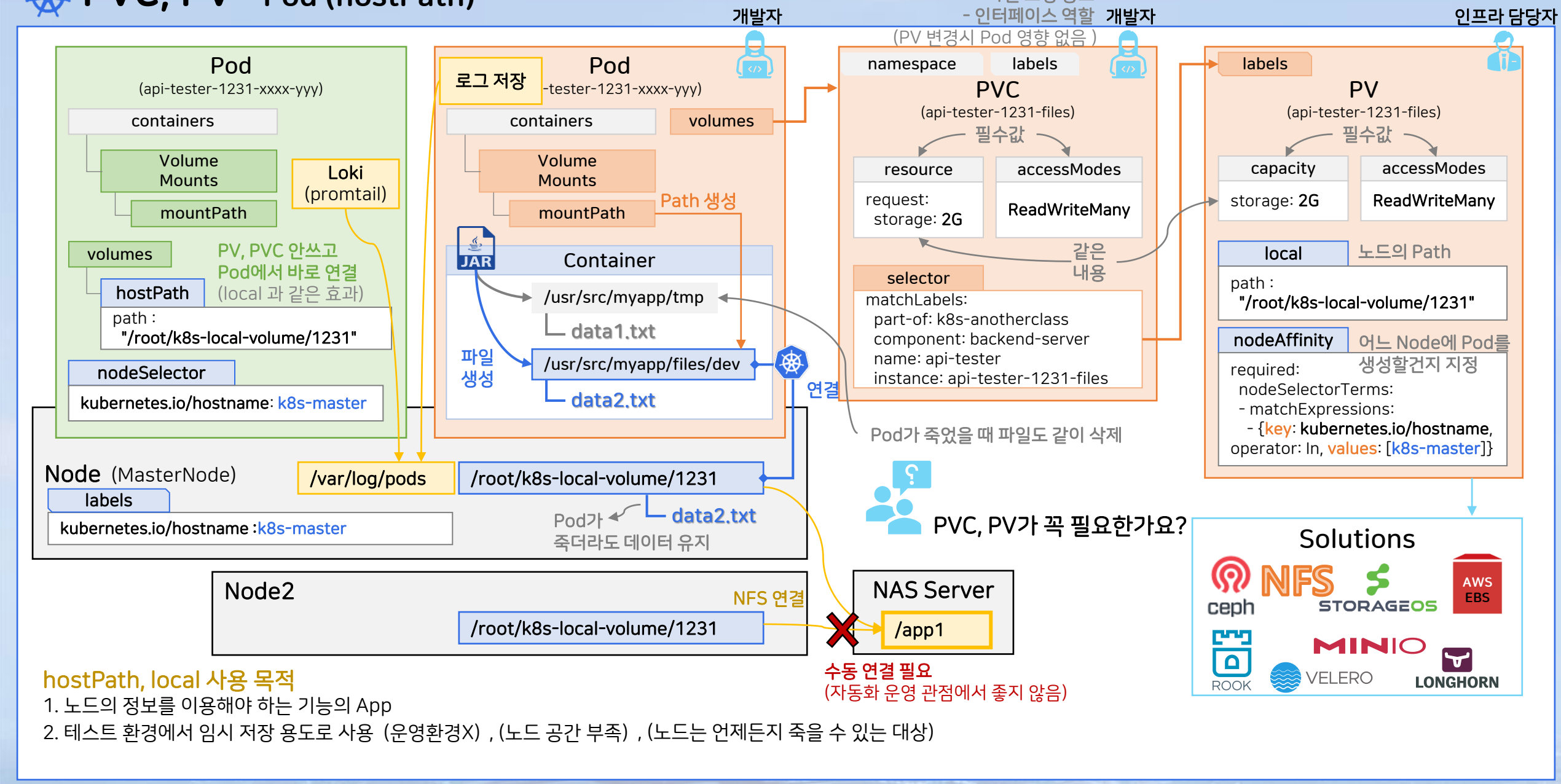

KubeOps
상위 1% 월급쟁이를 위한
쿠버네티스 어나더 클래스
Application 기능을 이해하기
PVC, PV / Deployment / HPA / Service

PVC, PV - Pod (hostPath)

Application 기능으로 이해하기



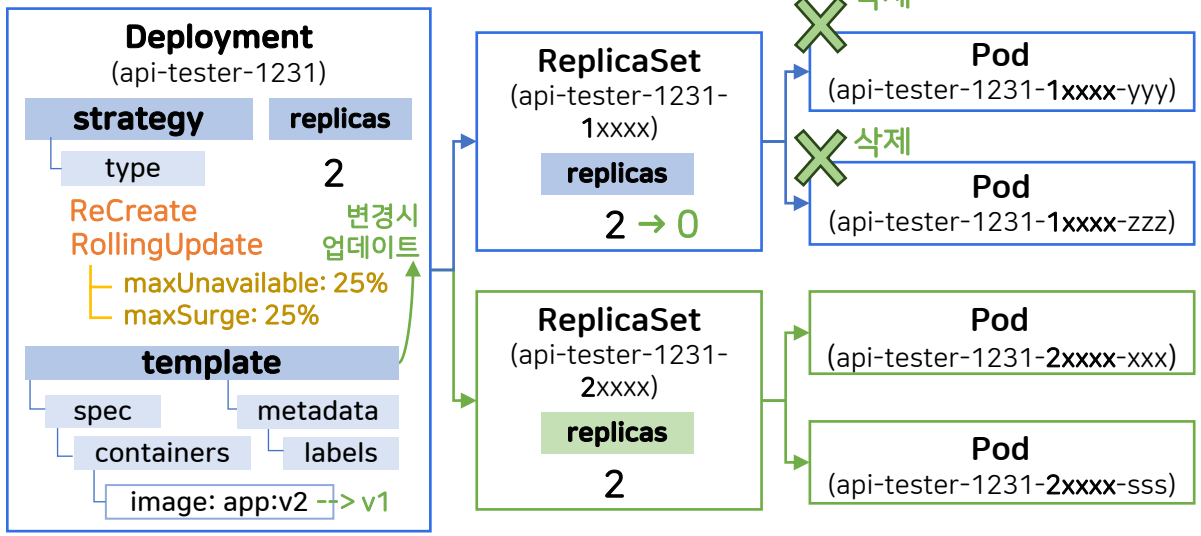
hostPath, local 사용 목적

1. 노드의 정보를 이용해야 하는 기능의 App
2. 테스트 환경에서 임시 저장 용도로 사용 (운영환경X) , (노드 공간 부족) , (노드는 언제든지 죽을 수 있는 대상)

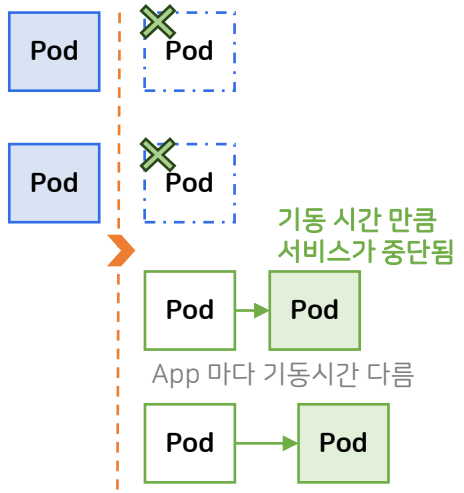
Deployment - update

Application 기능으로 이해하기

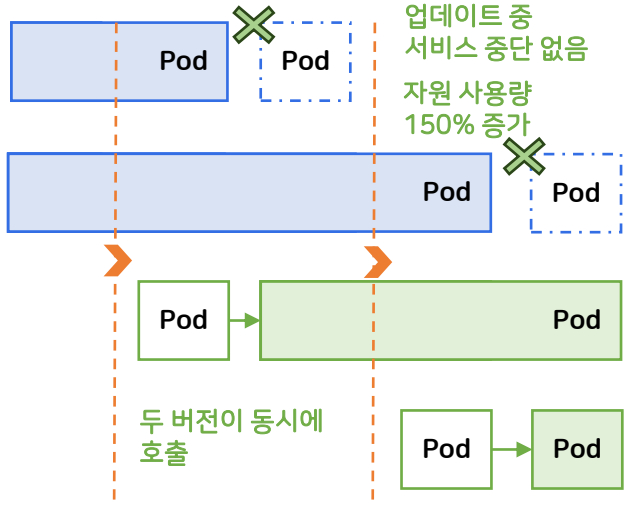
Deployment Udpate



Recreate



RollingUpdate

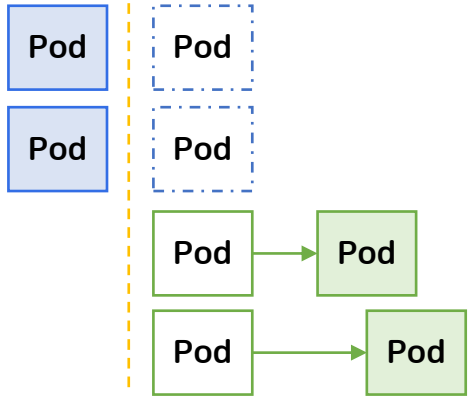


[Blue/Green]
- 업데이트 동안 두 버전이 동시에 호출되지 않음
- 자원 사용량 200% 증가
- 별도 배포 솔루션에서 제공

RollingUpdate - maxUnavailable, maxSurge

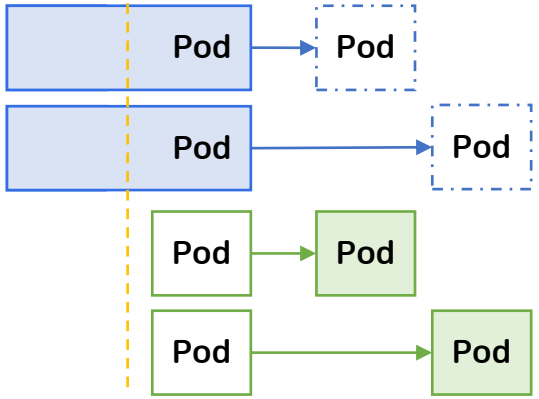
maxUnavailable: 100%
maxSurge: 100%

Recreate와
같은 효과

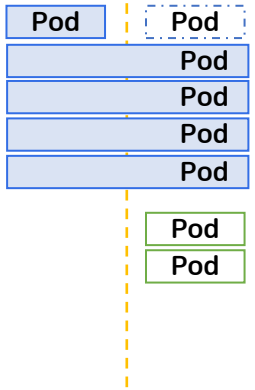


maxUnavailable: 0%
maxSurge: 100%

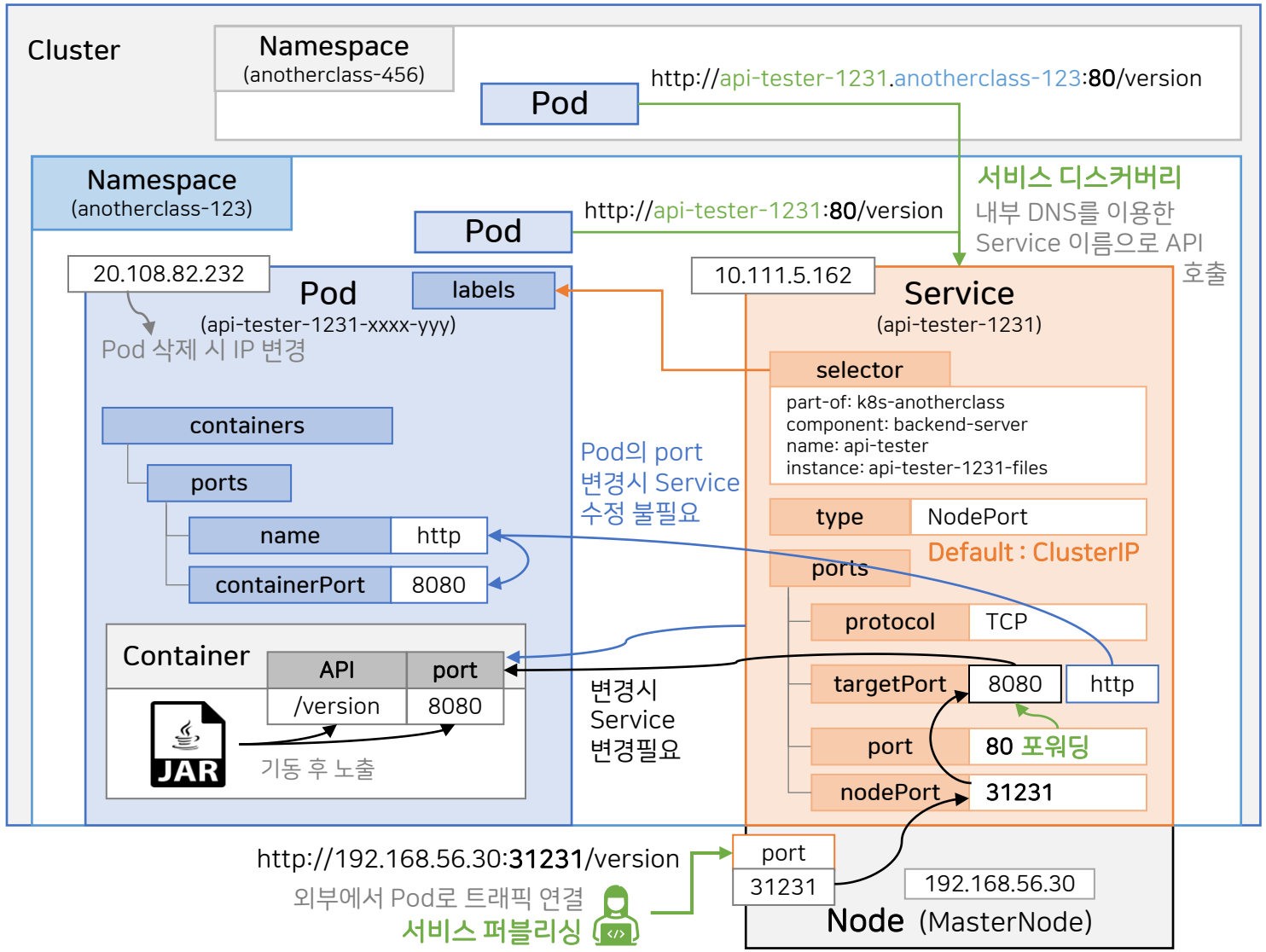
Blue/Green에
가까운 효과

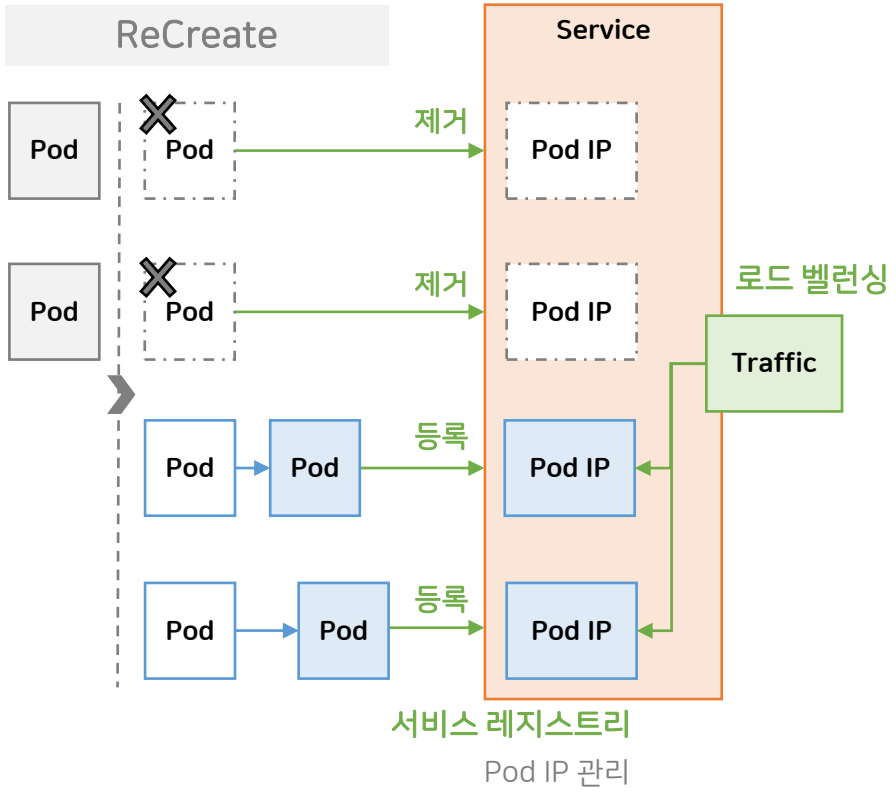


maxUnavailable: 25%
maxSurge: 25%



Service

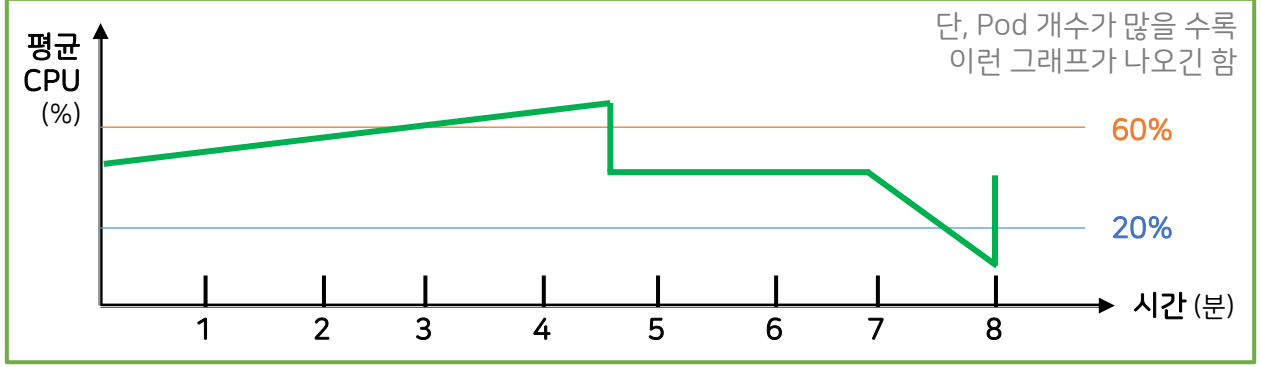

Cluster
Namespace
(anotherclass-456)
Pod
http://api-tester-1231.anotherclass-123:80/version
Namespace
(anotherclass-123)
Pod
http://api-tester-1231:80/version
서비스 디스커버리
내부 DNS를 이용한
Service 이름으로 API
호출
20.108.82.232
Pod
labels
(api-tester-1231-xxxx-yyy)
Pod 삭제 시 IP 변경
10.111.5.162
Service
(api-tester-1231)
selector
part-of: k8s-anotherclass
component: backend-server
name: api-tester
instance: api-tester-1231-files
containers
ports
name
http
containerPort
8080
Pod의 port
변경시 Service
수정 불필요
type
NodePort
Default : ClusterIP
ports
protocol
TCP
Container
API
port
/version
8080
JAR
기동 후 노출
변경시
Service
변경필요
targetPort
8080
http
port
80 포워딩
nodePort
31231
http://192.168.56.30:31231/version
외부에서 Pod로 트래픽 연결
서비스 퍼블리싱
port
31231
192.168.56.30
Node (MasterNode)


ReCreate
Service
Pod
Pod
제거
Pod IP
Pod
Pod
제거
Pod IP
로드 벨런싱
Traffic
Pod
Pod
등록
Pod IP
Pod
Pod
등록
Pod IP
서비스 레지스트리
Pod IP 관리

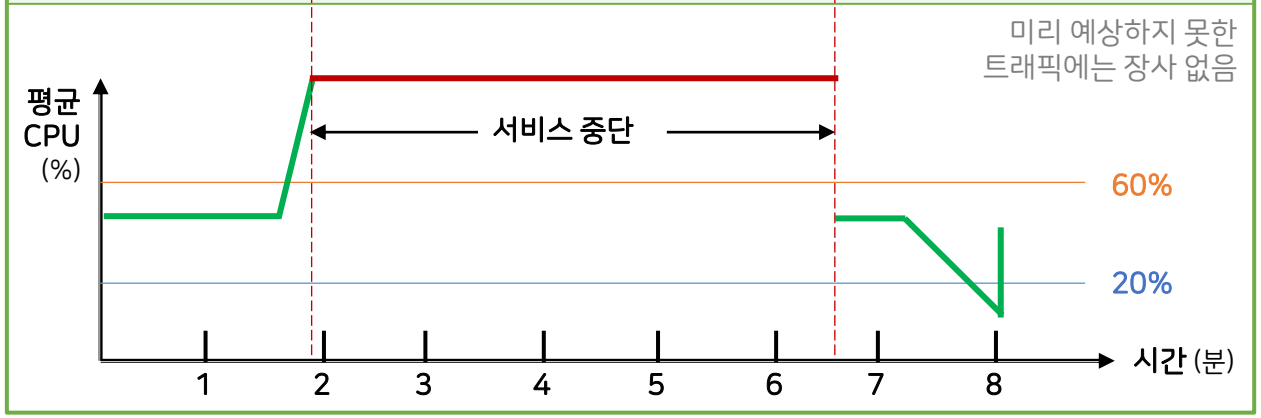
HPA

Application 기능으로 이해하기

[이상적인 스케일링]



[현실적인 스케일링]



HPA
(api-tester-1231-default)

scaleTargetRef

```
apiVersion: apps/v1
kind: Deployment
name: api-tester-1231
```

minReplicas	2
maxReplicas	4

metrics

```
- type: Resource
  resource:
    name: cpu
    target:
      type: Utilization
      averageUtilization: 60
```

수치를 결정하는데 영향도가 큼

현재 Pod 수 * (평균 CPU / HPA CPU)
→ 변경될 Pod 수

2 * (((150+1)/2) / 60) = 3 (2.51)

behavior 잦은 스케일링 방지

```
scaleUp:
  stabilizationWindowSeconds: 120
scaleDown
  stabilizationWindowSeconds: 600
  policies:
    - type: Pods
      value: 1
      periodSeconds: 60
```

1분에 1개씩 Pod 제거

2분 동안 60%이상 유지시 scaleOut

부하가 감소해도 10분동안은 Pod 수 개수

Deployment
(api-tester-1231)

replicas 2

ReplicaSet

Pod -Xmx 50M

resource	container
limits: cpu: 200m memory: 200Mi	200% 150m
requests: cpu: 100m memory: 100Mi	70m (CPU) 40Mi (Memory)

100% 계산을 위한 기준 값

부하에 따른 증감 대상 아님