

Módulo 4: Managing Cloud Instances at Scale

Introducción

En este módulo se aborda cómo **administrar grandes cantidades de instancias en la nube** de forma eficiente, utilizando herramientas y principios de automatización, monitoreo y escalabilidad. El foco está en **cómo pasar de administrar 5 servidores a cientos o miles**, manteniendo control, seguridad y rendimiento.

Conceptos clave

Escalabilidad en la nube

- Capacidad de aumentar o reducir recursos **según la demanda**.
 - Puede ser:
 - **Vertical (Scale Up)**: aumentar recursos en una instancia (CPU, RAM).
 - **Horizontal (Scale Out)**: agregar más instancias idénticas.
 - La automatización es esencial para escalar sin intervención manual.
-

Instancias y grupos gestionados

- En proveedores como **Google Cloud**, **AWS** o **Azure**, podés agrupar instancias en:
 - **Managed Instance Groups (MIGs)** o **Auto Scaling Groups (ASGs)**.
- Estas herramientas permiten:
 - Reemplazar automáticamente instancias fallidas.
 - Mantener una cantidad mínima/máxima deseada.
 - Aplicar actualizaciones controladas.

Ejemplo (conceptual):

```
# Crear un grupo de instancias gestionadas en GCP
gcloud compute instance-groups managed create web-servers \
  --base-instance-name web \
  --template web-template \
  --size 5
```

 Crea 5 instancias iguales basadas en una plantilla (**web-template**), todas administradas automáticamente.

Herramientas de gestión a gran escala

Herramienta	Función principal	Características
Puppet / Ansible / Chef	Configuración y orquestación	Aplican configuraciones de manera uniforme.
Terraform	Infrastructure as Code	Define recursos de múltiples proveedores.
Kubernetes	Orquestación de contenedores	Automatiza despliegues, escalado y actualizaciones.
Cloud SDKs / CLIs	Control directo por línea de comandos	Scripts para tareas recurrentes.

Automatización con scripts y APIs

- Cada proveedor cloud tiene **APIs REST o SDKs** (Python, Go, Java, etc.) que permiten:
 - Crear, eliminar y actualizar recursos.
 - Consultar estados del sistema.
 - Integrar con herramientas personalizadas.

Ejemplo (Python + Google Cloud SDK):


```
from googleapiclient import discovery
from oauth2client.client import GoogleCredentials

credentials = GoogleCredentials.get_application_default()
service = discovery.build('compute', 'v1', credentials=credentials)

project = 'mi-proyecto'
zone = 'us-central1-a'

request = service.instances().list(project=project, zone=zone)
response = request.execute()

for instance in response['items']:
    print(instance['name'])
```

 Script que lista todas las instancias activas en una zona.

Monitoreo y mantenimiento

- **Monitoreo continuo** → uso de CPU, RAM, disco, red, logs.
- Herramientas:
 - **Stackdriver (Google Cloud Monitoring)**
 - **AWS CloudWatch**
 - **Prometheus + Grafana**

- Alertas automáticas permiten escalar recursos o reiniciar servicios sin intervención manual.

Actualizaciones y despliegues a escala

- **Rolling Updates:** reemplazar instancias una por una sin downtime.
- **Blue-Green Deployment:** mantener dos entornos (actual y nuevo), con cambio de tráfico controlado.
- **Canary Releases:** desplegar primero a una pequeña porción de instancias para testeo.

Mejores prácticas para gestión a escala

Práctica	Descripción
Usar plantillas	Define configuraciones base reproducibles.
Control de versiones	Guarda scripts y archivos IaC en Git.
Monitoreo automatizado	Detecta fallos antes de que impacten usuarios.
Escalado automático	Ajusta recursos en tiempo real según métricas.
Automatizar parches y backups	Mantiene seguridad y resiliencia.

Resumen del módulo

Tema	Concepto clave
Escalabilidad	Capacidad de crecer automáticamente.
Managed Instance Groups	Administración automática de instancias idénticas.
APIs y SDKs	Control programático de infraestructura.
Monitoreo	Supervisión activa y alertas automáticas.
Rolling Updates	Despliegues sin interrupciones.

Referencias

- Google Cloud Docs: <https://cloud.google.com/compute/docs/instance-groups>
- AWS Auto Scaling: <https://aws.amazon.com/autoscaling/>
- Terraform Docs: <https://developer.hashicorp.com/terraform/docs>
- Google Cloud SDK: <https://cloud.google.com/sdk/docs>