA entre zonas e réplicas de leitura.	**Spanner** Relacional global com consistência forte entre regiões e escala horizontal.
Bigtable Milhões de gravações por segundo e leitura de baixa latência por chave (séries temporais, IoT).	**BigQuery** Consultas SQL analíticas sobre petabytes, sem gerenciar servidores.
Filestore Compartilhamento NFS gerenciado para várias VMs.	**NetApp Volumes** Aplicação que precisa de SMB (Windows) com snapshots, ou NFS, iSCSI e NVMe/TCP.
Cloud SQL HA regional Continuar disponível se a zona principal falhar.	**Réplica de leitura** Aliviar a instância principal de relatórios de leitura; não protege contra falha de zona.
Balanceador de aplicativo externo global IP anycast único direcionando usuários à região mais próxima (exige Premium).	**Balanceador de aplicativo externo regional** Exigência legal de terminar e processar o tráfego em uma única região.
Cloud VPN de alta disponibilidade Conexão criptografada pela internet com SLA de 99,99%.	**Interconexão dedicada** Conexão privada de alta capacidade (ex.: 10 Gbps) sem passar pela internet pública.

Números que caem

Item	Valor
Peso do domínio no guia vigente	**30%**
Duração mínima Nearline / Coldline / Archive	**30 / 90 / 365 dias**
SLA da Cloud VPN de alta disponibilidade	**99,99%**
Intervalos das sondas de verificação de integridade	**35.191.0.0/16 e 130.211.0.0/22**
Menor prioridade de regra de firewall (regras implícitas)	**65535**

Alertas de prova (pegadinhas)

⚠ **Alerta de prova:** Escolher Coldline ou Archive para conteúdo lido milhares de vezes por dia porque o armazenamento é mais barato.
Como não cair: Dados quentes vão para Standard, sem duração mínima nem taxa de recuperação.

⚠ **Alerta de prova:** Copiar 500 TB por gcloud storage cp ou console com internet lenta.
Como não cair: Pouca banda com grande volume local pede Transfer Appliance.

⚠ **Alerta de prova:** Usar Transfer Appliance para trazer dados de outro provedor de nuvem.
Como não cair: Appliance é para dados locais com pouca banda; de outra nuvem use Storage Transfer Service.

⚠ **Alerta de prova:** Criar réplica de leitura na mesma zona para alta disponibilidade.
Como não cair: Ela falha junto com a zona. HA do Cloud SQL usa instância em espera em outra zona.

⚠ **Alerta de prova:** Usar MIG zonal ou uma VM maior para sobreviver à falha de zona.
Como não cair: Só o MIG regional distribui instâncias entre zonas.

⚠ **Alerta de prova:** Recriar VMs travadas com autoscaling por CPU.
Como não cair: Autoscaling não detecta app travada; use recuperação automática com verificação de integridade.

⚠ **Alerta de prova:** Achar que a regra de prioridade 1000 vence a de 900 por ser número maior, ou que permitir vence negar.
Como não cair: Número menor tem prioridade maior; em empate, negar vence.

⚠ **Alerta de prova:** Confiar em tag de rede como alvo de firewall do banco.
Como não cair: Qualquer editor da VM muda tags. Conta de serviço como alvo é mais rígida; em políticas do Cloud NGFW, use tags seguras.

⚠ **Alerta de prova:** Usar Cloud VPN clássica de túnel único quando se exige 99,99%.
Como não cair: Só a VPN de alta disponibilidade, com dois túneis, atinge 99,99%.

⚠ **Alerta de prova:** Usar Pub/Sub para migrar sistema Kafka sem mudar código.
Como não cair: Pub/Sub exigiria adaptar o código; Managed Service for Apache Kafka mantém compatibilidade sem administrar brokers.

⚠ **Alerta de prova:** Apagar o bucket ou rodar terraform destroy e apply para passar recurso manual ao Terraform.
Como não cair: Declare o recurso e use terraform import; nada é recriado.

⚠ **Alerta de prova:** Pôr todo o cluster GKE em nós spot para economizar.
Como não cair: Crie pool spot só para os lotes tolerantes, com taints e tolerâncias.

⚠ **Alerta de prova:** Escolher Autopilot para DaemonSets privilegiados ou Cloud Run para módulos de kernel.
Como não cair: Autopilot restringe cargas privilegiadas (use Standard); módulos de kernel e controle total pedem Compute Engine.

⚠ **Alerta de prova:** Back-ends não íntegros: procurar rótulo de saúde, orçamento ou DNS.
Como não cair: Falta regra de firewall liberando as sondas do Google (35.191.0.0/16 e 130.211.0.0/22).

Como cai na prova

📝 **Como cai:** Cenário de requisito único e quatro produtos parecidos (qual banco, qual armazenamento, qual computação).
Como resolver: Grife a palavra decisiva (global, NoSQL por chave, offline, NFS, escala até zero, kernel) e elimine pelo que cada distrator não oferece: escopo regional, tipo de dado ou necessidade de gerenciar servidores.

📝 **Como cai:** Classe de armazenamento a partir da frequência de acesso.
Como resolver: Converta a frequência: diário = Standard; mensal = Nearline; trimestral = Coldline; anual ou menos = Archive; imprevisível sem regras = Autoclass.

📝 **Como cai:** Qual balanceador usar.
Como resolver: Pergunte: interno ou externo? HTTP ou TCP próprio? Global ou uma única região? Externo HTTP multirregião = aplicativo externo global; regra legal de região = regional; só dentro da VPC = interno; TCP próprio global = proxy de rede externo.

📝 **Como cai:** Alta disponibilidade: sobreviver a falha de zona ou região.
Como resolver: Zona: MIG regional, Cloud SQL HA, disco regional, cluster regional. Região: bucket birregional ou multirregional, snapshots em local multirregional, Spanner multirregional.

📝 **Como cai:** Qual comando ou flag implanta o recurso.
Como resolver: Lembre o verbo exato: set-autoscaling, get-credentials, kubectl apply -f, bq load gs://, gcloud run deploy --source ., --add-cloudsql-instances, networks create --subnet-mode=custom, terraform init.

Gatilhos do enunciado

Se o enunciado trouxer...	Pense em
Contêiner sem estado, tráfego irregular, escalar até zero	**Serviço Cloud Run**
Rodar até terminar uma vez por noite, sem HTTP	**Job do Cloud Run com Cloud Scheduler**
Arquivo novo no bucket dispara código	**Função do Cloud Run com gatilho do Eventarc**
Lote interrompível e mais barato	**Spot VMs**
Combinação de vCPU e memória que não existe	**Tipo de máquina personalizado**
Consistência forte entre regiões, relacional	**Spanner**
Séries temporais, IoT, baixa latência por chave	**Bigtable**
Mobile, tempo real e offline	**Firestore**
Mesmas consultas milhares de vezes, milissegundos	**Memorystore**
Protocolo TCP próprio com IP global	**Balanceador de rede de proxy externo**
DDoS e WAF no balanceador externo	**Cloud Armor**
VMs sem IP externo acessando Cloud Storage e BigQuery	**Acesso privado do Google na sub-rede**
Cloud SQL só com IP privado	**Acesso a serviços particulares**
Rotas aprendidas automaticamente com o datacenter	**Cloud Router com BGP**

🎯 Se você lembrar de só uma coisa deste capítulo...

- VPC sem regras extras: entrada negada, saída permitida.
- Disco regional replica de forma síncrona entre duas zonas; SSD local perde os dados ao parar a VM.
- Balanceador global exige nível Premium; nível Standard usa a internet pública e é regional.
- Assinatura push entrega ao endpoint; mensagens que falham repetidamente vão ao tópico de mensagens inativas.
- Para VM com conta de serviço: --service-account com --scopes=cloud-platform e acesso controlado pelo IAM.

Essenciais para a prova (checklist)

☐ Sei escolher entre Cloud Run (serviço, job, função), GKE Autopilot, GKE Standard, Compute Engine e Agent Runtime.

☐ Sei quando usar tipo personalizado, E2 de núcleo compartilhado, família M, Spot VMs, TPU e GPU.

☐ Sei a classe do Cloud Storage para cada frequência de acesso e a duração mínima de cada uma.

☐ Sei escolher entre Hyperdisk, disco regional, SSD local, Filestore, NetApp Volumes e Managed Lustre.

☐ Sei diferenciar Cloud SQL, Spanner, AlloyDB, Bigtable, Firestore, BigQuery e Memorystore.

☐ Sei escolher entre Transfer Appliance, Storage Transfer Service e BigQuery Data Transfer Service.

☐ Sei como prioridade e regras implícitas do firewall decidem o tráfego.

☐ Sei escolher o balanceador por alcance (interno/externo), protocolo e escopo (global/regional).

☐ Sei para que servem VPC compartilhada, peering, HA VPN, Cloud Router, Interconnect, acesso privado do Google e acesso a serviços particulares.

☐ Sei os comandos básicos de Terraform (init, import, back-end gcs) e quando usar Config Connector e Helm.

Questões de revisão do capítulo

R4.1 · Planejar e Implementar a Solução em Nuvem

Uma empresa precisa levar 500 TB para o Cloud Storage, mas a conexão com a internet é lenta. Qual opção é a mais adequada?

A. Dispositivo Transfer Appliance
B. Assinatura do Pub/Sub
C. gcloud storage cp pela rede
D. Upload pelo console web do Storage

R4.2 · Planejar e Implementar a Solução em Nuvem

Uma VM deve ser criada com a conta de serviço app-sa e acesso autorizado apenas pelo IAM, sem restringir por escopos. Quais opções usar?

A. --tags=app-sa e --labels=app-sa
B. --scopes=default sem conta definida
C. --service-account e --no-address na nova VM
D. --service-account e --scopes=cloud-platform

R4.3 · Planejar e Implementar a Solução em Nuvem

Uma aplicação precisa de VMs com 6 vCPUs e 40 GB de memória, combinação que não existe nos tipos predefinidos. O que usar?

A. VM spot do menor tipo
B. Tipo de máquina personalizado
C. Nó de locatário único
D. Tipo predefinido maior disponível

R4.4 · Planejar e Implementar a Solução em Nuvem

Um conteúdo de site é lido milhares de vezes por dia pelos usuários. Qual classe do Cloud Storage é a adequada?

A. Classe Standard
B. Classe Coldline
C. Classe Nearline
D. Classe Archive

Respostas comentadas

R4.1. A) Dispositivo Transfer Appliance

Por quê: O Transfer Appliance é indicado para grandes volumes com pouca banda disponível.

✗ B) Incorreta. Não transfere arquivos.
✗ C) Incorreta. Levaria tempo demais.
✗ D) Incorreta. Levaria tempo demais.

> **Dica de prova:** Muitos dados e pouca banda: Transfer Appliance.

R4.2. D) --service-account e --scopes=cloud-platform

Por quê: A recomendação é usar o escopo cloud-platform e controlar o acesso por papéis do IAM.

✗ A) Incorreta. Tags e rótulos não anexam contas.

✗ B) Incorreta. Usa a conta padrão.

✗ C) Incorreta. Remove o IP, não trata escopos.

> **Dica de prova:** Conta de serviço na VM: escopo cloud-platform e IAM.

R4.3. B) Tipo de máquina personalizado

Por quê: Tipos de máquina personalizados permitem escolher vCPUs e memória para evitar desperdício.

✗ A) Incorreta. Não atende à capacidade.

✗ C) Incorreta. Isola hardware, não ajusta o tamanho.

✗ D) Incorreta. Paga por recursos que não usa.

> **Dica de prova:** Combinação sob medida de CPU e memória: tipo personalizado.

R4.4. A) Classe Standard

Por quê: A classe Standard não tem duração mínima nem taxa de recuperação e serve a dados quentes.

✗ B) Incorreta. Tem custo de acesso e mínimo de 90 dias.

✗ C) Incorreta. Tem custo de acesso e mínimo de 30 dias.

✗ D) Incorreta. Cobra caro pelo acesso.

> **Dica de prova:** Dados acessados muito: Standard.

DIA DA PROVA

Estratégia de prova

A prova Google Cloud Associate Cloud Engineer tem de 50 a 60 questões de múltipla escolha e múltipla resposta em 2 horas, e o Google não divulga a nota de aprovação. Os enunciados são cenários curtos de operação: um requisito, uma restrição e a pergunta pelo produto, recurso, papel ou comando que resolve. Os distratores costumam ser produtos reais do mesmo universo que fazem outra coisa (rótulo em vez de política, réplica em vez de HA, scale em vez de resize). A resposta certa é a mais gerenciada, de menor privilégio e que atende ao requisito sem trabalho manual. Planejar e Implementar e Garantir a Operação pesam 30% cada; Configurar o Ambiente e Configurar Acesso e Segurança, 20% cada.

Como responder

✓ Grife o requisito decisivo do enunciado (escala até zero, uma única região, sem chaves, sem recriar, sem desligar) antes de olhar as alternativas.

✓ Elimine primeiro as alternativas que fazem outra coisa: rótulo e orçamento nunca bloqueiam, firewall não controla IAM, réplica de leitura não é HA.

✓ Entre duas certas, prefira a solução gerenciada e automática em vez de script manual, VM própria ou ação humana diária.

✓ Em questões de IAM, descarte proprietário, editor e papéis na organização; escolha o papel mais específico no menor recurso.

✓ Em questões de comando, associe o verbo ao comando exato (resize x scale, rollout undo x rollout status, get-iam-policy x add-iam-policy-binding).

✓ Em Escolha DUAS, avalie cada alternativa isoladamente como verdadeira ou falsa e marque exatamente duas.

✓ Com 2 horas para 50 a 60 questões, o ritmo é de cerca de 2 minutos por questão: não gaste tempo demais em um único enunciado.

Erros que derrubam candidatos

✗ Achar que orçamento impede gasto ou desliga recursos sozinho.

✗ Confundir HPA, VPA e cluster autoscaler, e kubectl scale com gcloud container clusters resize.

✗ Escolher classe de armazenamento só pelo preço do armazenamento, ignorando frequência de acesso e duração mínima.

✗ Confundir serviceAccountUser (usar e anexar) com Token Creator (personificar) e Key Admin (chaves).

✗ Esquecer que registros de acesso a dados vêm desativados na maioria dos serviços, enquanto atividade do administrador e eventos do sistema são sempre ativos.

✗ Trocar Transfer Appliance por Storage Transfer Service e balanceador global por regional.

✗ Ignorar herança do IAM e achar que remover o papel no projeto basta.

Na semana da prova

- ☐ Revise cada erro pelo motivo de a alternativa estar errada, não só pela resposta certa.
- ☐ Revise as tabelas de decisão: classes de armazenamento, bancos, balanceadores, escalonadores do GKE e papéis de contas de serviço.
- ☐ Repasse os comandos gcloud, kubectl, bq e terraform mais cobrados até reconhecê-los pelo verbo.
- ☐ Releia os números fixos: 30/90/365 dias das classes, _Default 30 e _Required 400 dias, 35.235.240.0/20 do IAP e prioridades de firewall.
- ☐ Na véspera, só revisão leve das pegadinhas e dos gatilhos; durma bem e confira os requisitos do dia da prova.

REVISÃO FINAL

Folha de cola

Leia na véspera e na manhã da prova: os gatilhos e as pegadinhas mais importantes de cada domínio.

1. Configurar Acesso e Segurança

‣ Descobrir qual permissão falta e por quê → **Policy Troubleshooter**

‣ Usuários com permissões que nunca usam → **Recomendações de papéis do IAM**

‣ Acesso só até o fim do contrato → **Condição do IAM com data de expiração**

‣ Ninguém exclui produção, nem proprietário → **Política de negação do IAM**

‣ CI em outro provedor sem chaves → **Federação de identidade da carga de trabalho**

⚠ A política efetiva é a união com as herdadas; remova na pasta.

⚠ No projeto vale para todas as contas; conceda na política da própria conta de serviço.

⚠ Token Creator permite personificar; o papel certo é Service Account Key Admin.

2. Configurar o Ambiente da Solução em Nuvem

‣ Impedir em toda a organização, de forma central → **Restrição de política da organização**

‣ Nenhum recurso fora das regiões da UE → **gcp.resourceLocations**

‣ Projetos novos sem rede default → **compute.skipDefaultNetworkCreation**

‣ Só imagens de projetos aprovados → **Restrição de imagens confiáveis (trustedImageProjects)**

‣ Separar custos por equipe no mesmo projeto → **Rótulos nos recursos**

⚠ Orçamento somente de alertas avisa e não bloqueia. Para VMs, o controle exige automação via notificações no Pub/Sub.

⚠ Rótulos não bloqueiam ações. O controle preventivo é restrição de política da organização.

⚠ Firewall não impede criar o IP externo; use compute.vmExternalIpAccess na política da organização.

3. Garantir a Operação Bem-Sucedida da Solução em Nuvem

‣ Voltar ao estado exato das 14h05 → **Recuperação a um ponto no tempo**

‣ Backup diário automático de discos com retenção → **Programação de snapshots**

‣ Memória da VM não aparece no Monitoring → **Ops Agent**

‣ Site fora do ar visto de vários lugares do mundo → **Verificação de tempo de atividade**

‣ Alerta quando a mensagem X aparecer nos registros → **Métrica baseada em registros com alerta**

⚠ A réplica replica o erro e a exportação perde dados recentes; use recuperação a um ponto no tempo.

⚠ Discos e IPs estáticos reservados continuam cobrados.

⚠ scale muda réplicas; nós do pool mudam com gcloud container clusters resize.

4. Planejar e Implementar a Solução em Nuvem

‣ Contêiner sem estado, tráfego irregular, escalar até zero → **Serviço Cloud Run**

‣ Rodar até terminar uma vez por noite, sem HTTP → **Job do Cloud Run com Cloud Scheduler**

‣ Arquivo novo no bucket dispara código → **Função do Cloud Run com gatilho do Eventarc**

‣ Lote interrompível e mais barato → **Spot VMs**

‣ Combinação de vCPU e memória que não existe → **Tipo de máquina personalizado**

⚠ Dados quentes vão para Standard, sem duração mínima nem taxa de recuperação.

⚠ Pouca banda com grande volume local pede Transfer Appliance.

⚠ Appliance é para dados locais com pouca banda; de outra nuvem use Storage Transfer Service.

GABARITO

Respostas do diagnóstico

D1. A) gcloud iam roles describe → revise o capítulo 1

Por quê: gcloud iam roles describe mostra título, estágio e permissões de um papel.

✗ B) Incorreta. Revoga credenciais locais.
✗ C) Incorreta. Exclui papéis personalizados.
✗ D) Incorreta. Lista projetos.

> **Dica de prova:** Permissões de um papel: gcloud iam roles describe.

D2. C) --service-account com o e-mail da conta → revise o capítulo 1

Por quê: A identidade do serviço do Cloud Run é definida com --service-account em gcloud run deploy ou services update; o recomendado é uma conta dedicada com papéis mínimos.

✗ A) Incorreta. Rótulos não definem identidade.
✗ B) Incorreta. Trata de quem pode invocar o serviço, não da identidade dele.
✗ D) Incorreta. Expor uma chave em variável de ambiente é inseguro e desnecessário.

> **Dica de prova:** Identidade do Cloud Run: --service-account.

D3. A) Papel de administrador de chaves de contas de serviço → revise o capítulo 1

Por quê: O papel Service Account Key Admin gerencia chaves de contas de serviço; ainda assim, o recomendado é evitar chaves e preferir federação ou personificação.

✗ B) Incorreta. Concede permissões amplas demais.
✗ C) Incorreta. Permite gerar tokens e personificar a conta.
✗ D) Incorreta. Permite anexar e usar a conta, não gerenciar chaves.

> **Dica de prova:** Gerenciar só as chaves: Service Account Key Admin.

D4. C) Projetos separados sob pastas por ambiente → revise o capítulo 2

Por quê: Projetos são a fronteira de isolamento; pastas agrupam projetos para aplicar políticas comuns.

✗ A) Incorreta. Rótulos não isolam permissões.
✗ B) Incorreta. Não isola acessos.
✗ D) Incorreta. Mistura acessos e cotas.

> **Dica de prova:** Isolar dev e prod: projetos separados em pastas.

D5. A) No nível da organização do Google Cloud → revise o capítulo 2

Por quê: Na federação de identidade de colaboradores, os pools ficam na organização e os provedores, OIDC ou SAML 2.0, descrevem a relação com o IdP, como Okta ou Microsoft Entra ID.

✗ B) Incorreta. Pools de colaboradores não são criados por projeto.
✗ C) Incorreta. O pool é definido na organização, não em pastas.
✗ D) Incorreta. A conta de faturamento não gerencia identidades.

> **Dica de prova:** Pool de colaboradores: nível da organização.

D6. B) Rótulos nos recursos com equipe e ambiente → revise o capítulo 2

Por quê: Rótulos (labels) são pares chave-valor exportados com o faturamento, úteis para alocar custos.

✗ A) Incorreta. É impraticável e desnecessário.
✗ C) Incorreta. Cotas limitam uso, não separam custos.
✗ D) Incorreta. Não classificam custos.

> **Dica de prova:** Custos por equipe no mesmo projeto: rótulos.

D7. A) Sink de logs para o BigQuery → revise o capítulo 3

Por quê: Coletores (sinks) roteiam registros ao BigQuery, Cloud Storage, Pub/Sub ou buckets de registros com retenção definida.

✗ B) Incorreta. Não guarda por cinco anos por padrão.
✗ C) Incorreta. Não é viável.
✗ D) Incorreta. Não guardam logs.

> **Dica de prova:** Guardar logs e consultar com SQL: sink para BigQuery.

D8. B) Namespaces por equipe → revise o capítulo 3

Por quê: Namespaces dividem o cluster e podem ter cotas de recursos e RBAC próprios.

✗ A) Incorreta. Não separam recursos.
✗ C) Incorreta. É inviável e caro.
✗ D) Incorreta. Não é possível.

> **Dica de prova:** Separar equipes no cluster: namespaces.

D9. B) Política de firewall hierárquica na organização → revise o capítulo 3

Por quê: Políticas de firewall hierárquicas são associadas à organização ou a pastas; regras inferiores não as sobrepõem, e a ação goto_next delega a avaliação aos níveis abaixo.

✗ A) Incorreta. Depende de cada equipe e pode ser alterada no projeto.
✗ C) Incorreta. Políticas de negação do IAM controlam permissões, não tráfego de rede.
✗ D) Incorreta. Rotas direcionam tráfego, mas não são o controle de firewall indicado.

> **Dica de prova:** Bloqueio de rede em toda a organização: política hierárquica.

D10. A) VMs do Compute Engine → revise o capítulo 4

Por quê: O Compute Engine oferece controle total do sistema operacional das VMs.

✗ B) Incorreta. Restringe acesso privilegiado aos nós.
✗ C) Incorreta. Não permitem módulos do kernel.
✗ D) Incorreta. Não permitem módulos do kernel.

> **Dica de prova:** Controle do SO e do kernel: Compute Engine.

D11. B) Aceleradores Cloud TPU → revise o capítulo 4

Por quê: A documentação do Cloud TPU recomenda TPUs para modelos grandes dominados por cálculos de matriz e treinos de semanas; GPUs são mais flexíveis para operações personalizadas.

✗ A) Incorreta. Isolam o hardware por licença ou conformidade, sem acelerar o treino.
✗ C) Incorreta. GPUs são preferíveis quando há muitas operações personalizadas ou não suportadas na TPU.
✗ D) Incorreta. CPUs atendem protótipos e modelos pequenos, não treinos longos de modelos grandes.

> **Dica de prova:** Treino longo dominado por matrizes: TPU.

D12. A) Serviço de transferência de dados do BigQuery → revise o capítulo 4

Por quê: O BigQuery Data Transfer Service agenda cargas de SaaS, outras nuvens e do Cloud Storage.

✗ B) Incorreta. Não carrega dados.
✗ C) Incorreta. Depende de ação humana.
✗ D) Incorreta. Não carrega dados.

> **Dica de prova:** Carga agendada no BigQuery: Data Transfer Service.

Agora, treine no formato real

Faça as 6 provas completas do curso Google Associate Cloud Engineer: 300 questões inéditas, tempo do exame, explicação de cada alternativa e link da documentação para revisar.

Ver as 6 provas completas e o cupom do mês →

Até o dia da prova, uma revisão por dia no seu feed

Instagram @strecht.simulados

Um post de prova por dia: questão comentada, pegadinhas e o "não confunda" de cada certificação.

Seguir no Instagram →

YouTube Strecht Simulados

Shorts diários com pegadinhas de prova e, aos sábados, um simulado comentado.

Inscrever-se no canal →

Material independente de preparação, sem afiliação ou endosso de Google Cloud. As questões são inéditas e não reproduzem a prova real.