

VCC Assignment 3

Report • Mar 23, 2025

Objective

Assignment 3: Create a Local VM and Auto-Scale It to GCP or Any Other Public Cloud When Resource Usage Exceeds 75%

By – Shreyash Kadam (G24AI1012)

Document Report

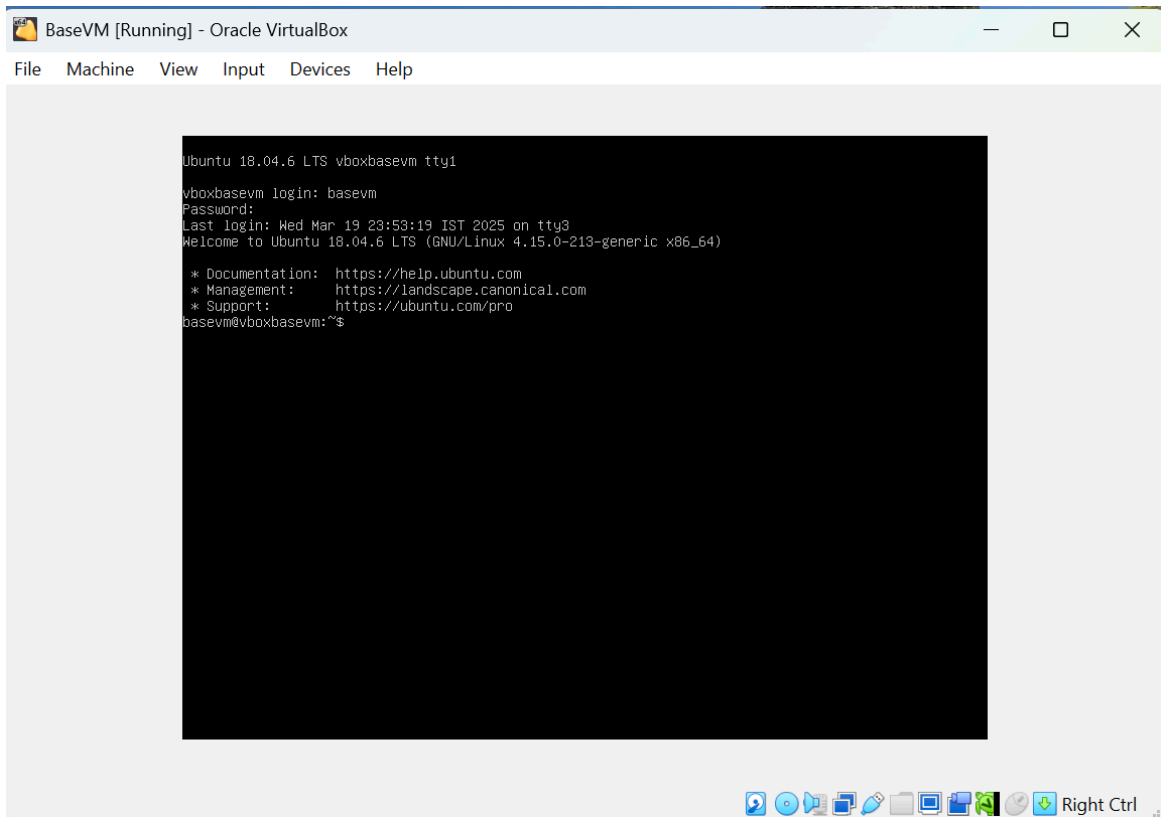
Step-by-Step Instructions for Implementation:

Step 1: Creation of a local VM using VirtualBox

- Install Virtual box
- Click New >> Provide the configuration of VM like Name(BaseVM), ISO Image (Ubuntu server Mini), Hardware details (2 CPUs and ram 2 GB) and HDD size 20 GB.

Note - Opted for Ubuntu server mini as it is lite weight and supports applications like Prometheus & Grafana

- Click on Finish.
- Start your VM.



Step 2: Implementation of resource monitoring

- Install Prometheus

```
sudo apt update && sudo apt install prometheus-node-exporter -y
```

- Verify if prometheus is installed

```
prometheus-node-exporter --version
```

```
basevm@vboxbasevm:~$ prometheus-node-exporter --version
node_exporter, version 0.15.2+ds (branch: debian/sid, revision: 0.15.2+ds-1)
 build user:      pkg-go-maintainers@lists.alioth.debian.org
 build date:      20171214-15:26:08
 go version:      go1.9.2
basevm@vboxbasevm:~$
```

- Configure Prometheus

```
cd /etc/prometheus
```

```
nano prometheus.yml
```

```
global:
```

```
scrape_interval: 15s
```

```
scrape_configs:
```

```
- job_name: "prometheus"
```

```
static_configs:
```

```
- targets: ["localhost:9090"]
```

- **Start Node Exporter**

```
systemctl start prometheus-node-exporter
systemctl enable prometheus-node-exporter
```

- **Verify prometheus is working**

```
curl http://localhost:9090/metrics
```

Step 3: Install Grafana for Visualization

- **Install Grafana**

```
sudo apt install grafana -y
sudo systemctl start grafana-server
sudo systemctl enable grafana-server
```

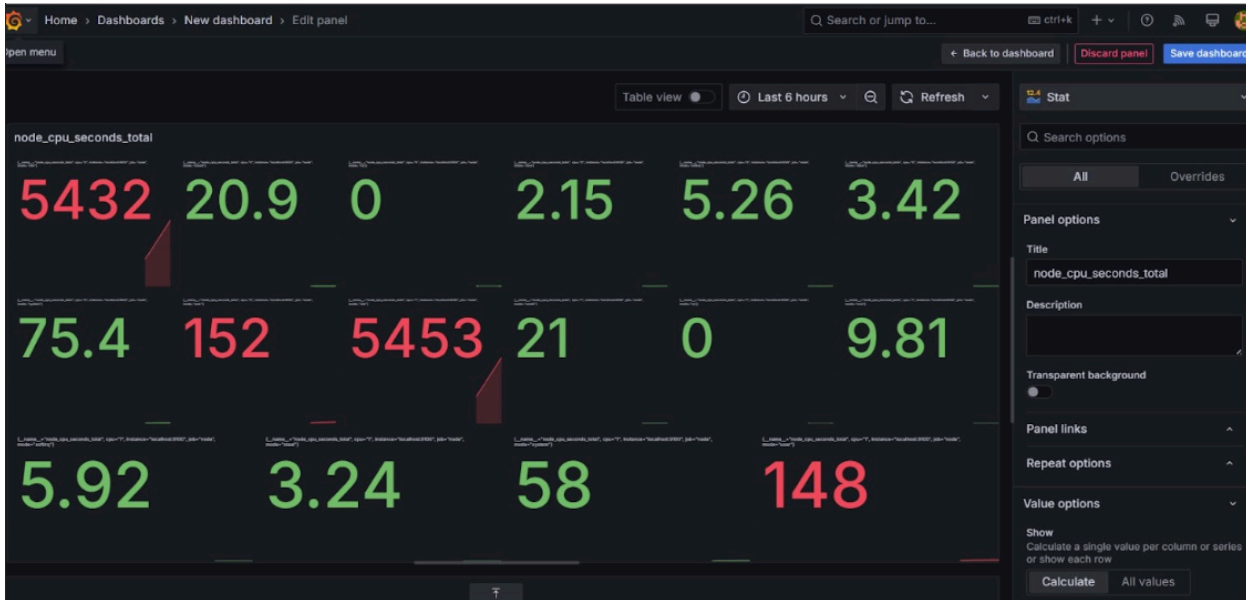
```
basevm@vboxbasevm:~$ sudo systemctl start grafana-server
basevm@vboxbasevm:~$ sudo systemctl enable grafana-server
Synchronizing state of grafana-server.service with SysV service script with /lib/systemd/systemd-sysv-install.
Executing: /lib/systemd/systemd-sysv-install enable grafana-server
basevm@vboxbasevm:~$ sudo systemctl status grafana-server
● grafana-server.service - Grafana instance
   Loaded: loaded (/usr/lib/systemd/system/grafana-server.service; enabled; vendor preset: enabled)
   Active: active (running) since Sun 2025-03-23 01:15:52 IST; 1h 17min ago
     Docs: http://docs.grafana.org
```

- Once grafana is up and running setup the grafana dashboard

- Open a browser and go to:

<http://localhost:3000>

- Login using admin as username and admin as password
- Navigate to configuration and add data source and add prometheus
- Provide the prometheus url
- Click save and test
- Then create a dashboard and visualise the dashboard



Step 4: Create VM in google cloud

- Login into GCP
- Navigate to compute engine >> vm instance
- Provide the details like name, operating system, ram, disk size, zone and allow http and https
- Click on create

| VM instances | | | | | | | |
|---|--------|---------------|-----------------|-----------|--------------------------------------|---|---------|
| Create instance Import VM Refresh Learn | | | | | | | |
| Instances Observability Instance schedules | | | | | | | |
| VM instances | | | | | | | |
| Filter <input type="text"/> Enter property name or value | | | | | | | |
| <input type="checkbox"/> Status | Name ↑ | Zone | Recommendations | In use by | Internal IP | External IP | Connect |
| <input checked="" type="checkbox"/> | vm2 | us-central1-c | | | 10.128.0.2 (nic0) | 35.222.19... (nic0) | SSH |

Step 5: Enable SSH Key-Based Authentication

- Once the vm2 is created authenticate it via ssh key using the following commands.
- Copy the key generated

```
ssh-keygen -t rsa -b 2048 -f ~/.ssh/gcp_key
cat ~/.ssh/gcp_key.pub
```

```
basevm@boxbasevm:~$ cat ~/.ssh/gcp_key.pub
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDAQX8VTc8M6UYUJA0giPdcJHF/kmnnjXZr16LU4Ci40SLMP8wmugvNlnxvmH/pqWQEkWlJ3Xx2v+91+h1eESCyn9f+ANFhwHokCvYH81dNmHQyuz57c101xj55ebb1yFvOra5U
1DOKAZtnQY1DKDtdgldSU1et2AiPwn10ccyQUCj/SPKk5XiggNoVy7e0z8eN8TdIAOYCTtcYoktbAZg4izGYGSXrumXfkRw89R9B4oT0nW1IoRUwgdIp0nQGi/W3LWYWhq5ESYNrnXvjUQZC9UR9Ka40LX13uUmaBqEXAXZgrbR
yBK05AX/VXv3o25gYD8gT177gQvFGS+30gQh basevm@boxbasevm
```

- Add the ssh key created in local vm (basevm) to the ssh key in your cloud vm (vm2)
- Open the vm2
- Click edit
- Navigate to ssh key then add your copied ssh key and authenticate

vm2
Edit
Reset
Create machine image
Create similar
⋮
Equivalent code
Learn

Details
Observability
OS Info
Screenshot

SSH Keys

SSH keys

| Username | Key |
|-----------|------------------------|
| g24ai1012 | ecdsa-sha2-nistp256... |
| g24ai1012 | ssh-rsa... |
| basevm | ssh-rsa... |

Step 6: Create an Instance Template

- In gcp navigate to compute engine >> instance template
- Click create instance template
- Provide all the machine configurations

Instance tem...
Create instance template
Refresh
Create VM
Show info panel

Instance templates are saved VM configurations used to create identical VMs, either individually or as part of managed instance groups. [Learn more](#)

Filter
Filter instance templates

| <input type="checkbox"/> | Name ↑ | Machine type | Image | Disk type | Location ? | Placement policy ? | In u | Actions |
|--------------------------|-------------------------|--------------|-----------------------------|--------------------------|------------|--------------------|------|---------|
| <input type="checkbox"/> | vm-auto | e2-medium | ubuntu-2004-focal-v20250313 | Balanced persistent disk | global | No policy | | ⋮ |

- Create a Managed Instance Group for auto scaling
- Navigate to compute engine >> instance group
- Select new managed instance group
- Provide the name, select instance template created earlier

- Set initial instance count 1
- Choose auto scaling on
- Set Minimum instances to 1 and Maximum instances to 3
- select Scaling based on CPU usage
- Set CPU utilization target to 75%
- Click Create

The screenshot shows the Google Cloud Console interface for 'Instance groups'. The left sidebar lists navigation options like 'Overview', 'Virtual machines', and 'Instance templates'. The main content area shows a table of instance groups. One group, 'auto-instance', is listed with 1 instance, using the 'vm-auto' template, in a 'Managed' group type. A status bar indicates 'Instance group is ready'. A bottom notification confirms the successful creation of the 'auto-instance' group.

| Status | Name | Instances | Template | Group type | Creation time | Recommendation | Autoscaling |
|-------------------------------------|-------------------------------|-----------|-------------------------|------------|------------------------------------|----------------|------------------------|
| <input checked="" type="checkbox"/> | auto-instance | 1 | vm-auto | Managed | Mar 23, 2025, 5:57:19 PM UTC+05:30 | | On: Target utilization |

Step 7: Configure Auto-Scaling Alerts in Prometheus

- Open local vm

```
sudo nano /etc/prometheus/alert.rules.yml
```

- Edit the file alert.rules.yml

```
basevm@vboxbasevm:~$ cat /etc/prometheus/alert.rules.yml
groups:
- name: auto-scale-alerts
  rules:
  - alert: HighCPUUsage
    expr: 100 - (avg by (instance) (rate(node_cpu_seconds_total{mode="idle"}[5m])) * 100) > 75
    for: 2m
    labels:
      severity: critical
    annotations:
      summary: "High CPU Usage Detected"
```

- Reload prometheus

```
curl -X POST http://localhost:9090/-/reload
```

Step 8: Verify Auto-Scaling Behavior

- Install stress in local vm (basevm)
- Apply stress by executing
- Stress -cpu 2 -timeout 300
- Check logs in prometheus

```
basevm@vboxbasevm:~$ sudo apt install stress -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  stress
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.
Need to get 17.5 kB of archives.
After this operation, 46.1 kB of additional disk space will be used.
Get:1 http://in.archive.ubuntu.com/ubuntu bionic/universe amd64 stress amd64 1.0.4-2 [17.5 kB]
Fetched 17.5 kB in 1s (13.2 kB/s)
Selecting previously unselected package stress.
(Reading database ... 70518 files and directories currently installed.)
Preparing to unpack .../stress_1.0.4-2_amd64.deb ...
Unpacking stress (1.0.4-2) ...
Setting up stress (1.0.4-2) ...
Processing triggers for man-db (2.8.3-2ubuntu0.1) ...
Processing triggers for install-info (6.5.0.dfsg.1-2) ...
basevm@vboxbasevm:~$ stress --cpu 2 --timeout 300
stress: info: [5483] dispatching hogs: 2 cpu, 0 io, 0 vm, 0 hdd
```

- Check logs in prometheus and grafana
- Open gcp console
- You can see multiple vm instance created after stress was applied

| VM instances | | | | | | | | | |
|---|------------------------------------|---------------|-----------------|-----------------------|----------------------|---|---------|--|--|
| Create instance Import VM Refresh Learn | | | | | | | | | |
| Instances Observability Instance schedules | | | | | | | | | |
| VM instances | | | | | | | | | |
| Filter Enter property name or value | | | | | | | | | |
| <input type="checkbox"/> Status | Name ↑ | Zone | Recommendations | In use by | Internal IP | External IP | Connect | | |
| <input type="checkbox"/> | auto-instance-1xv1 | us-central1-c | | auto- | 10.128.0.3 (nic0) | 34.56.123.45 34.56.123.45 (nic0) | SSH | | |
| <input type="checkbox"/> | auto-instance-9kgf | us-central1-c | | auto- | 10.128.0.4 (nic0) | 34.56.123.45 34.56.123.45 (nic0) | SSH | | |
| <input type="checkbox"/> | auto-instance-xml3 | us-central1-c | | auto- | 10.128.0.5 (nic0) | 34.56.123.45 34.56.123.45 (nic0) | SSH | | |
| <input type="checkbox"/> | vm2 | us-central1-c | | | 10.128.0.2 (nic0) | 35.245.123.45 35.245.123.45 (nic0) | SSH | | |

Google Cloud VCC 2025 Search (/) for resources, docs, products, and more Search

Compute Engine / Instance groups / Instance group: auto-instance

Overview auto-instance Edit Update VMs Restart/replace VMs Delete Group

Virtual machines

- VM instances
- Instance templates
- Sole-tenant nodes
- Machine images
- TPUs
- Committed use discou...
- Reservations
- Migrate to Virtual Mach...
- Marketplace
- Release Notes

Overview Details Monitoring Errors

Instances by status
3 instances 1
3

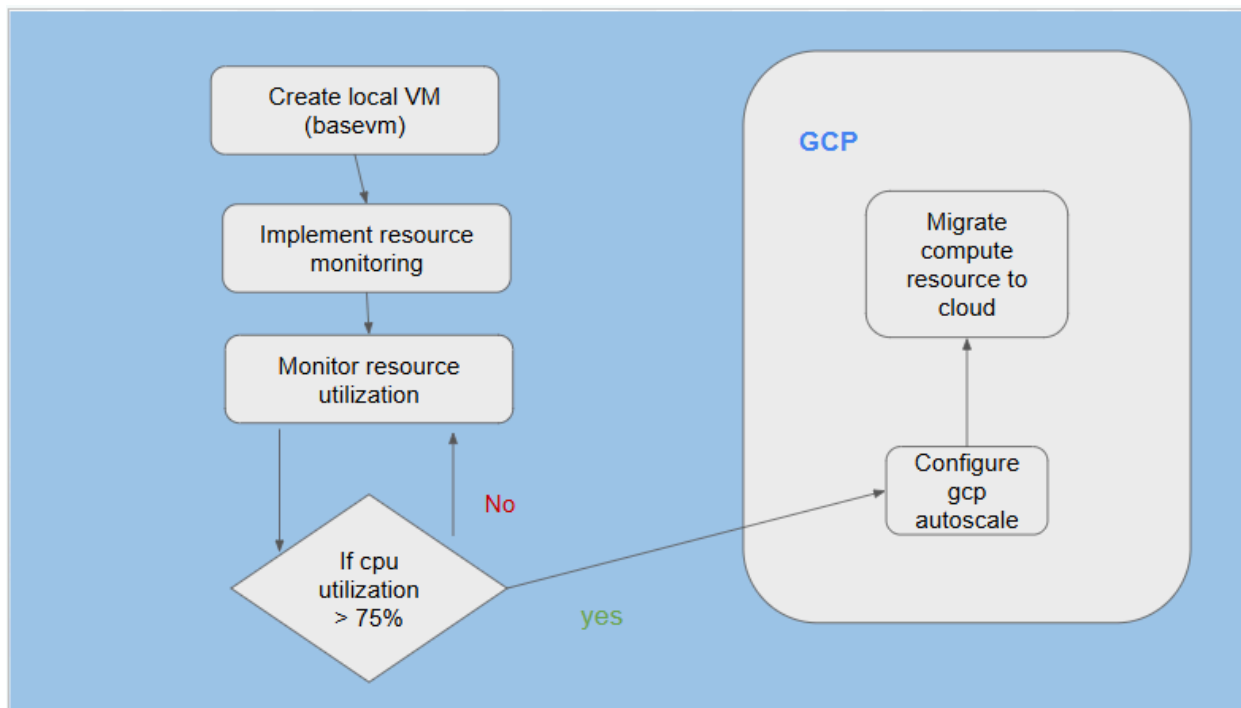
Instance by health ?
Not configured
Autohealing off. [Configure](#)

Autoscaling
On (min 1, max 3)
Based on 1 metric and 0 schedules 1

Status Ready
Creation Time Mar 23, 2025, 5:57:19 PM UTC+05:30
Description
Target running size 3
Template [vm-auto](#)
Location us-central1-c
Resize requests None

VM instances

Architecture Diagram



Link to Recorded Video Demo:

Link to Source Code Repo:

Thank you!