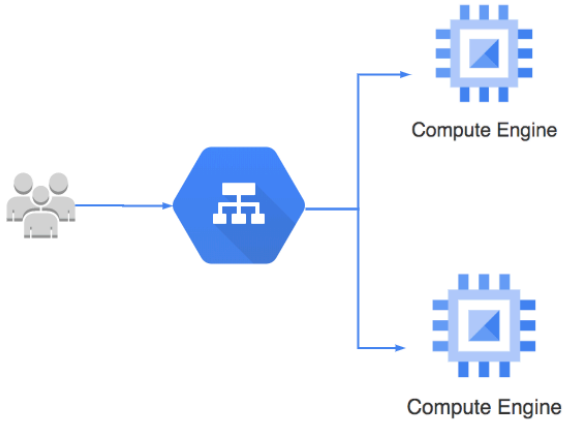
Instance Groups

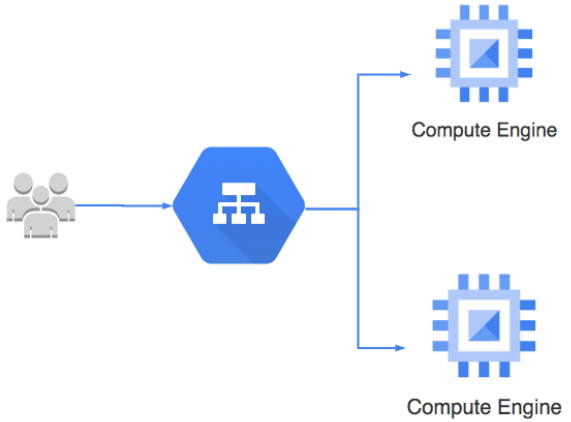
Instance Groups

- How do you create a group of VM instances?
  - **Instance Group** - Group of VM instances managed as a single entity
    - **Manage group** of similar VMs having similar lifecycle as **ONE UNIT**
- **Two Types of Instance Groups:**
  - **Managed** : Identical VMs created using a template:
    - Features: Auto scaling, auto healing and managed releases
  - **Unmanaged :** Different configuration for VMs in same group:
    - Does NOT offer auto scaling, auto healing & other services
    - NOT Recommended unless you need different kinds of VMs
- **Location** can be Zonal or Regional
  - Regional gives you higher availability (RECOMMENDED)



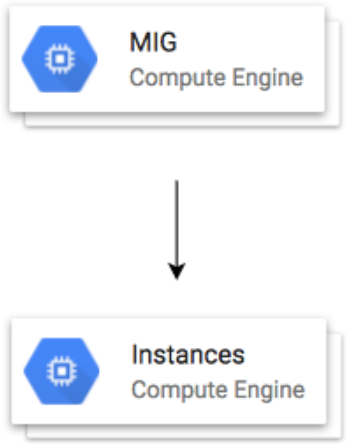
Managed Instance Groups (MIG)

- **Managed Instance Group** - Identical VMs created using an **instance template**
- Important Features:
  - **Maintain** certain number of instances
    - If an instance crashes, MIG launches another instance
  - **Detect application failures** using health checks (Self Healing)
  - Increase and decrease instances based on load (**Auto Scaling**)
  - Add **Load Balancer** to distribute load
  - Create instances in multiple zones (regional MIGs)
    - Regional MIGs provide higher availability compared to zonal MIGs
  - **Release** new application versions without downtime
    - **Rolling updates**: Release new version step by step (gradually). Update a percentage of instances to the new version at a time.
    - **Canary Deployment**: Test new version with a group of instances before releasing it across all instances.



Creating Managed Instance Group (MIG)

- **Instance template** is mandatory
- Configure **auto-scaling** to automatically adjust number of instances based on load:
  - **Minimum** number of instances
  - **Maximum** number of instances
  - **Autoscaling metrics:** CPU Utilization target or Load Balancer Utilization target or Any other metric from Stack Driver
    - **Cool-down period**: How long to wait before looking at auto scaling metrics again?
    - **Scale In Controls**: Prevent a sudden drop in no of VM instances
      - **Example**: Don't scale in by more than 10% or 3 instances in 5 minutes
  - **Autohealing**: Configure a Health check with Initial delay (How long should you wait for your app to initialize before running a health check?)
- Time for a **Demo**



Updating a Managed Instance Group (MIG)

In28 Minutes

- **Rolling update** - Gradual update of instances in an instance group to the new instance template
  - Specify new template:
    - (OPTIONAL) Specify a template for canary testing
  - Specify how you want the update to be done:
    - When should the update happen?
      - Start the update immediately (Proactive) or when instance group is resized later(Opportunistic)
    - How should the update happen?
      - **Maximum surge**: How many instances are added at any point in time?
      - **Maximum unavailable**: How many instances can be offline during the update?
- **Rolling Restart/replace**: Gradual restart or replace of all instances in the group
  - No change in template BUT replace/restart existing VMs
  - Configure Maximum surge, Maximum unavailable and What you want to do? (Restart/Replace)