Resource Management

Essential Cloud Infrastructure: Core Services

CLOUD RESOURCE MANAGER

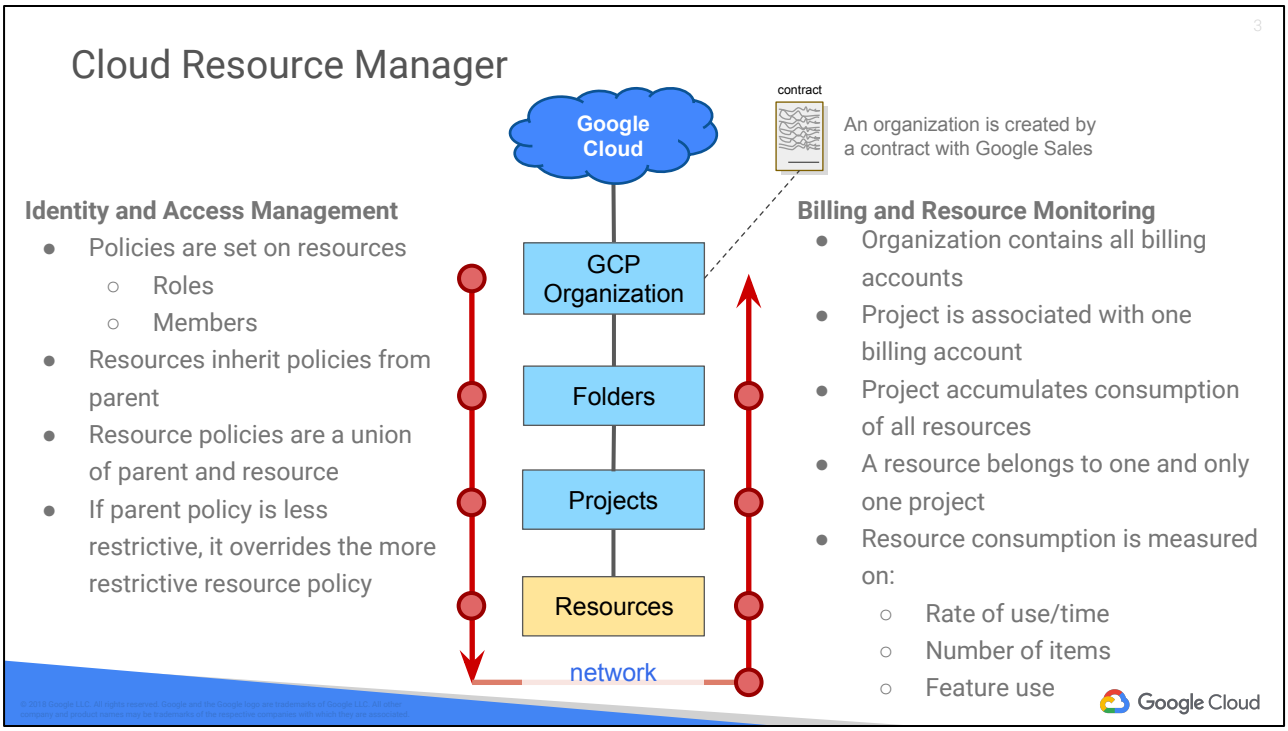
EXAMINING BILLING DATA WITH BIGQUERY

Last modified 2018-1-29

Agenda

- **Cloud Resource Manager**
- Quotas
- Labels and names
- Billing
- Lab
- Quiz

Google Cloud

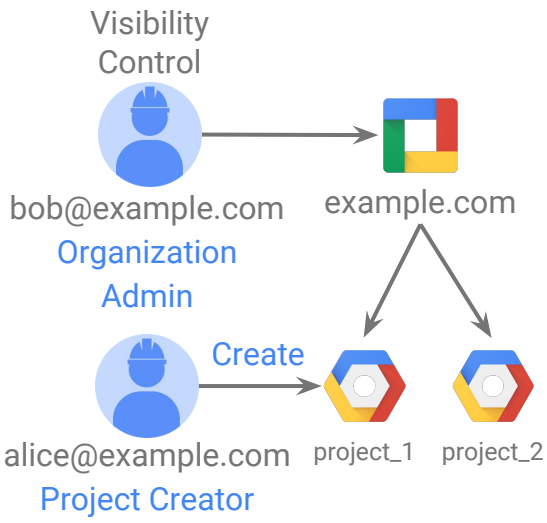


- IAM Policies are inherited from the **top down.**
- IAM Policies determine which **users** (or groups) have which access **roles** to which **resources.**
- Billing is accumulated from the **bottom up.** Reporting of billing is per project.

- You can set policies at the organization, folder, project, or individual resource level, to override policies set at the higher level in the hierarchy
- By default, all resources within the same project are "trusted."
- There is isolation between projects
    - Isolation is implemented in the network using Software Defined Networking (SDN), so that packets can be inspected and not permitted to pass between projects without authorization.

Organization node

- Organization node is root node for Google Cloud resources
- 2 organization roles:
  - Organization Admin: Control over all cloud resources
  - Project Creator: Controls project creation



Google Cloud

An organizational node is the root node for all Google Cloud Platform resources. The diagram shows a couple of examples where we have an individual is in control of the organizational domain. In the diagram, the individuals have delegated privileges and access to the individual projects.

Projects

- Track resource and quota usage
  - Enable billing
  - Manage permissions and credentials
  - Enable services and APIs
- Projects use three identifying attributes:
  - Project Name
  - Project Number
  - Project ID, also known as Application ID
- Google Cloud Platform Console or the Cloud Resource Manager API

Google Cloud

API actions:

- Get a list of all projects associated with an account.
- Create new projects.
- Update existing projects.
- Delete projects.
- Undelete, or recover, projects that you don't want to delete.

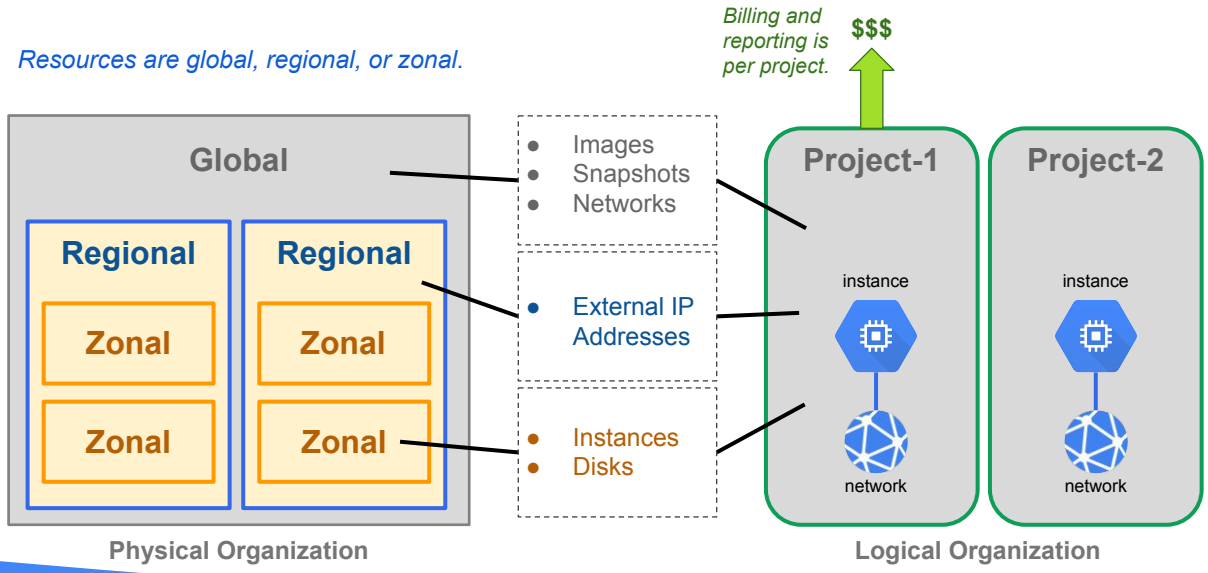
Projects
https://cloud.google.com/compute/docs/projects

https://cloud.google.com/compute/docs/networks-and-firewalls
A network can belong to only one project. An instance can attach to only one network. A project can contain up to 5 networks and 24 CPUs. A single network can contain a maximum of 7000 instances.

If you want servers to communicate over the Private IP, they have to be in the same project on the same network.
Because networks are homed to projects, as soon as two servers are located in different projects or on different networks in the same project, they must communicate over external IP addresses (which can be ephemeral or static). Internal DNS resolver of FQDN is part of Compute Engine.
https://cloud.google.com/compute/docs/networking

Resource hierarchy

Resources are global, regional, or zonal.



Google Cloud

Agenda

- Cloud Resource Manager
- **Quotas**
- Labels and names
- Billing
- Lab
- Quiz

Google Cloud

Project quotas

- All resources are subject to project quotas or limits.
  - Typically fall into one of three categories:
    - How many resources you can create per project
    - How quickly you can make API requests in a project—rate limits
    - Some quotas are per region
  - Quota examples:
    - 5 networks per project
    - 300 admin requests per minute (Cloud Spanner)
    - 24 CPUs region/project
  - Most quotas can be increased through self-service form or a support ticket
    - IAM & admin -> Quotas

Google Cloud

Compute Engine enforces quotas on resource usage for a variety of reasons. For example, quotas protect the community of Google Cloud Platform users by preventing unforeseen spikes in usage. Special quotas limit access for projects that are just exploring Google Cloud Platform on a free trial basis.

Not all projects have the same quotas. As your use of Google Cloud Platform expands over time, your quotas may increase accordingly. If you expect a notable upcoming increase in usage, you can proactively request quota adjustments from the Quotas page in the Google Cloud Platform Console.

How quotas are applied
Resource quotas are the maximum amount of resources you can create for that resource type, if those resources are available. Quotas do not guarantee that resources will be available at all times. If a resource is not available, you won't be able to create new resources of that type, even if you still have remaining quota in your region or project. This is particularly relevant for regional quotas; if a particular region is out of a resource, you won't be able to create a resource of that type, even if you still have quota. For example, if a region is out of local SSDs, you cannot create local SSDs in that region, even if you still had quota for local SSDs. In such cases, you should deploy regional resources in another region.

Check your quota
To check the available quota for resources in your project, go to the Quotas page in the Google Cloud Platform Console. If you are using gcloud, run the following

command to check your quotas. Replace myproject with your own project ID:
gcloud compute project-info describe --project myproject
To check your used quota in a region, run:
gcloud compute regions describe example-region

Why use project quotas?

- Prevent runaway consumption in case of an error or malicious attack
- Prevent billing spikes or surprises
- Forces sizing consideration and periodic review

Agenda

- Cloud Resource Manager
- Quotas
- **Labels and names**
- Billing
- Lab
- Quiz

Labels

- A utility for organizing Google Cloud Platform resources
  - Attached to resources: VM, disk, snapshot, image
  - Console, `gcloud` or API
- *Example uses of labels:*
  - Search and list all resources (inventory)
  - Filter resources (ex: separate production from test)

    Labels used in scripts
    - Help analyze costs
    - Run bulk operations

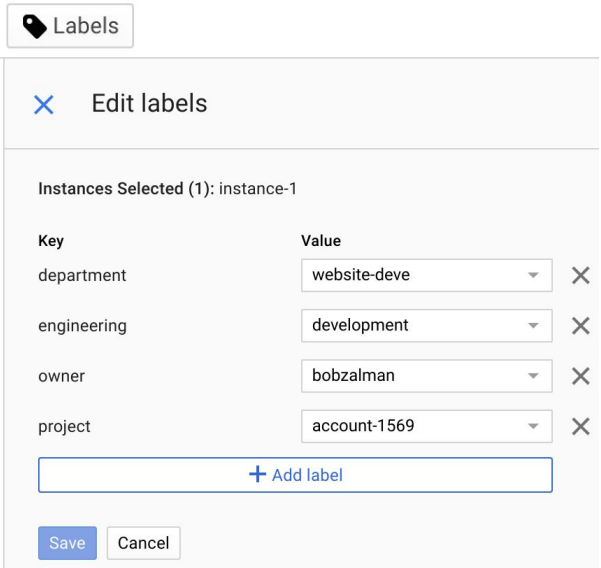
Google Cloud

The purpose of labels is to identify your resources. It's a simple text field that you can enter whatever information that might help you describe your resources.

For more information, see:
https://cloud.google.com/resource-manager/docs/using-labels

Label specification

- A label is a key-value pair.
- Label keys and non-empty label values can contain lowercase letters, digits, and hyphens, must start with a letter, and must end with a letter or digit. The regular expression is: [a-z]([-a-z0-9]*[a-z0-9])
- The maximum length of label keys and values is 63 characters.
- There can be a maximum of 64 labels per resource.



Google Cloud

Labels can only be managed through the Compute Engine Beta API or through the Google Cloud Platform Console. Labels are currently not supported through the gcloud command-line tool.
https://cloud.google.com/compute/docs/label-or-tag-resources

A label is a key-value pair.
Label keys and non-empty label values can contain lowercase letters, digits, and hyphens, must start with a letter, and must end with a letter or digit. The regular expression is: [a-z]([-a-z0-9]*[a-z0-9])
The maximum length of labels keys and values is 63 characters.
There can be a maximum of 64 labels per resource.

Label practices

- Team or Cost Center
  - Distinguish projects owned by different teams.
  - Useful in cost accounting or budgeting.
  - Examples: `team:marketing`, `team:research`
- Components
  - Examples: `component:redis`, `component:frontend`
- Environment or stage
  - Examples: `environment:prod`, `environment:test`
- Owner or contact
  - Person responsible for resource or primary contact for the resource
  - Examples: `owner:gaurav`, `contact:opm`
- State
  - Examples: `state:inuse`, `state:readyfordeletion`

Comparing labels and tags

- Labels are a way to organize resources across GCP
  - disks, image, snapshots...
- User-defined strings in `key-value` format
- Propagated through billing

- Tags are applied to instances only
- User-defined strings
- Tags are primarily used for networking (applying firewall rules)

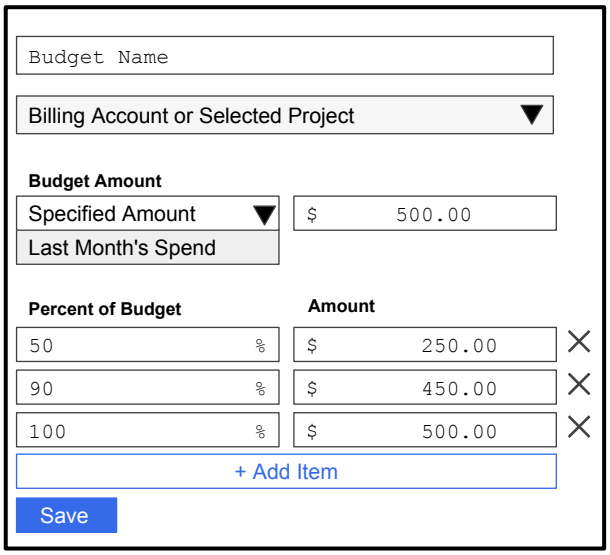
Google Cloud

Before March 7, 2017 VM tags were automatically created and synchronized with labels, appearing as value-less labels.
Create a tag, a label was created. Create a label, a tag was created. After that date, the two systems were separate: two name spaces.

Agenda

- Cloud Resource Manager
- Quotas
- Labels and names
- **Billing**
- Lab
- Quiz

Budgets and alerts



Google Cloud

To set a budget you must be a Billing Administrator.

- Budget lets you track how spend is approaching specified amount
- Monthly budget (period is fixed)
- Can only be created by Billing Administrators
- Budget applies to either an entire billing account or to a single project
- Budget amount is specified or can be automatically set to last month's spend on the account or project
- Alerts are set as a % of budget (0.005% to 100%) rounds up to the cent
- Notification is triggered when spend is greater than alert amount (frequency ~hourly)
- Notification is sent by email to the Billing Administrator that created it (it does not appear in the console notifications)
- You can choose whether to include consumption of credits in the calculation (credits are promotions or grants)

For more information, see: https://cloud.google.com/billing/docs/how-to/billing-access

Example notification email

Billing Alert Notification

Dear Google customer,

You are receiving this email because you are a Google Cloud Platform, Firebase, or API customer.

This is an automated notification to inform you that the project: **deadpool-cpb100** has exceeded **0.05%** of the monthly budget of **$100.00**.

You are receiving this message because there is an alert configured on this project's budget. To disable this alert or modify the budget's threshold, please edit your budget.

Billing export

JSON Field	CSV Field	Data Type	Description
accountID	Account ID	string	Billing account ID
lineItemID	Line Item	string	URI of the resource
startTime	Start Time	dateTime	Start of measured period of use
endTime	End Time	dateTime	End of measured period of use
projectNumber	Project Number	integer	Project number
projectID	Project ID	string	Project ID
projectName	Project Name	string	**Project Name**
projectLabels	Project Labels	string	**Project Labels**
measurementID	Measurement	string	URI of the resource
sum	Measurement Total Consumption	integer	Measured time of use
unit	Measurement Units	string	Time period units (ie seconds)
creditID	Credit	string	Credit grant ID
amount	Credit Amount	decimal	Amoiunt of the credit
currency	Credit Currency	string	Currency code (ie USD)
cost	Amount	decimal	Calculated cost
currency	Currency	string	Currency code (ie USD)

Google Cloud

- Export to either a file or a BigQuery dataset
- Create a Cloud Storage bucket or BigQuery dataset first, to specify when enabling
- Access is set via IAM on bucket or dataset
- File export is either in CSV or JSON format (not both)
- File export prefix name is appended with date-time-stamp
- Report is generated daily; there is no on-demand generation
- Project name and project labels are your primary post-export parsing tools

Fields that are exported to BigQuery:
https://support.google.com/cloud/answer/7237695

Agenda

- Cloud Resource Manager
- Quotas
- Labels and names
- Billing
- **Lab**
- Quiz

Google Cloud

Lab: Examining Billing Data with BigQuery

Objectives

In this lab, you learn how to perform the following tasks:

- Sign in to BigQuery from the GCP Console
- Create a dataset
- Create a table
- Import data from a billing CSV file stored in a bucket
- Run complex queries on a larger dataset

Completion: 45 minutes

Access: 90 minutes

BigQuery

Lab Review

In this lab you:

- Imported billing data into BigQuery that had been generated as a CSV file.
- Ran a simple query on the file.
- Accessed a shared dataset containing more than 22,000 records of billing information.
- Ran a variety of queries on that data to explore how you can use BigQuery to ask and answer questions by running queries.

Agenda

- Cloud Resource Manager
- Quotas
- Labels and names
- Billing
- Lab
- **Quiz**

Quiz

No resources in GCP can be used without being associated with...

1. A user
2. A virtual machine
3. A bucket
4. A project

Quiz

A budget is set at $500 and an alert is set at 100%. What happens when the full amount is used?

1. Everything in the associated project is suspended because there is not more budget to spend.
2. A notification email is sent to the Billing Administrator.
3. You have a 4-hour courtesy period before Google shuts down all resources.
4. Nothing. There is no point in sending a notification when there is no budget remaining.

Quiz

How do quotas protect GCP customers?

1. By preventing resource use in too many zones in a region.
2. By preventing resource use by unknown users.
3. By preventing resource use of too many different GCP services.
4. By preventing uncontrolled consumption of resources.

More resources

Cloud Resource Manager
https://cloud.google.com/resource-manager/

Quotas
https://cloud.google.com/compute/quotas

Labels
https://cloud.google.com/resource-manager/docs/using-labels

Choosing labels or tags
https://cloud.google.com/compute/docs/label-or-tag-resources

Google Cloud