



Compute: Configure Metric-Based Autoscaling

Lab 09-1 Practices

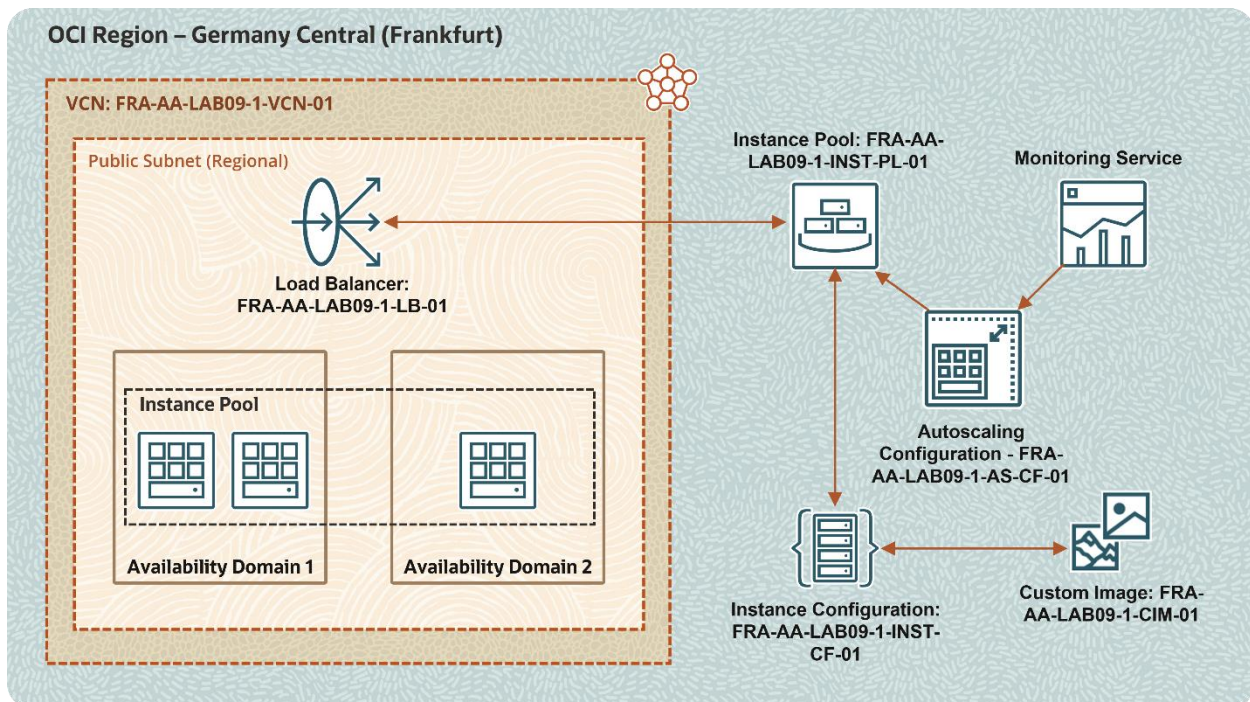
Get Started

Overview

Autoscaling lets you automatically adjust the number or the lifecycle state of compute instances in an instance pool. This helps you provide consistent performance for your end users during periods of high demand, and reduce your costs during periods of low demand.

In this lab, you will:

- Create a Virtual Cloud Network (VCN) and its components
- Create a load balancer
- Create a compute instance and a custom image
- Create an instance configuration
- Create an instance pool
- Create a metric-based autoscaling configuration
- Test autoscaling



Prerequisites

- You must have access to the OCI Console.

Assumptions

- You must be familiar with navigating the OCI Console.
- In this lab we are using Germany Central (Frankfurt) region.

Create a Virtual Cloud Network and Its Components

In this practice, you will create a Virtual Cloud Network (VCN), subnet, and Internet gateway, and add route rules in the route table.

Tasks

1. Sign in to your Oracle Cloud Infrastructure (OCI) Console.
2. From the navigation menu, under **Networking**, select **Virtual Cloud Networks**.
3. Click **Create VCN**.
4. In the **Create a Virtual Cloud Network** dialog box, populate the following information:

- **Name:** FRA-AA-LAB09-1-VCN-01
- **Create In Compartment:** *<your compartment>*
- **IPv4 CIDR Blocks:** 10.0.0.0/16

5. Keep all the other options default and click **Create VCN**.

Note: You can see that the VCN is created successfully.

6. Click **FRA-AA-LAB09-1-VCN-01** to view the details and click **Create Subnet**.

7. In the **Create Subnet** dialog box, populate the following information:

- **Name:** FRA-AA-LAB09-1-SNET-01
- **Create In Compartment:** *<your compartment>*
- **Subnet Type:** Regional
- **IPv4 CIDR Blocks:** 10.0.1.0/24
- **Subnet Access:** Public Subnet

8. Keep all the other options default and click **Create Subnet**.
9. Under **Resources** in the left navigation panel, click **Internet Gateways**.

10. Click **Create Internet Gateway** and populate the following information:

- **Name:** FRA-AA-LAB09-1-IG-01
- **Create In Compartment:** <your compartment>

11. Click **Create Internet Gateway**.

You can see that the internet gateway is created successfully, and the state is Available.

12. Under **Resources** in the left navigation panel, click **Route Tables**.

13. Click **Default Route Table** for FRA-AA-LAB09-1-VCN-01.

14. Click **Add Route Rules** and populate the following information:

- **Target Type:** Internet Gateway
- **Destination CIDR Block:** 0.0.0.0/0
- **Target Internet Gateway:** FRA-AA-LAB09-1-IG-01

15. Click **Add Route Rules**.

16. Using the breadcrumb list at the top of the screen, return to the VCN page by selecting **FRA-AA-LAB09-1-VCN-01**.

17. Under **Resources** in the left navigation panel, click **Security Lists**.

18. Click **Default Security List** for FRA-AA-LAB09-1-VCN-01.

19. Click **Add Ingress Rule** and populate the following information:

- **Source Type:** CIDR
- **Source CIDR:** 0.0.0.0/0
- **IP Protocol:** TCP
- **Source Port Range:** All
- **Destination Port Range:** 80

Note: Do not select Stateless.

20. Click **Add Ingress Rules**.

Create a Load Balancer

The Oracle Cloud Infrastructure (OCI) Load Balancer provides automated traffic distribution from one entry point to multiple servers reachable from your VCN. A load balancer improves resource utilization, facilitates scaling, and helps ensure high availability.

In this practice, you will create a public load balancer.

Tasks

1. From the navigation menu, under **Networking**, select **Load Balancers**.
2. Click **Create Load Balancer**.
3. Select **Load Balancer Type** as **Load Balancer** and click **Create Load Balancer**.
4. In the **Add Details** section, populate the following information:
 - **Load Balancer Name:** FRA-AA-LAB09-1-LB-01
 - **Choose visibility type:** Public
 - **Assign a public IP address:** Ephemeral IP Address
 - **Shapes:** Flexible Shapes
 - **Choose the minimum bandwidth:** 10
 - **Choose the maximum bandwidth:** 20
 - **Virtual Cloud Network in <your compartment>:** FRA-AA-LAB09-1-VCN-01
 - **Subnet in <your compartment>:** FRA-AA-LAB09-1-SNET-01 (regional)
5. Click **Next**.
6. In the **Choose Backends** section, select **Weighted Round Robin** in the **Specify a Load Balancing Policy** field.

Note: Do not click **Add Backends** and keep the **Specify Health Check Policy** field default.
7. Click **Show Advanced Options** and enter FRA-AA-LAB09-1-LB-BS-01 in the **Backend Set Name** field.

8. Click **Next**.
9. In the **Configure Listener** section, populate the following information:
 - **Listener Name:** FRA-AA-LAB09-1-LB-LS-01
 - **Specify the type of traffic your listener handles:** HTTP
 - **Specify the port your listener monitors for ingress traffic:** 80
10. Click **Next**.
11. In the **Manage Logging** section, disable **Error Logs** and **Access Logs**.
12. Click **Submit**.

Create a Compute Instance and a Custom Image

In this practice, you will create SSH keys, launch a compute instance, install Apache HTTP server, and create a custom image.

Tasks

1. Click the **Cloud Shell** icon in the console header next to the Region icon.
2. After the Cloud Shell has started, run the following command:

```
$ mkdir .ssh
```

Important: In case you get an error that says, “cannot create director: File exists”, you can skip running the first command.

```
$ cd .ssh
```

```
$ ssh-keygen -b 2048 -t rsa -f <<sshkeyname>>
```

Replace `<sshkeyname>` with **ociaalab9key**. Select the key name you can remember. This will be the key name you will use to connect to the compute instance you create.

Remember:

- After entering the third command, press **Enter** twice for no passphrase.
- Do not include the angle brackets «» and \$ symbol when pasting code into Cloud Shell.

3. Examine the two files that you just created by running the following command:

```
$ ls
```

Note: In the output there are two files, a private key: `<sshkeyname>` and a public key: `<sshkeyname>.pub`. Keep the private key safe and don't share its content with anyone. The public key will be needed for various activities and can be uploaded to certain systems as well as copied and pasted to facilitate secure communications in the cloud.

4. To list the contents of the public key, run the following command:

```
$ cat <<sshkeyname>>.pub
```

Replace `<<sshkeyname>>` with **ociaalab9key**.

5. Copy the contents of the public key as you will need this in a subsequent step. Make sure that you remove any hard returns that may have been added when copying. The `.pub` key should be one line.

6. From the navigation menu, under **Compute**, click **Instances**.
7. Click **Create instance** and populate the following information:

- **Name:** FRA-AA-LAB09-1-VM-01
- **Create in compartment:** *<your compartment>*
- **Availability Domain:** AD 1

Click **Show advanced options** and select **On-demand capacity** under Capacity type.

- **Image:** Oracle Linux 8
- **Shape:** Select VM.Standard.A1.Flex (1 OCPU, 6GB Memory) [Shape series: Ampere]
- **Primary Network:** Select existing virtual cloud network.
- **Virtual cloud network in *<your compartment>*:** FRA-AA-LAB09-1-VCN-01
- **Subnet:** Select existing subnet.
- **Subnet in *<your compartment>*:** FRA-AA-LAB09-1-SNET-01 (regional)
- **Public IP address:** Assign a public IPv4 address.
- **Add SSH keys:** Paste public keys.
- **SSH Keys:** *<contents of the public key>* (which is copied in Step 5 of this practice)

8. Keep the **Boot Volume** default and click **Create**.

Note: In a couple of minutes, you will see that the Instance is created successfully, and the state is Running.

9. Open **Cloud Shell** and log in to your instance by running the following command:

```
$ ssh -i <private_key_file> <username>@<public-ip-address>
```

Remember:

- *<private_key_file>* is the full path and name of the file that contains the private key associated with the instance you want to access.
- *<username>* is the default user `opc`.

- `<public-ip-address>` is the public IP address of the instance.

Note: Enter `yes` in response to “Are you sure you want to continue connecting (yes/no)?”.

10. While connected to your Compute instance via SSH, run the following commands:

- Install Apache http:

```
$ sudo yum install httpd -y
```
- Start the Apache server and configure it to start after system:

```
$ sudo apachectl start  
$ sudo systemctl enable httpd
```
- Run a quick check on Apache configurations:

```
$ sudo apachectl configtest
```
- Create firewall rules to allow access to the ports on which the HTTP server listens:

```
$ sudo firewall-cmd --permanent --zone=public --add-service=http  
$ sudo firewall-cmd --reload
```
- Create an index file for your web server:

```
$ sudo bash -c 'echo $(hostname) >> /var/www/html/index.html'
```

11. Now that you have a compute instance and Apache web server installed, you need to create a custom image from it.

12. From the navigation menu, under **Compute**, select **Instances**.

13. Click the **FRA-AA-LAB09-1-VM-01** instance.

14. From the **More Actions** drop-down list, select **Create custom image**.

15. In the **Create custom image** dialog box, populate the following information:

- **Create in compartment:** `<your compartment>`
- **Name:** `FRA-AA-LAB09-1-CIM-01`

16. Click **Create custom image**.

Note: Now, you need to create a new compute instance based on the custom image. Once the custom image has been successfully created, you have to delete the instance **FRA-AA-LAB09-1-VM-01** as it is no longer required.

17. From the navigation menu, under **Compute**, select **Instances**.

18. Click **Create instance** and populate the following information:

- **Name:** FRA-AA-LAB09-1-VM-02
- **Create in compartment:** <your compartment>
- **Availability Domain:** AD 1

Click **Show advanced options** and select **On-demand capacity** under Capacity type.

- **Image:** FRA-AA-LAB09-1-CIM-01

Note: To select the custom image, click **Change image** and select **Custom images** in the **Image source** field. Select <your compartment> in the **Compartment** field. Then, select the custom image you just created from the list.

- **Shape:** Select VM.Standard.A1.Flex (1 OCPU, 6GB Memory) [Shape series: Ampere]
- **Primary Network:** Select existing virtual cloud network.
- **Virtual cloud network in <your compartment>:** FRA-AA-LAB09-1-VCN-01
- **Subnet:** Select an existing subnet.
- **Subnet in <your compartment>:** FRA-AA-LAB09-1-SNET-01 (regional)
- **Public IP address:** Assign a public IPv4 address.
- **Add SSH keys:** Paste public keys.
- **SSH Keys:** <contents of the public key> (which is copied in Step 5 of this practice)

19. Keep the Boot Volume default and click **Create**.

Create an Instance Configuration

The instance configurations let you define the settings to use when creating compute instances.

In this practice, you will create an instance configuration and subsequently use it to create one or more instances in an instance pool.

Tasks

1. From the navigation menu, under **Compute**, select **Instances**.
2. Click **FRA-AA-LAB09-1-VM-02** (the instance of an image you want to use as a template to create the instance configuration).
3. From the **More Actions** drop-down list, select **Create instance configuration**.
4. In the **Create instance configuration** dialog box, populate the following information:
 - **Create in compartment:** *<your compartment>*
 - **Name:** FRA-AA-LAB09-1-INST-CF-01
5. Click **Create instance configuration**.

Note: You will see that the instance configuration is created successfully.

Create an Instance Pool

The instance pools let you create and manage multiple compute instances within the same region as a group. Before you create an instance pool, you need an instance configuration and optionally a load balancer and back-end set.

In this practice, you will create an instance pool.

Tasks

1. From the navigation menu, under **Compute**, select **Instance Pools**.
2. Click **Create instance pool**.
3. In the **Add basic details** section, populate the following information:
 - **Name:** FRA-AA-LAB09-1-INST-PL-01
 - **Create in compartment:** *<your compartment>*
 - **Instance configuration in <your compartment>:** FRA-AA-LAB09-1-INST-CF-01
 - **Number of instances:** 2
4. Click **Next**.
5. In the **Configure pool placement** section, you will select the location where you want to place the instances. Populate the following information:
 - **Availability domain:** AD 1
 - **Select a virtual cloud network in <your compartment>:** FRA-AA-LAB09-1-VCN-01
 - **Select a subnet in <your compartment>:** FRA-AA-LAB09-1-SNET-01

Note: You can leave the Fault domains field blank.
6. To create instances in more than one availability domain, click **+ Another availability domain** and populate the following information:
 - **Availability domain:** AD 2
 - **Select a virtual cloud network in <your compartment>:** FRA-AA-LAB09-1-VCN-01

- **Select a subnet in <your compartment>:** FRA-AA-LAB09-1-SNET-01

Note: You can leave the Fault domains field blank.

7. Select the **Attach a load balancer** check box and populate the following information:

- **Load balancer type:** Load Balancer
- **Load balancer in <your compartment>:** FRA-AA-LAB09-1-LB-01
- **Backend set:** FRA-AA-LAB09-1-LB-BS-01
- **Port:** 80

Note: This field is the server port on the instances to which the load balancer must direct traffic.

- **VNIC:** Primary VNIC

8. Click **Next**.

9. Review the instance pool details and click **Create**.

Note: You will see that the instance pool creation is successful.

10. Under **Resources** in the left navigation panel, click **Attached instances**. You should see two instances in the Running state.

Create a Metric-Based Autoscaling Configuration

In metric-based autoscaling, you select a performance metric to monitor, and set thresholds that the performance metric must reach to trigger an autoscaling event. When system usage meets a threshold, autoscaling dynamically resizes the instance pool in near-real time. As load increases, the pool scales out. As load decreases, the pool scales in.

In this practice, you will create a metric-based autoscaling configuration.

Tasks

1. From the navigation menu, under **Compute**, click **Autoscaling Configurations**.
2. Click **Create autoscaling configuration**.
3. In the **Add basic details** section, populate the following information:
 - **Name:** FRA-AA-LAB09-1-AS-CF-01
 - **Create in compartment:** *<your compartment>*
 - **Instance Pool:** FRA-AA-LAB09-1-INST-PL-01
4. Click **Next**.
5. In the **Configure autoscaling policy** section, select **Metric-based autoscaling**. Then populate the following information:
 - **Autoscaling policy name:** FRA-AA-LAB09-1-AS-POL-01
 - **Cooldown in seconds:** 300
 - **Performance metric:** CPU utilization
6. In the **Scale-out rule** subsection, populate the following information:
 - **Scale-out operator:** Greater than (>)
 - **Threshold percentage:** 70
 - **Number of instances to add:** 1

7. In the **Scale-in rule** subsection, populate the following information:

- **Scale-in operator:** Less than (<)
- **Threshold percentage:** 20
- **Number of instances to remove:** 1

8. In the **Scaling limits** subsection, populate the following information:

- **Minimum number of instances:** 1
- **Maximum number of instances:** 3
- **Initial number of instances:** 2

9. Click **Next**.

10. Review the autoscaling configuration and click **Create**.

Note: You will see that the autoscaling configuration is created successfully and the state is Enabled.

Test Autoscaling

In the metric-based autoscaling, you choose a performance metric to monitor, and set thresholds that the performance metric must reach to trigger an autoscaling event. When system usage meets a threshold, autoscaling dynamically resizes the instance pool in near-real time. As load increases, the pool scales out. As load decreases, the pool scales in.

In this practice, you will test a metric-based autoscaling configuration.

Tasks

1. From the navigation menu, under **Compute**, select **Instance Pools**.
2. Click **FRA-AA-LAB09-1-INST-PL-01**.
3. Under **Resources** in the left navigation panel, click **Attached instances**.

Note: After a few minutes, you will notice that one of the instances is terminated due to the scale-in rule you defined. Please note that initially two instances were in Running state.

Note: You might have to wait for a couple of minutes before the scale in occurs.

4. Click the instance. (You will see only one instance here.)
5. Open Cloud Shell and use SSH to log in to your instance by running the following command:

```
$ ssh -i <private_key_file> <username>@<public-ip-address>
```

Remember:

- *<private_key_file>* is the full path and name of the file that contains the private key associated with the instance you want to access.
- *<username>* is the default user `opc`.
- *<public-ip-address>* is the public IP address of the instance.

Note: Enter `yes` in response to “Are you sure you want to continue connecting (yes/no)?”

6. To install the stress package, run the following command:

```
$ sudo dnf makecache
$ sudo dnf install stress-ng-0.14.00-1.el8.aarch64
```

7. To generate stress, run the following command:

```
$ stress-ng --cpu 2 -t 5m
```

8. From the navigation menu, under **Compute**, select **Instance Pools**.
9. Click **FRA-AA-LAB09-1-INST-PL-01**.
10. Under **Resources** in the left navigation panel, click **Attached instances**.

Note: After a few minutes, you will notice that one of the instances will be added to the instance pool due to the scale-out rule you defined.