

# Autoscalling infrastruktur oracle cloud

## Contents

1. Pengantar .....	2
2. Arsitektur .....	2
3. Membuat Konfigurasi Instance .....	3
4. Membuat instance pool .....	4
5. Membuat konfigurasi Autoscalling.....	5
6. Tugas.....	8

## 1. Pengantar

Kemampuan autoscaling server merupakan kemampuan yang sangat penting dimiliki oleh infrastruktur komputasi yang memiliki trafik yang sering berubah-ubah.

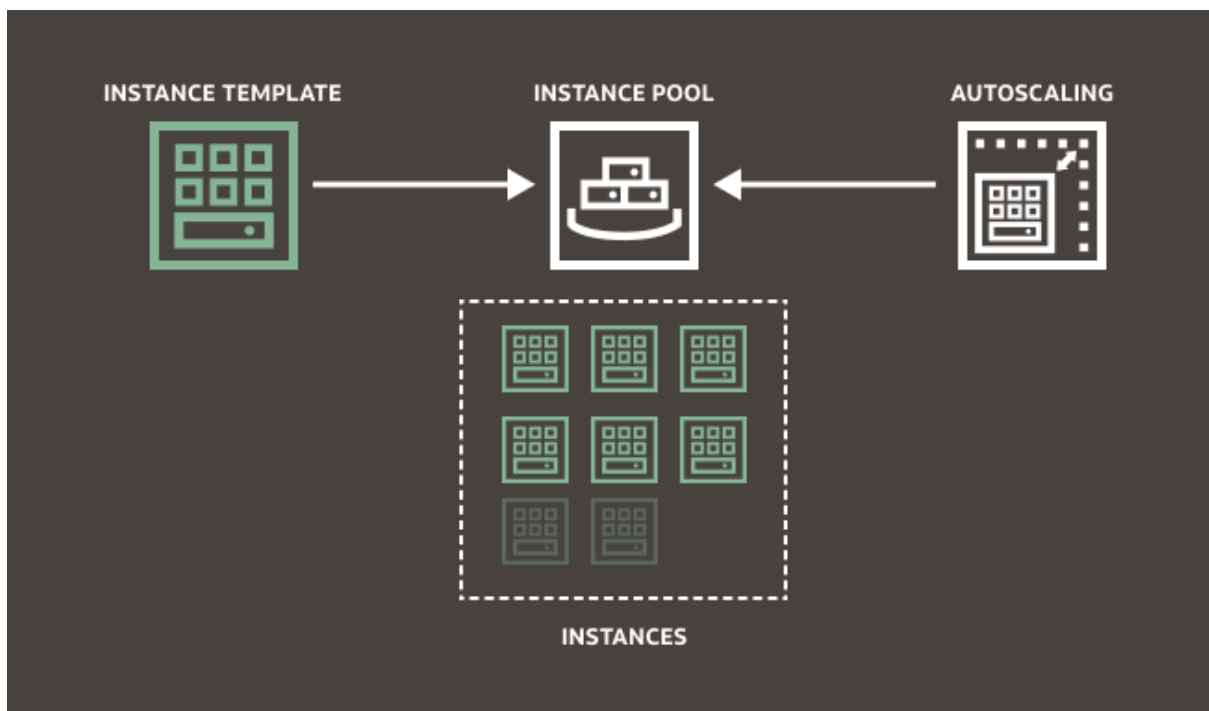
Misalnya suatu toko online, dimana trafik saat malam rendah akibat pengguna relatif tidak mengakses toko online saat malam. Namun, ketika sudah mulai pagi sekitar pukul 06.00 pagi trafiknya meningkat hingga mencapai puncaknya saat pukul 10.00. Saat weekend pola trafik pun mirip dengan hari biasa, hanya saja dengan beban lebih tinggi. Dan mencapai beban trafik tertinggi ketika saat musim liburan.

Untuk menghandle variasi kebutuhan komputasi tersebut, toko online bisa saja menyediakan infrastruktur komputasi yang cukup untuk menghandle beban saat trafik sedang tinggi-tingginya. Akan tetapi menjalankan infrastruktur komputasi tersebut sepanjang waktu tentu sangat pemborosan. Karena kebutuhan infrastruktur komputasi yang sedemikian tinggi barangkali hanya di waktu-waktu tertentu yang tidak cukup sering terjadi.

Maka, dengan autoscaling, barangkali infrastruktur komputasi yang dijalankan akan bisa menyesuaikan dengan kebutuhan komputasi pada saat itu. Sehingga biaya yang diperlukan tentu akan lebih hemat.

## 2. Arsitektur

Berikut komponen dalam kemampuan autoscaling

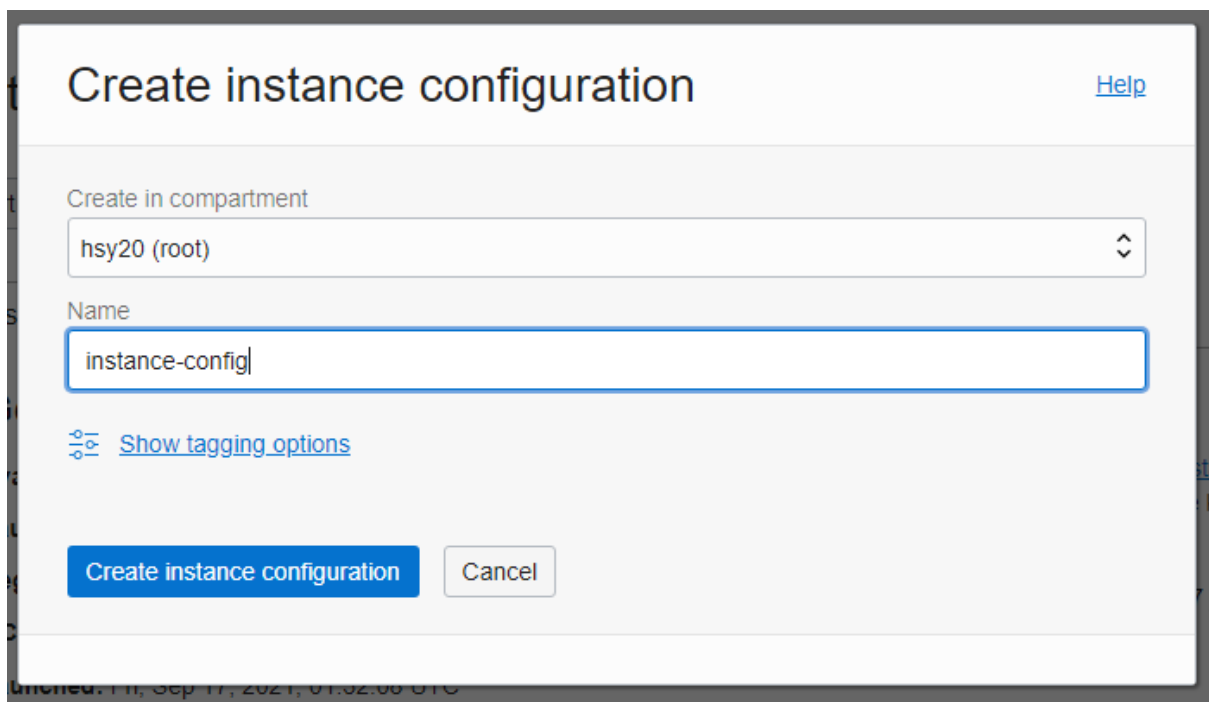


Terdapat instance template, yang akan digunakan untuk membangkitkan sebuah instance secara otomatis yang kelak akan menjadi kolam instance. Kolam instance inilah yang kelak sebagai persediaan sumberdaya dalam melakukan autoscaling.

### 3. Membuat Konfigurasi Instance

Kita akan gunakan instance yang telah kita miliki sebagai template untuk konfigurasi instance. Jika belum memiliki instance, buatlah instance komputasi seperti yang telah dilakukan pada pertemuan awal.

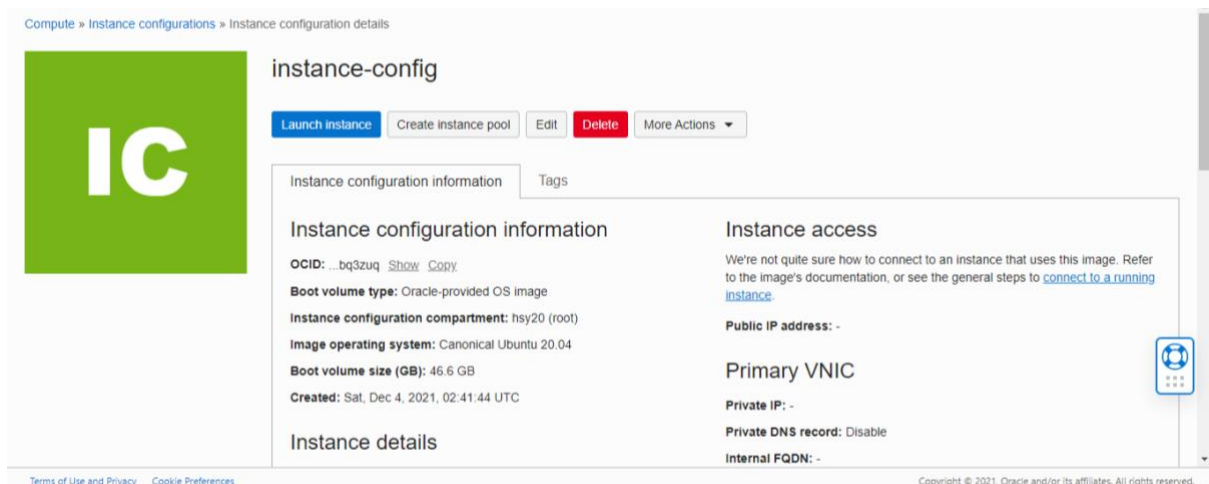
1. Pada halaman detail instance, klik menu More Actions dan pilih Create Instance Configuration.
2. Pilih kompartemen Anda.
3. Untuk nama, masukkan instance-config.



The screenshot shows a 'Create instance configuration' dialog box. At the top, it says 'Create in compartment' with a dropdown menu showing 'hsy20 (root)'. Below that is a 'Name' field with the text 'instance-config' entered. There is a link 'Show tagging options' with a tag icon. At the bottom, there are two buttons: 'Create instance configuration' (blue) and 'Cancel' (grey).

4. Click Create Instance Configuration.

Setelah Anda membuat konfigurasi instance, halaman detailnya akan ditampilkan, seperti yang ditunjukkan pada gambar berikut.



The screenshot shows the 'Instance configuration details' page for 'instance-config'. On the left is a green square with 'IC' in white. The main content area has a tab 'Instance configuration information' selected. Below the tab, there is a section 'Instance configuration information' with the following details: OCID: ...bq3zuq, Boot volume type: Oracle-provided OS image, Instance configuration compartment: hsy20 (root), Image operating system: Canonical Ubuntu 20.04, Boot volume size (GB): 46.6 GB, and Created: Sat, Dec 4, 2021, 02:41:44 UTC. To the right of this section is a section 'Instance access' with the text: 'We're not quite sure how to connect to an instance that uses this image. Refer to the image's documentation, or see the general steps to connect to a running instance.' Below this, there are fields for 'Public IP address: -', 'Primary VNIC', 'Private IP: -', 'Private DNS record: Disable', and 'Internal FQDN: -'. At the bottom, there is a section 'Instance details'.

## 4. Membuat instance pool

Anda dapat membuat kolam instans langsung dari halaman konfigurasi instans dengan mengikuti langkah-langkah berikut:

1. klik **Create Instance pool**
2. Dari daftar Kompartemen, pilih kompartemen tempat Anda membuat konfigurasi instans.
3. Untuk nama, masukkan my-instance-pool.
4. Dari daftar Konfigurasi Instance, pilih instance-config.
5. Untuk number of instans, masukkan 3. Jumlah instans menunjukkan jumlah maksimum instans yang dapat disediakan di kolam instans. Jumlah instans yang dapat disediakan bergantung pada batas layanan penyewa Anda dan ketersediaan bentuk komputasi di wilayah Anda.

Create instance pool [Help](#)

1 Add basic details

2 Configure pool placement

3 Review

Name

my-instance-pool

Create in compartment

hsy20 (root)

Instance configuration in hsy20 (root) [\(Change Compartment\)](#)

instance-config

Number of instances

3

The number of instances that can be provisioned is limited by your tenancy's service limits.

> Instance configuration details

[Show tagging options](#)

Next

Cancel

### 6. Klik next

Pada halaman **Configure Pool Placement**, Anda dapat menambahkan satu atau beberapa availability domain. Untuk setiap domain, Anda dapat menentukan fault domain, primary virtual cloud network (VCN), dan subnet. Secara default, instans dalam kolam didistribusikan di semua fault domain. Jika kapasitas tidak tersedia di satu fault domain, instans ditempatkan di fault domain lain untuk memungkinkan kumpulan instans berhasil diluncurkan. Anda dapat meminta agar instans didistribusikan secara merata di setiap fault domain yang Anda pilih untuk skenario ketersediaan tinggi.

Selain itu, Anda dapat mengaitkan **loadbalancer** dengan kumpulan instans dengan memilih kotak centang Lampirkan loadbalancer. Untuk menggunakan fitur ini, Anda harus memiliki loadbalancer dulu.

Untuk melanjutkan, pilih AD1 untuk availability domain, VCN dan subnet. Klik **Next** dan kemudian **Create** untuk membuat kolam instans.

## Create instance pool

[Help](#)

1 Add basic details

2 **Configure pool placement**

3 Review

Availability domain selection 1

Availability domain

AD 1

Primary VNIC

Select a virtual cloud network in **hsy20 (root)** [\(Change Compartment\)](#)

vcn-20210917-0827

Fault domains

FAULT-DOMAIN-1 x

Select a subnet in **hsy20 (root)** [\(Change Compartment\)](#)

subnet-20210917-0827 (regional)

+ Another availability domain

☐ Attach a load balancer

Previous

Next

Cancel

Untuk fault domain pilih FAULT-DOMAIN-1, FAULT-DOMAIN-2, FAULT-DOMAIN-3

Karena ada keterbatasan layanan penyewa nampaknya instance bisa dibuat hanya 1 buah.

ORACLE Cloud

Search for resources, services, and documentation

Australia Southeast (Melbourne)

PROVISIONING

OCID: ...eurbbq [Show](#) [Copy](#)

Instance configuration: [instance-config](#)

Created: Sat, Dec 4, 2021, 03:05:37 UTC

Target instance count: 2

Compartment: hsy20 (root)

Availability domain: AD-1

Fault domain: FAULT-DOMAIN-1, FAULT-DOMAIN-2, FAULT-DOMAIN-3

Primary VNIC virtual cloud network: [vcn-20210917-0827](#)

Primary VNIC subnet: [subnet-20210917-0827](#)

Secondary VNIC subnet: -

Resources

Metrics

**Attached instances**

[Load balancers](#)

[Work requests](#)

Attached instances

Attach instance

Search...

Name	State	Availability domain	Fault domain	Shape	Instance configuration	Attached or created
<a href="#">inst-gzjsj-my-instance-pool</a>	Running	AD-1	FD-1	VM.Standard.E2.1.Micro	<a href="#">instance-config</a>	Sat, Dec 4, 2021, 03:05:51 UTC

Showing 1 Item < Page 1 >

## 5. Membuat konfigurasi Autoscaling

Skenarionya adalah barangkali suatu saat instans yang ada tidak dapat menangani kebutuhan komputasi. Maka dibuat konfigurasi penskalaan otomatis yang secara otomatis menskalakan jumlah instans di kolam instance.

Kita akan buat konfigurasi penskalaan otomatis dari halaman detail instance pool.

1. Klik menu **More Actions** dan pilih **Create Autoscaling Configuration**.
2. Untuk nama, masukkan my-autoscaling-config.
3. Dari daftar Buat di kompartemen, pilih kompartemen tempat Anda membuat kumpulan instans.
4. Klik **Next**.

ORACLE Cloud Search for resources, services, and documentation Australia Southeast (Melbourne)

## Create autoscaling configuration

- 1 Add basic details
- 2 Configure autoscaling policy
- 3 Review

Name  
my-autoscaling-config

Create in compartment  
hsy20 (root)

Instance pool: my-instance-pool

[Show tagging options](#)

Next Cancel

Pada halaman **Configure Autoscaling Policy**, Anda dapat memilih **Metric-based Autoscaling** atau **Schedule-based Autoscaling**.

**Metric-based Autoscaling** penskalaan otomatis berbasis metrik, seperti CPU atau penggunaan memori untuk memicu peristiwa penskalaan otomatis. Anda kemudian dapat menentukan aturan scale-out dan scale-in. Aturan ini mengambil persentase ambang batas. Untuk memastikan bahwa instance tidak ditambahkan atau dihapus terlalu cepat, pengaturan cooldown dalam detik memastikan waktu minimum antara peristiwa pen-skalaan (300 detik, atau 5 menit). Gambar berikut menunjukkan contoh aturan penskalaan otomatis berbasis penggunaan CPU. Kebijakan ini menambahkan satu instance saat penggunaan CPU lebih besar dari 70%. Saat penggunaan CPU turun di bawah 40%, kebijakan penskalaan otomatis menghapus satu instans.

**Metric-based Autoscaling** digunakan saat Anda tidak dapat memprediksi jumlah lalu lintas, dan Anda ingin mengotomatiskan penskalaan berdasarkan penggunaan CPU atau memori. Ketika Anda dapat memprediksi permintaan atau Anda tahu bahwa peningkatan permintaan akan terjadi (acara peluncuran, misalnya), Anda menggunakan **Schedule-based Autoscaling**.

## Contoh Penskalaan otomatis berbasis penggunaan CPU

ORACLE Cloud Search for resources, services, and documentation Australia Southeast (Melbourne) Help

### Create autoscaling configuration

[Help](#)

1 Add basic details  
2 **Configure autoscaling policy**  
3 Review

**Metric-based autoscaling**  
Performance metrics such as CPU utilization trigger autoscaling events. ✓

**Schedule-based autoscaling**  
Autoscaling events take place at the specific times that you schedule.

> Instance pool and instance configuration information

**Configure autoscaling policy**

Autoscaling policy name  
autoscaling-policy-20211204-1023

Cooldown in seconds ⓘ  
300  
Minimum value is 300 seconds.

Performance metric ⓘ  
CPU utilization

**Scale-out rule**

Scale-out operator  
Greater than (>)

Threshold percentage ⓘ  
70

Number of instances to add  
1

**Scale-in rule**

Scale-in operator  
Less than (<)

Threshold percentage ⓘ  
40

Number of instances to remove  
1

**Scaling limits**

Minimum number of instances  
1

Maximum number of instances ⓘ  
3

Initial number of instances  
2

Previous **Next** Cancel

Terms of Use and Privacy Cookie Preferences Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

**Schedule-based Autoscaling.** Misalnya, Anda merencanakan serangkaian acara peluncuran yang terjadi setiap hari Senin pukul 14.00, dari Januari hingga Maret. Anda ingin memastikan bahwa Anda memiliki 10 instans yang berjalan pada waktu tersebut. Gambar berikut menunjukkan cara mengonfigurasi kebijakan tersebut dengan menyetel ukuran kumpulan target ke 10 dan memasukkan nilai untuk menit (0), jam (14), hari dalam sebulan (?), bulan (1-3), hari minggu (2), dan tahun (\*).

Autoscaling Policy 1

AUTOSCALING POLICY NAME  
autoscaling-policy-9f3e775d-20201015-1138

TARGET POOL SIZE  
10

Execution schedule

Use a cron expression to schedule when the pool should scale to the target pool size for this policy.  
**All times are in UTC.**

QUICK START  
Custom

MINUTE  
0

HOUR  
14

DAY OF THE MONTH  
?

MONTH  
1-3

DAY OF THE WEEK  
2

YEAR  
\*

Summary: At 02:00 PM, only on Monday, January through March, UTC

- Setelah Anda mengonfigurasi Autoscaling policy, klik **Create**. Anda dapat menentukan beberapa konfigurasi penskalaan otomatis untuk kumpulan instance yang sama, dan Anda dapat mengaktifkan atau menonaktifkannya salah satunya.

ORACLE Cloud
Search for resources, services, and documentation
Australia Southeast (Melbourne)

Compute > Autoscaling configurations > Autoscaling configuration details

AS

ENABLED

my-autoscaling-config

Edit
Disable
Move resource
Add Tags
Delete

Autoscaling configuration information
Tags

OCID: ...h5qhnq
Show
Copy

Compartment: hsy20 (root)

Created: Sat, Dec 4, 2021, 03:39:12 UTC

Instance pool: my-instance-pool

Cooldown period: 300 seconds

Resources

Autoscaling policies

Autoscaling policies define the criteria that trigger autoscaling actions and the actions to take.

Edit

Policy name	Policy type	Performance metric
autoscaling-policy-20211204-1023	Threshold	CPU utilization

Showing 1 item

## 6. Tugas

- Lakukan langkah-langkah di atas, dan buat laporan.
- Jelaskan Apa kaitan Autoscaling dengan Availability?