Ch09

User and Project Management in GCP

CHAPTER ASSIGNMENT

Name _____ Class _____ No. _____ Date _____

> Answer every question in the space given. Show your working. You may use the chapter page to help you.

Questions from *Data Engineering with Google Cloud Platform*, 1st edition, Adi Wijaya (Packt, 2022), and from the Guru Jazzy chapter page, for classroom use.

Exercises and scenarios from the book

Ex1 **Use-case scenario ? planning a BigQuery ACL on an e-commerce organization**

An e-commerce company has four user groups (**Data Engineer**, **Marketing**, **Sales**, **Head of Analysts**) and six tables (`user_profile`, `user_address`, `customer_checkouts`, `customer_carts`, `event_campaigns`, `campaign_metrics`) with these access requirements:

? The Data Engineers group can create and view all tables.

? The Head of Analysts group can view all tables except `user_profile` and `user_address`.

? Sales can view data in `customer_checkouts` and `customer_carts`.

? Marketing can view data in `event_campaigns`, `campaign_metrics`, and `user_profile`.

? Marketing cannot access credit card number columns that are stored in the `user_profile` table.

(1) How would you group the tables into datasets and grant predefined BigQuery IAM roles at the project, dataset, or table level to satisfy the first four requirements scalably?

(2) Explain how Column-level security in BigQuery (using Data Catalog Policy tags and the Fine-Grained Reader role) satisfies the fifth requirement so that Marketing can query `user_profile` without accessing the `credit_card_number` column.

Ex2 **Exercise ? creating and running basic Terraform scripts**

Practice Infrastructure as Code (IaC) in Cloud Shell using Terraform (`chapter-9/code`):

(1) Step 1 ? Downloading Terraform in the Cloud Shell environment: verify Terraform in Cloud Shell (`terraform -help`).

(2) Step 2 ? Configuring the Terraform backend to GCS: create a `[your project id]-tfstate` GCS bucket and configure `backend.tf`. Explain why storing Terraform state in GCS is better than storing it locally.

(3) Step 3 ? Using variables in Terraform: configure `variables.tf` (`project_id`, `location`), `terraform.tfvars`, and `main.tf` (`google_storage_bucket`).

(4) Step 4 ? Running the Terraform commands: run `terraform init`, `terraform plan`, and `terraform apply`. Describe what each of the three commands does.

(5) Step 5 ? Configuring a BigQuery dataset: uncomment `module "bigquery"` in `main.tf` (`dataset_id = "demo_terraform_dataset"`), re-run `terraform init`, `terraform plan`, and `terraform apply`, and verify the dataset in BigQuery.

Ex3 **Self-exercise ? managing a GCP project and resources using Terraform**

Develop a new Terraform project on your own to strengthen your understanding of Terraform. Write down the Terraform resources/modules and configuration needed to accomplish the following four tasks:

(1) Create a new GCP project.

(2) Enable the BigQuery API and the Cloud Composer API on the new project.

(3) Create a BigQuery dataset and table inside the new project.

(4) Grant a BigQuery role to a user on the new project.

Sorting activities from the chapter page

S1 **Which kind of role?**

Identity and Access Management is the central authorisation service for all of GCP. The idea is simple: grant roles to accounts.

Options: **Basic** ? **Predefined** ? **Custom**

No.	Scenario	Answer
1	Your own owner role in a throwaway practice project.	
2	A data engineer who needs to build BigQuery pipelines, and nothing else.	
3	No predefined role has quite the permission set your organisation needs.	

S2 **How many projects should you have?**

There is no single right answer, and every company differs. Small companies start with one project and grow. The decision comes down to three benefits, weighed against the overhead of splitting.

Options: **One project** ? **Split projects**

No.	Scenario	Answer
1	A start-up with a small team building an MVP and no one to run operations.	
2	A company where the app, data and ML teams use different services and need different roles.	
3	More than 1,000 active BigQuery users, and a per-project limit of 100 concurrent interactive queries.	
4	Finance wants to see cost per team on the billing dashboard.	

S3 **Where should the grant go?**

An access control list is simply IAM applied to data. BigQuery has two kinds of permission: job permissions (may you run a query) and access permissions (may you read this data).

Options: **Project** ? **Dataset** ? **Table** ? **Column**

No.	Scenario	Answer
1	Data engineers must create and view every table in the warehouse, including tables that do not exist yet.	
2	Sales needs the checkout and cart tables. Group those two tables into one dataset and grant on it.	
3	One analyst needs to read a single table and nothing else.	
4	Marketing may read the user profile table but never its credit card number column.	

Quiz from the chapter page: Check yourself

1 Can you grant a permission directly to an account?

A. Yes, always

B. No: you grant roles, which contain permissions

C. Only to service accounts

D. Only at organisation level

---

2 Why is granting roles/editor to a data engineer for BigQuery work risky?

A. It removes BigQuery access

B. Basic roles apply to every resource in the project, far beyond BigQuery

C. Editors cannot run queries

D. It costs extra

---

3 What is the best reason to create a parent folder?

A. It tidies projects like folders on your laptop

B. A group of projects should share inherited permissions from one grant

C. Folders reduce cost

D. Organizations require at least three

---

4 How many organizations does a company typically have in GCP?

A. One per team

B. One per project

C. One: it is the single root of everything

D. One per region

---

5 To hide a credit card column, whose permission do you manage?

A. The BigQuery column directly

B. The policy tag in the taxonomy, not the column

C. The bucket

D. The project owner only

---

6 You run terraform plan. What does it do?

A. Creates the resources immediately

B. Shows what Terraform would add, change or destroy, without doing it

C. Downloads the providers

D. Deletes the backend bucket

End of assignment · Check your answers against the chapter at gurujazzy.com/subjects/gcp-data-engineering/chapter-09-users-and-projects-iam/