

Scale-In/Out

정책 설정

(Configuring Scaling Policy)

October 2024

시나리오 구성

Scale-In/Out 정책 설정

학습 목표

#Virtual Private Cloud(VPC) #Subnet #Security Group #Virtual Server #VM Auto-Scaling #Load Balancer #Scale-In #Scale-Out

본 **Scale-In/Out 정책 설정** 시나리오는 Auto-Scaling에서 정책을 설정하여 Scale-In, Scale-Out의 과정을 확인하기 위한 시나리오입니다.

CPU, 메모리 등 자원 사용률에 기반한 Scale-In/Out 정책을 설정하는 방법을 이해하고, 사전에 정의한 조건에 따라 동일한 스펙의 VM이 자동으로 추가 생성 및 삭제되는 것을 직접 확인할 수 있는 가장 기본적인 방법을 소개합니다.

Auto-Scaling 생성 시나리오에서 실습을 통해 생성된 서버는 애플리케이션 없이 기본 OS만 설치된 서버이기 때문에 부하를 일으켜 서버의 증감을 확인하는 것이 불가능합니다. 따라서 강제적으로 낮은 임계치를 두어 이벤트를 발생시키는 형태로 실습을 진행합니다.

따라서 Scale-Out 테스트를 위해 서버의 평균 메모리 사용률 10% 이상일 때 서버를 증가시키고, Scale-In 테스트를 위해 서버의 평균 CPU 사용률이 10% 이하일 때 서버를 반납하도록 Auto-Scaling 정책을 각각 추가합니다.

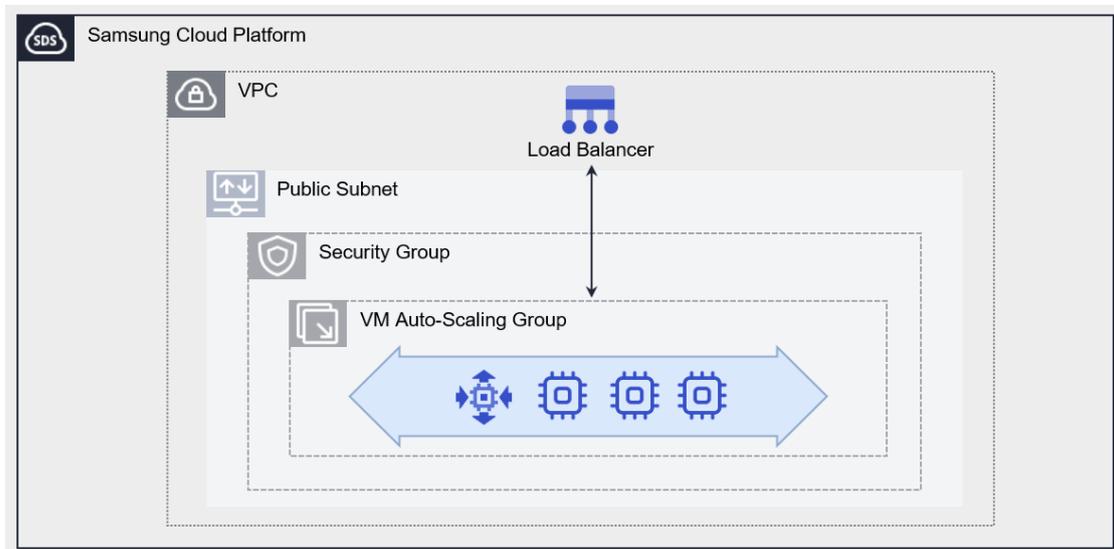


그림 1. Scale-In/Out 정책 설정 시나리오 구성도

사전 작업

<사용되는 상품 리스트>

- VPC
- Security Group
- Load Balancer
- Auto-Scaling

따라하기

사전 작업 준비하기

1. Auto-Scaling 생성하기

Scale-In/Out 정책 설정은 Auto-Scaling과 Load Balancer가 구성된 환경에서 자원 사용률 기반 정책을 생성하여 실제 서버의 증감을 확인하는 과정으로서, Auto-Scaling Group 과 Load Balancer 가 사전에 구성되어 있어야 합니다. 자세한 환경 구성 방법은 **[시작 가이드] Auto-Scaling 생성**을 참고하시기 바랍니다.

Scale-Out 확인하기

2. Scale-Out 정책 추가하기

- ① **Compute > Auto-Scaling > Auto-Scaling** 메뉴에서 **자원관리** 버튼을 클릭하세요. **Auto-Scaling 목록** 화면으로 이동합니다.
- ② **Auto-Scaling 목록** 화면에서 미리 생성한 Auto-Scaling Group을 선택하세요.
- ③ **Auto-Scaling - Auto-Scaling 상세** 화면의 **정책** 탭에서 **정책 추가** 버튼을 클릭하세요. **정책 추가** 팝업창이 열립니다.
- ④ **정책 추가** 팝업창에서 해당 정책 관련 필수 정보를 입력하고, **확인** 버튼을 클릭하세요.
 - 본 테스트에서는 Scale-out 테스트만을 위해 강제적으로 낮은 임계치를 두어 이벤트를 발생시키지만, 실제 환경에서는 이렇게 정책 설정을 하지 않습니다.
 - 여러 번의 테스트와 안정화 기간 동안 모니터링을 통해 적당한 임계치 도출 후에 적용할 필요가 있습니다.

화면	필수 입력 요소 항목	입력 값
정책 추가	구분	Scale Out
	정책명	asgout-test
	실행조건	Memory usage >= 10
	쿨다운	120

- ⑤ **Auto-Scaling - Auto-Scaling 상세** 화면의 **Virtual Server** 탭을 클릭하여 Scale-Out이 동작하는지 확인합니다.
 - 최종적으로 Max 로 지정한 3 Node 가 **Running** 상태가 되는지 확인합니다.
 - 추가 서버가 기동되는 것을 기다리는 동안 미리 생성한 Load Balancer 의 **서비스** 탭에서 해당 노드의 상태가 **Up** 으로 변경되는지도 같이 확인해 볼 수 있습니다.

Scale-In 확인하기

3. Scale-Out 정책 삭제하기

- ① **Compute > Auto-Scaling > Auto-Scaling** 메뉴에서 **자원관리** 버튼을 클릭하세요. **Auto-Scaling 목록** 화면으로 이동합니다.
- ② **Auto-Scaling 목록** 화면에서 미리 생성한 Auto-Scaling Group을 선택하세요.
- ③ **Auto-Scaling - Auto-Scaling 상세** 화면의 **정책** 탭에서 미리 생성한 Scale-Out 정책을 선택한 후 **삭제** 버튼을 클릭하세요.
 - 테스트만을 위한 임계치를 설정한 것으로, 낮은 자원에서도 Scale-Out 되도록 설정하기 때문에 삭제하지 않을 경우 Scale-In 정책과 충돌하게 됩니다.

4. Scale-In 정책 추가하기

- ① **Compute > Auto-Scaling > Auto-Scaling** 메뉴에서 **자원관리** 버튼을 클릭하세요. **Auto-Scaling 목록** 화면으로 이동합니다.
- ② **Auto-Scaling 목록** 화면에서 미리 생성한 Auto-Scaling Group을 선택하세요.
- ③ **Auto-Scaling - Auto-Scaling 상세** 화면의 **정책** 탭에서 **정책 추가** 버튼을 클릭하세요. **정책 추가** 팝업창이 열립니다.
- ④ **정책 추가** 팝업창에서 해당 정책 관련 필수 정보를 입력하고, **확인** 버튼을 클릭하세요.
 - 본 테스트에서는 Scale-In 테스트만을 위해 강제로 낮은 임계치를 두어 이벤트를 발생시키지만, 실제 환경에서는 이렇게 정책 설정을 하지 않습니다.
 - 여러 번의 테스트와 안정화 기간 동안 모니터링을 통해 적당한 임계치 도출 후에 적용할 필요가 있습니다.

화면	필수 입력 요소 항목	입력 값
정책 추가	구분	Scale In
	정책명	asgin-test
	실행조건	CPU usage <= 10
	쿨다운	120

- ⑤ **Auto-Scaling - Auto-Scaling 상세** 화면의 **Virtual Server** 탭을 클릭하여 Scale-In이 동작하는지 확인합니다.
 - 최종적으로 Min 으로 지정한 1 Node 까지 Virtual Server 의 수가 줄어드는 것을 확인합니다.

이벤트 확인하기

5. Scale-In/Out 관련 발생 이벤트 내역 확인하기
 - ① **Compute > Auto-Scaling > Auto-Scaling** 메뉴에서 **자원관리** 버튼을 클릭하세요. **Auto-Scaling 목록** 화면으로 이동합니다.
 - ② **Auto-Scaling 목록** 화면에서 미리 생성한 Auto-Scaling을 선택하세요.
 - ③ **Auto-Scaling - Auto-Scaling 상세** 화면의 **이벤트** 탭에서 Scale-In/Out 관련 발생 이벤트 내역을 확인합니다.

정리하기

- 모니터링 대상에 설정한 임계치 기준을 충족하거나, 사전 예약된 일정에 맞춰 자동으로 서버 수량을 확장(Scale-Out) 및 축소(Scale-In)할 수 있습니다.
- Scale-In/Out 각각의 이벤트 발생 시, 이벤트 발생 내역을 수신할 수 있는 알림 서비스를 제공합니다.