Lab3. Auto Scaling 구현하기

1. 목적

- 이번 Lab에서는 Tencent Cloud에서 제공하는 Auto Scaling을 구현한다. Auto Scaling 서비스는 서비스에 필요한 인스턴스 수를 탄력적으로 유지할 뿐만 아니라 사용자 요청에 대해 서비스 지연 현상을 예방할 수 있는 서비스이다.

2. 사전 준비물

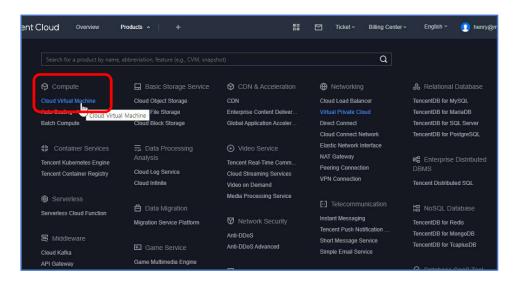
- Tencent Cloud Account

3. 목차

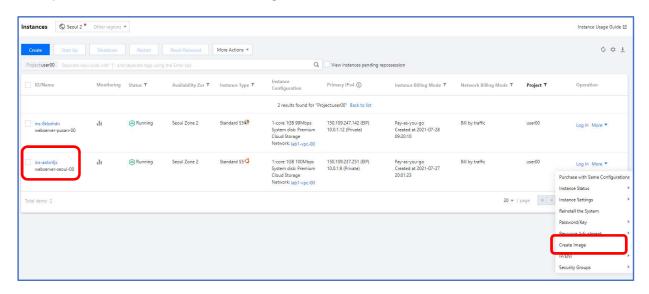
- Task1. Auto Scaling을 구성하기 전 사전 준비하기
- Task2. Launch Configuration 구성하기
- Task3. Scaling Group 설정하기
- Task4. Scaling Policy 생성하기
- Task5. Auto Scaling 테스트하기

Task1. Auto Scaling을 구성하기 전 사전 준비하기

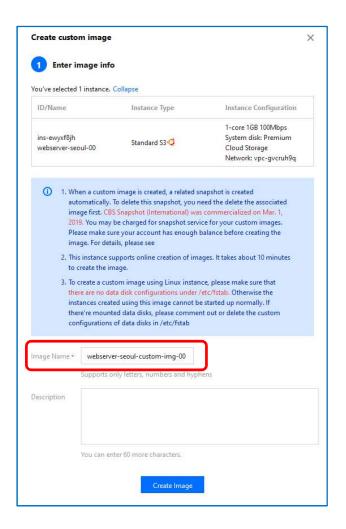
1. Lab1에서 생성한 webserver-seoul-xx(xx는 당일 부여된 번호) 인스턴스를 이용해서 Custom Image로 만드는 작업을 한다. 페이지 상단의 메뉴 중 [Products] > [Compute] > [Cloud Virtual Machine]을 클릭한다.



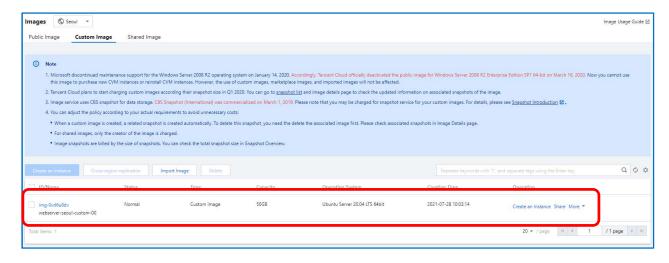
2. [Cloud Virtual Machine] Dashboard 페이지이다. 이미 Lab1에서 생성한 webserver-seoul-xx가 있다. 이 가상머신의 Custom Image를 생성하기 위해 목록에서 webserver-seoul-xx을 찾아서 제일 오른쪽 메뉴인 [Operations] > [More] > [Create Image] 메뉴를 클릭한다.



3. [Create custom image] 페이지이다. [Image Name]에 webserver-seoul-custom-img-xx(xx는 당일 부여 된 번호) 라고 입력한 다음, [Create Image] 파란색 버튼을 클릭한다.

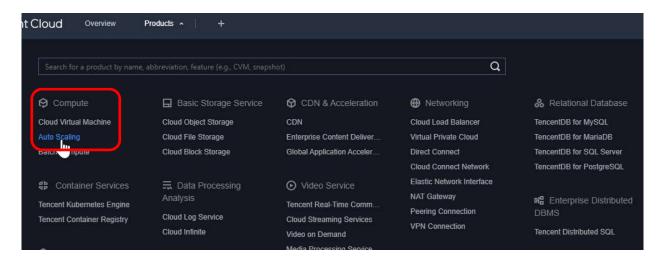


4. 잠시 후, Custom Image가 생성되는데, 확인을 위해서 [Cloud Virtual Machine] > [Images] > [Custom Image]으로 이동한다.

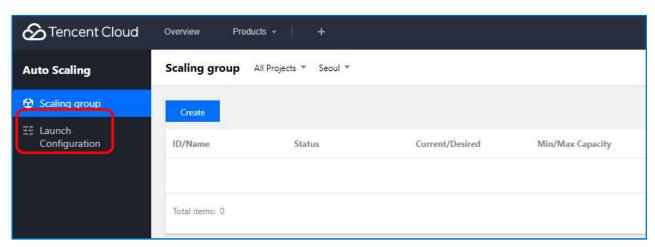


Task2. Launch Configuration 설정하기

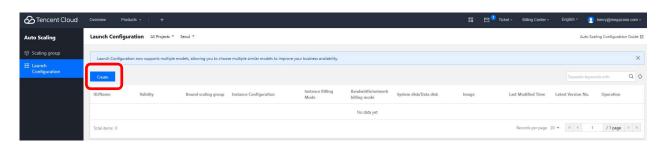
1. 먼저 Auto Scaling을 구성하기 위해 페이지 상단 메뉴에서 [Products] > [Compute] > [Auto Scaling] 메 뉴를 클릭한다.



2. [Scaling group] 페이지이다. 먼저 좌측 메뉴에서 [Launch Configuration] 메뉴를 클릭한다.



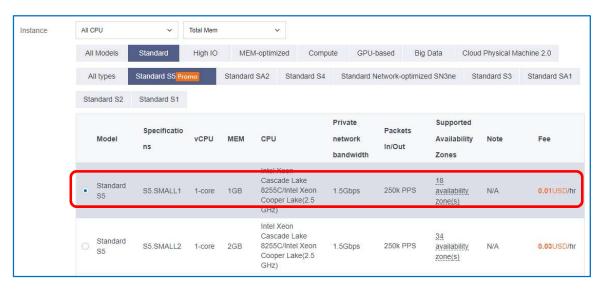
3. [Launch Configuration] 페이지이다. 인스턴스를 필요한 순간에 생성하게 하려면, <u>어떤 스펙으로 인스턴스를 생성할 것인지</u>를 미리 결정해야 한다. 그래서 [Launch Configuration]을 통해 생성할 인스턴스의스펙을 결정한다. 생성하기 위해 [Create] 파란색 버튼을 클릭한다.



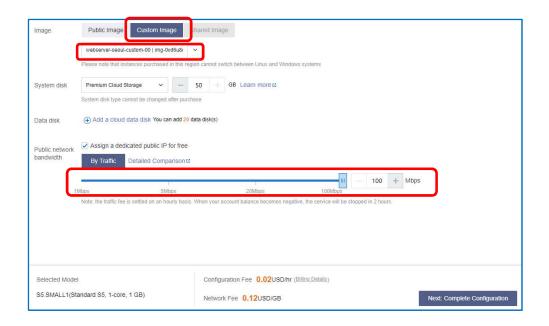
- 4. [Create Launch Configuration] 페이지이다. CVM을 생성하는 것처럼 모두 3단계에 걸쳐 생성된다. 먼저 1단계로 Select Model 단계이다. 다음의 각 값을 입력하고 [Next:Complete Configuration] 버튼을 클릭한다.
 - ① [Launch configuration name] : webserver-launch-config-xx(xx는 당일 부여된 번호)
 - ② [Billing Mode]: Pay as you go
 - ③ [Region] : Seoul
 - ④ [Availability Zone] : Seoul Zone 2



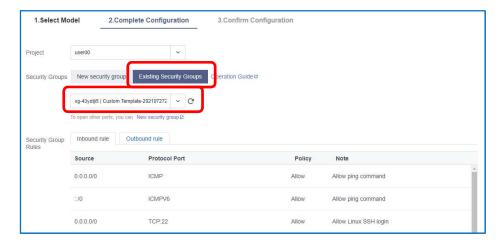
(5) [Instance] : Standard | Standard S5 | S5.SMALL1 | 1-core | 1GB | 0.01USD/hr



- ⑥ [Image] : Custom Image | webserver-seoul-custom-img-xx(xx는 당일 부여된 번호)
- ② [System disk] : Premium Cloud Storage 50GB
- [Public network bandwidth]: By Traffic | 100Mbps



- 5. 2단계 Complete Configuration 단계이다. 다음의 각 값을 입력한 다음, [Next:Confirm Configuration] 버튼을 클릭한다.
 - ① [Project] : userxx
 - **②** [Security Groups] : Existing Security Groups
 - ③ [Security Group Rules]: ICMP, TCP:22, TCP:80, TCP:443



④ [Instance Name] : 생략

(5) [Login Methods] : Set Password

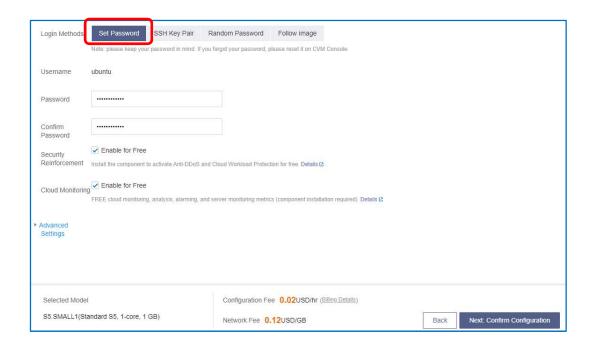
6 [Username] : ubuntu

⑦ [Password] : P@\$\$W0rd1234

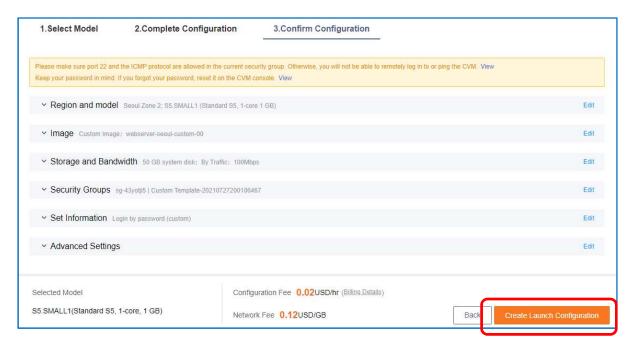
(a) [Confirm Password]: P@\$\$W0rd1234

[Security Reinforcement]: Enable for Free

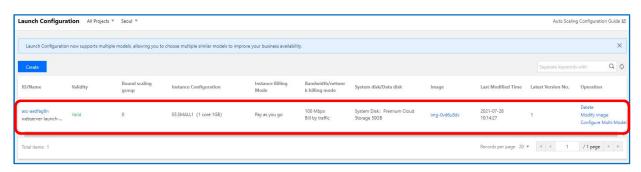
(I) [Cloud Monitoring]: Enable for Free



6. 마지막 3단계 [Confirm Configuration] 단계이다. 내용을 확인한 다음, [Create Launch Configuration] 주 황색 버튼을 클릭한다.

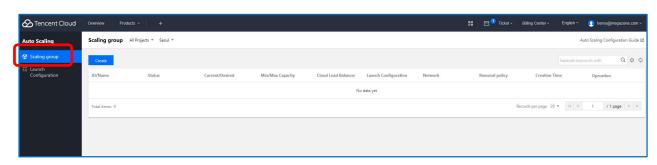


7. 잠시 후, 방금 생성한 Launch Configuration 이 목록에 보인다.

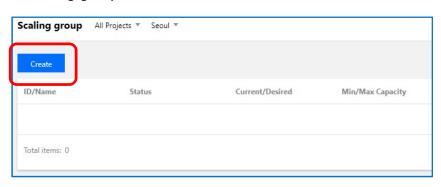


Task3. Scaling Group 설정하기

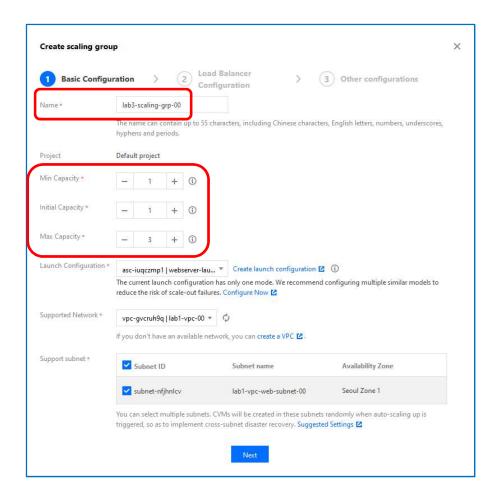
1. 이제 [Scaling group]을 설정할 차례이다. 좌측 메뉴에서 [Scaling group] 메뉴를 클릭한다.



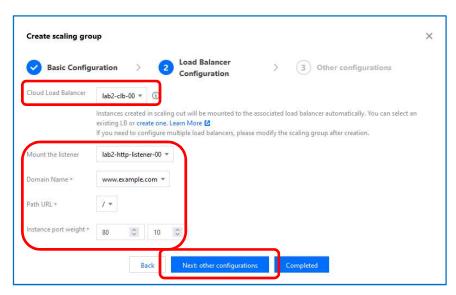
2. 새 Scaling group을 생성하기 위해 [Create] 파란색 버튼을 클릭한다.



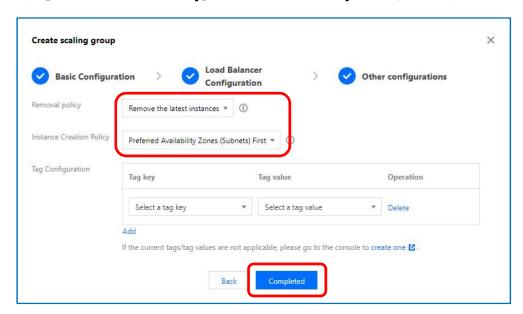
- 3. [Create scaling group]은 3단계를 거쳐야 한다. 먼저 1단계로 Basic Configuration 단계이다. 각 값을 설정한 후, [Next] 파란색 버튼을 클릭한다.
 - ① [Name]: lab3-scaling-grp-xx(xx는 당일 부여된 번호)
 - ② [Min Capacity]: 1
 - ③ [Initial Capacity]: 1
 - (4) [Max Capacity]: 3
 - ⑤ [Launch Configuration] : webserver-launch-config-xx(xx는 당일 부여된 번호)
 - ⑥ [Supported Network] : lab1-vpc-xx(xx는 당일 부여된 번호)
 - ⑦ [Support subnet] : lab1-vpc-web-subnet-xx(xx는 당일 부여된 번호)



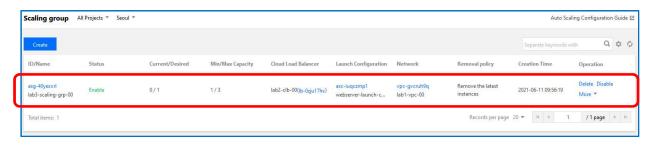
- 4. 다음은 2단계 Load Balancer Configuration 단계이다. 각각의 값을 입력 후, [Next:other configurations] 파란색 버튼을 클릭한다.
 - ① [Cloud Load Balancer] : lab2-clb-xx(xx는 당일 부여된 번호)
 - ② [Mount the listener] : lab2-http-listener-xx(xx는 당일 부여된 번호)
 - 3 [Domain Name] : www.example.com
 - ④ [Path URL]:/
 - 5 [Instance port weight]: 80 | 10



- 5. 다음 단계는 마지막 3단계로 Other configurations 단계이다. 각각의 값을 설정한 후, [Completed] 파란 색 버튼을 클릭한다.
 - ① [Removal policy] : Remove the latest instances
 - ② [Instance Creation Policy] : Preferred Availability Zones(Subnets) First

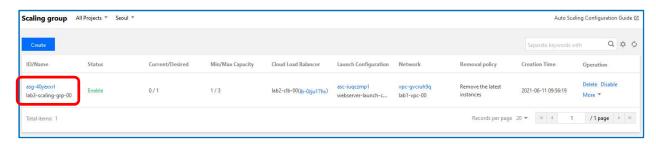


6. 이제 [Scaling group] 목록에 방금 생성한 Scaling group이 올라온 것을 확인할 수 있다.

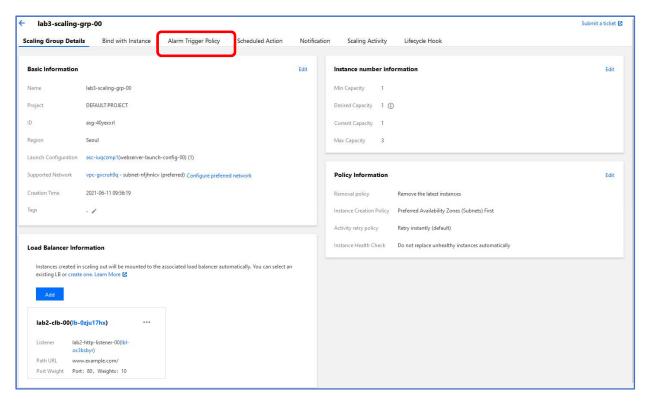


Task4. Scaling Policy 생성하기

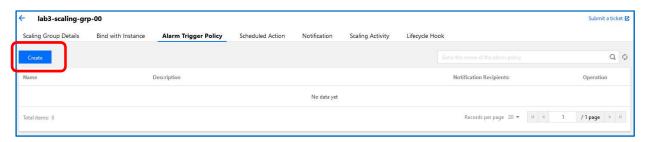
1. [Scaling group] 목록에서 방금 생성한 group을 클릭한다.



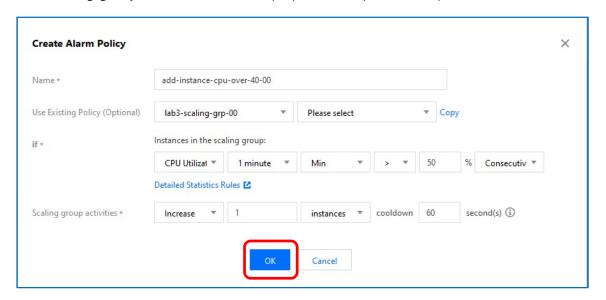
2. 해당 group의 상세 페이지가 나타나면, 상단 메뉴 중 [Alarm Trigger Policy] 메뉴를 클릭한다.



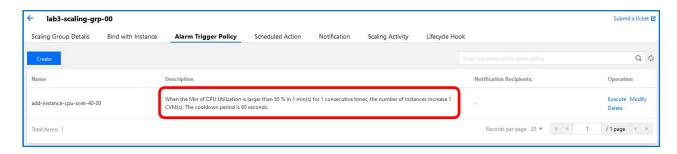
3. [Alarm Trigger Policy] 페이지에서 새 Policy를 생성하기 위해 [Create] 파란색 버튼을 클릭한다.



- 4. [Create Alarm Policy] 페이지이다. 각각의 값을 입력한 후, [OK] 파란색 버튼을 클릭한다.
 - ① [Name]: add-instance-cpu-over-40-xx(xx는 당일 부여된 번호)
 - ② [Use Existing Policy (Optional)] : lab3-scaling-grp-xx(xx는 당일 부여된 번호)
 - 3 [if]: CPU Utilization | 1 minute | Min | > | 50 | Consecutive 1 time
 - **④** [Scaling group activities] : Increase | 1 | instances | cooldown | 60 second(s)

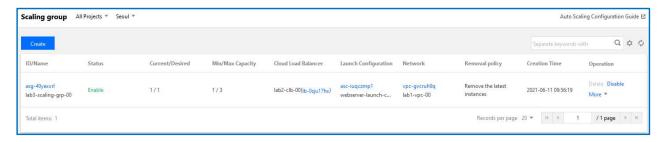


5. [Alarm Trigger Policy] 목록에 방금 생성한 Policy를 확인할 수 있다. [Description]을 보면 When the Min of CPU Utilization is larger than 50 % in 1 min(s) for 1 consecutive times, the number of instances increase 1 CVM(s). The cooldown period is 60 seconds. 즉, CPU의 최소 사용량이 1분동안 50%를 초과 하면 가상 머신(인스턴스) 1개 증가한다. 쿨다운 시간은 60초이다. 로 설정한 것을 확인할 수 있다.

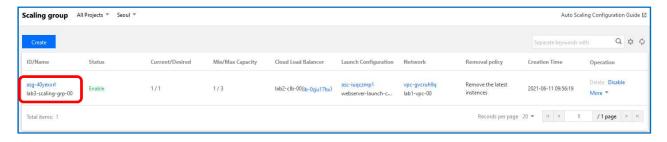


Task5. Auto Scaling 테스트하기

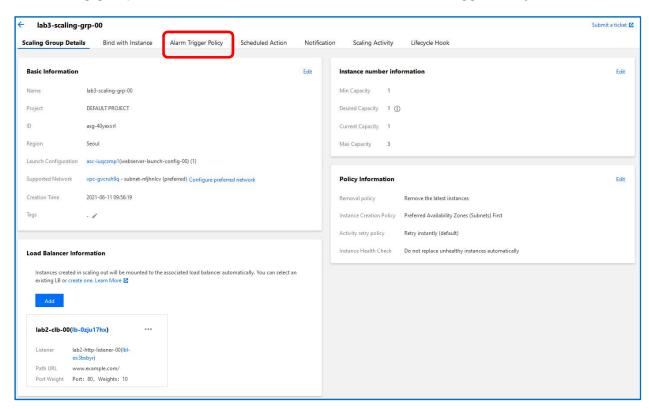
1. [Auto Scaling] 페이지에서 [Scaling group] 메뉴를 클릭한다.



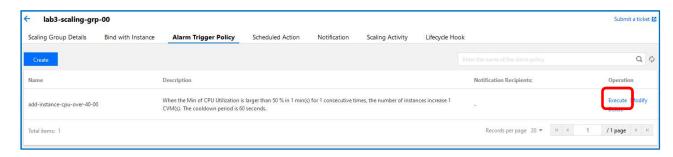
2. 목록에서 생성한 Scaling group의 이름을 클릭한다.



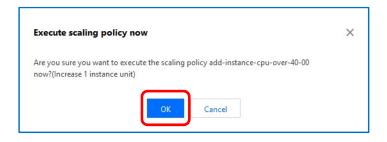
3. 해당 Scaling group의 상세페이지가 나타나면, 상단 메뉴에서 [Alarm Trigger Policy]를 클릭한다.



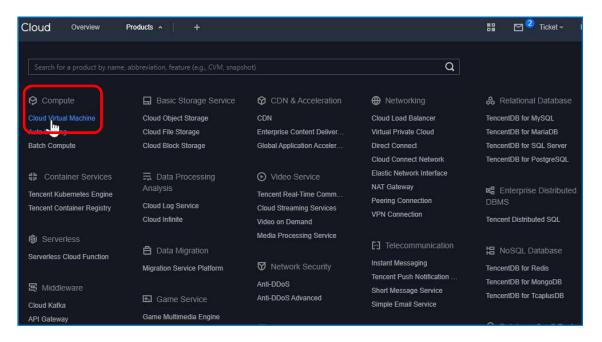
4. 목록에서 방금 생성한 Policy를 찾고, 해당 Policy의 제일 오른쪽의 [Operation] 중 [Execute] 링크를 클릭한다.



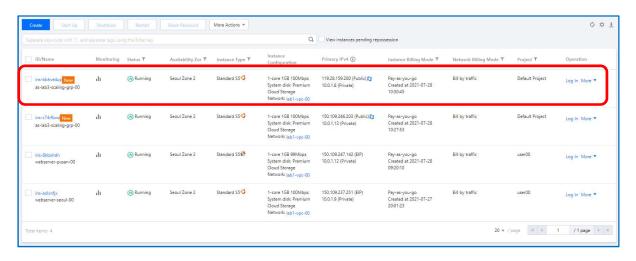
5. [Execute scaling policy now] 창이 나타나면, [OK] 파란색 버튼을 클릭한다.



6. 확인을 위해, 페이지 상단 메뉴의 [Products] > [Compute] > [Cloud Virtual Machine] 클릭하여 해당 페이지로 이동한다.



7. 인스턴스가 추가된 것을 확인할 수 있다.



8. 메일과 핸드폰 문자로(메일 및 핸드폰 등록시) 새 인스턴스가 증가된 것을 확인할 수 있다.

