



Simple Deployment on Kubernetes



kubernetes

- Start by creating a Kubernetes deployment YAML file. You can name it `my-app-deployment.yaml`. In this file, you'll define the deployment configuration for your application. Here's an example configuration:

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-app
spec:
  replicas: 3
  selector:
    matchLabels:
      app: my-app
  template:
    metadata:
      labels:
        app: my-app
    spec:
      containers:
      - name: my-app-container
        image: my-app-image:latest
        ports:
        - containerPort: 80
```

- Next, create a Kubernetes service YAML file. You can name it my-app-service.yaml. In this file, you'll define the service configuration for your application. Here's an example configuration:

```
apiVersion: v1
kind: Service
metadata:
  name: my-app-service
spec:
  selector:
    app: my-app
  ports:
    - name: http
      port: 80
      targetPort: 80
  type: LoadBalancer
```

- Now, let's create a Kubernetes Horizontal Pod Autoscaler (HPA) YAML file. You can name it my-app-hpa.yaml. In this file, you'll define the autoscaling configuration for your application. Here's an example configuration:

```
apiVersion: autoscaling/v1
kind: HorizontalPodAutoscaler
metadata:
  name: my-app-hpa
spec:
  scaleTargetRef:
    apiVersion: apps/v1
    kind: Deployment
    name: my-app
  minReplicas: 3
  maxReplicas: 10
  targetCPUUtilizationPercentage: 80
```

This YAML file defines an HPA named my-app-hpa that will scale the deployment named my-app. The HPA will ensure that there are at least three replicas of the deployment running at all times, and up to a maximum of ten replicas. The target CPU utilization percentage is set to 80%.

- Finally, let's create a Kubernetes Deployment strategy YAML file. You can name it my-app-strategy.yaml. In this file, you'll define the deployment strategy configuration for your application. Here's an example configuration:
yaml

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: my-app
spec:
  strategy:
    type: RollingUpdate
    rollingUpdate:
      maxUnavailable: 1
      maxSurge: 1
```

This YAML file defines a RollingUpdate strategy for the deployment named my-app. The strategy will ensure that there is always at least one pod available during the deployment process, and at most one extra pod beyond the desired number of replicas.

Once you have all the YAML files, you can deploy your application by running the following command:

```
> kubectl apply -f my-app-deployment.yaml
```

```
> kubectl apply -f my-app-service.yaml
```

```
> kubectl apply -f my-app-hpa.yaml
```

```
> kubectl apply -f my-app-strategy.yaml
```

This will create your application deployment, service, HPA