



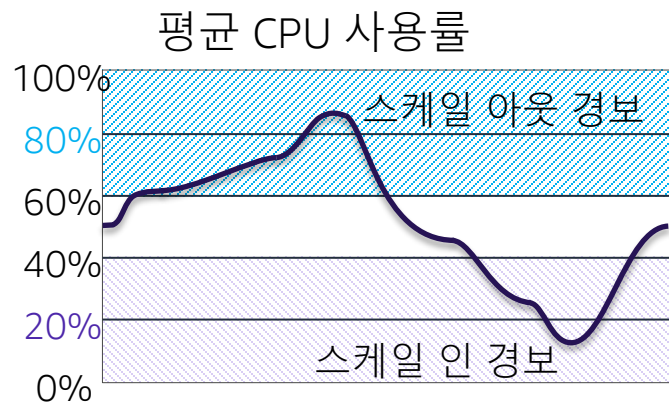
# 오토스케일링 예측 당면 과제 이끌기

컴퓨팅(크기 조정 및 이름 확인)

당면 과제

# 예측해야 할 항목

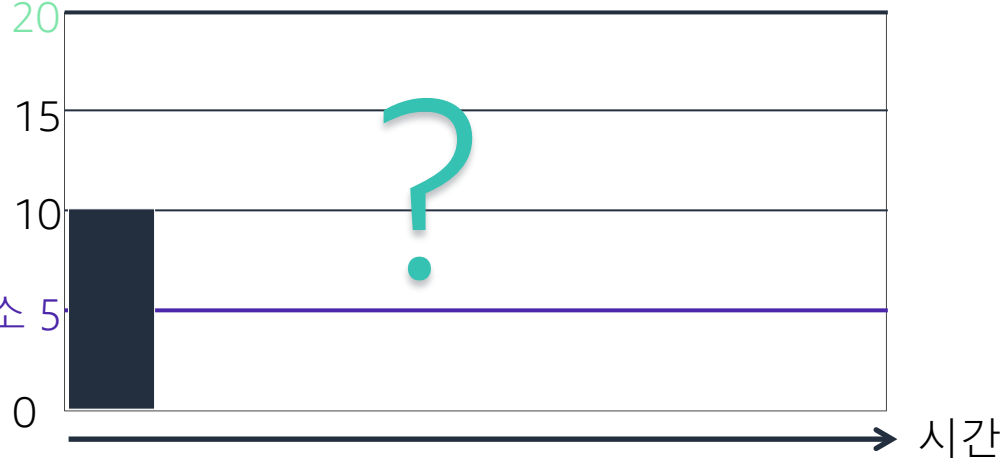
- 다음 Auto Scaling 그룹에 대한 단계 조정을 위한 Amazon EC2 Auto Scaling 정책을 생성하기로 결정했습니다.
  - 최대 용량: 20
  - 원하는 용량: 10
  - 최소 용량: 5
- 그러나 동작을 제어하기 위해 정의한 현재 정책을 기반으로 Amazon Elastic Compute Cloud(Amazon EC2) 인스턴스가 스케일 아웃 또는 스케일 인되는 방식은 확실하지 않습니다.
- 인스턴스 동작을 테스트하기 위해 정책 및 해당 경보를 기반으로 언제든지 해당 인스턴스 수를 예측할 수 있기를 원합니다.



인스턴스 개수

최대: 20

최소: 5



# 당면 과제: 조건 #1

## 단계 조정 정책 정보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
  - **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
  - **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
  - **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가
- 각 조건에 대해 추가되거나 제거되는 인스턴스 수를 예측합니다.
  - **조건 #1:** 2분 넘게 평균 CPU 사용률은 63~70%였습니다.

**참고:** 인스턴스 워밍업 기간이 5분으로 설정되어 있다고 가정합니다.

## 당면 과제: 조건 #2

### 단계별 크기 조정 정책 정보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
  - **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
  - **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
  - **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가
- 각 조건에 대해 추가되거나 제거되는 인스턴스 수를 예측합니다.
  - **조건 #2:** CPU 압력은 계속 증가하여 60~80%를 유지하고 있습니다. 조건 #1이 발생한 지 2분이 지났습니다.

**참고:** 인스턴스 워밍업 기간이 5분으로 설정되어 있다고 가정합니다.

# 당면 과제: 조건 #3

## 단계별 크기 조정 정책 정보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
  - **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
  - **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
  - **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가
- 각 조건에 대해 추가되거나 제거되는 인스턴스 수를 예측합니다.
  - 조건 #3: 조건 #1이 발생한 후 아직 5분이 지나지 않았지만 CPU 사용률은 현재 85%입니다.

**참고:** 인스턴스 워밍업 기간이 5분으로 설정되어 있다고 가정합니다.

# 당면 과제: 조건 #4

## 단계별 크기 조정 정책 경보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
  - **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
  - **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
  - **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가
- 추가되거나 제거되는 인스턴스 수를 예측합니다.
  - 조건 #4: 시간이 더 지났습니다. 추가된 2개 인스턴스 중 첫 번째 인스턴스는 워밍업되어 로드の一部를 처리합니다. 두 번째 인스턴스는 아직 워밍업 상태에 있습니다. 그러나 CPU 사용량은 지난 3분 동안 약 53%로 떨어졌습니다.

# 당면 과제: 조건 #5

## 단계별 크기 조정 정책 경보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
  - **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
  - **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
  - **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가
- 추가되거나 제거되는 인스턴스 수를 예측합니다.
  - 조건 #5: 이제 애플리케이션 배포에 대한 수요가 감소하여 CPU 사용량이 꾸준히 32%로 떨어졌습니다.

# 당면 과제: 조건 #6

## 단계별 크기 조정 정책 경보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
  - **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
  - **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
  - **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가
- 추가되거나 제거되는 인스턴스 수를 예측합니다.
  - 조건 #6: CPU 사용량이 훨씬 더 줄어들어 현재 약 17%입니다.



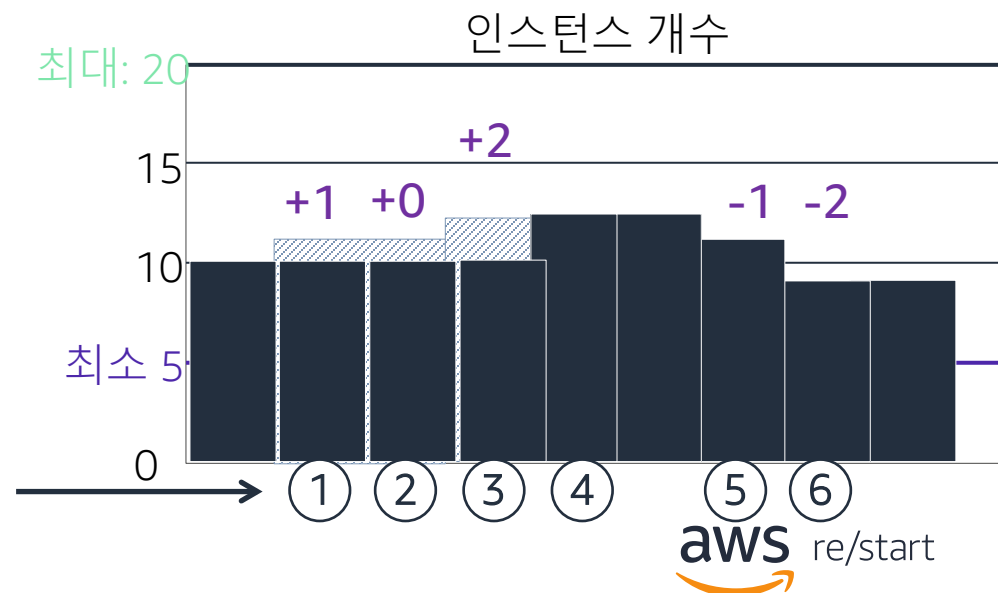
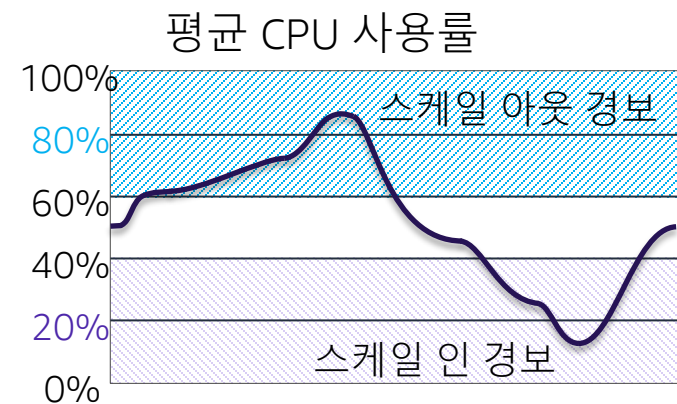
# 단계 조정 정책 당면 과제: 정답 키

예측이 이 두 그래프에 표시된 결과와 일치했습니까?

단계별 크기 조정 정책 경보:

- **스케일 인:** 2분 넘게 평균 CPU가 20%보다 크고 40%보다 작을 때 인스턴스 1개 제거
- **스케일 인:** 2분 넘게 평균 CPU가 0%보다 크고 20%보다 작을 때 인스턴스 2개 제거
- **스케일 아웃:** 2분 넘게 평균 CPU가 60%보다 크고 80%보다 작을 때 인스턴스 1개 추가
- **스케일 아웃:** 2분 넘게 평균 CPU가 80%보다 크고 100%보다 작을 때 인스턴스 2개 추가

당면 과제에서 해당 번호의 조건이 발생한 경우



# 스래싱 방지



경보 지속 기간:

예: 10분 동안 90%를 초과하는 CPU 사용률



휴지 기간:

- 스케일 인 또는 스케일 아웃을 실행하면 휴지 기간 동안 더 많은 크기 조정 작업을 유예함
- 단순 조정 정책에 사용됨

예: 5분 동안 크기 조정 중지


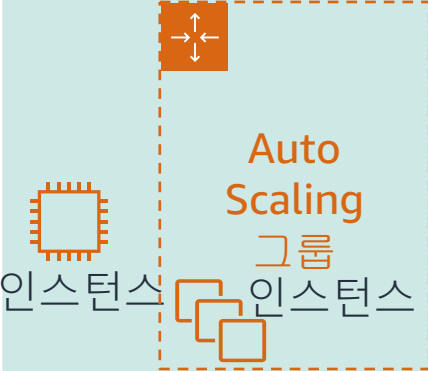

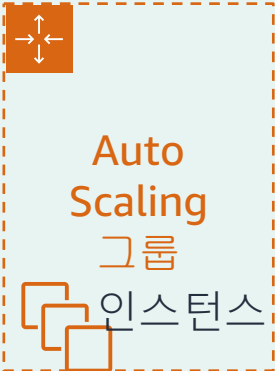




인스턴스 워밍업 기간:

- 새로 시작된 인스턴스가 워밍업되는 데 걸리는 시간(초)
- 단계 조정 정책에 사용됨

예: 워밍업은 5분이 걸립니다.

# Amazon EC2 Auto Scaling 그룹 수명 주기 후크

이벤트	이전	Auto Scaling 작업	사용자/프로그래밍된 작업		이후
스케일 아웃		<ol style="list-style-type: none"> <li>1. 인스턴스를 시작합니다.</li> <li>2. 알림을 보냅니다.</li> </ol>		<ol style="list-style-type: none"> <li>3. 인스턴스 작업을 수행합니다(예: 소프트웨어 설치).</li> <li>4. 그룹에 추가합니다.</li> </ol>	
스케일 인		<ol style="list-style-type: none"> <li>1. Auto Scaling 그룹에서 인스턴스를 제거합니다.</li> <li>2. 알림을 보냅니다.</li> </ol>		<ol style="list-style-type: none"> <li>3. 인스턴스 작업을 수행합니다(예: 로그 검색).</li> <li>4. 종료합니다.</li> </ol>	

# 학습 내용 확인

- 스케일 인 또는 스케일 아웃 이벤트 중에 스래싱을 방지하는 3가지 방법은 무엇입니까?
- 스케일 아웃 이벤트가 발생하면 새로 시작된 인스턴스에 대한 세부 정보가 DevOps 팀에서 사용하는 모니터링 애플리케이션으로 전달된다는 새로운 요구 사항이 생성되었습니다. 요구 사항에 따라 인스턴스가 시작되기 전에 통계가 전송되어야 합니다. 스케일 아웃 이벤트 중에 스크립트를 어떻게 시작할 수 있습니까?

# 핵심 사항



- 크기 조정 규칙을 명확하게 이해하고 많은 잠재적 워크로드에 걸쳐 크기 조정에 대한 영향을 이해해야 합니다.
- 스래싱은 크기가 조정되고 인스턴스가 너무 빨리 생성되거나 제거되는 상태입니다. 스래싱을 방지하려면 지속, 휴지 및 워밍업 기간을 사용합니다.
- 수명 주기 후크는 ASG에서 EC2 시작을 일시 중지합니다. 이 시간 동안 인스턴스는 애플리케이션 또는 스크립트를 사용하여 구성될 수 있습니다. 수명 주기 이벤트가 완료된 후 인스턴스를 사용할 수 있게 됩니다.