**Horizonal pod autoscaler HPA**

* HPA automatically scales number of pods in Replica Set, deployment based on observed CPU utilization or with custom metrics support.
* Deploy any application and execute the following command

kubectl autoscale deployment admin-service-deployment --cpu-percent=50 --min=1 --max=5



Link: <https://kubernetes.io/docs/tasks/run-application/horizontal-pod-autoscale/>

Note:

Matrix resources: CPU, Memory or Number of incoming requests.

We can achieve node scaling as well if infrastructure is cloud.

desiredReplicas = ceil[currentReplicas \* ( currentMetricValue / desiredMetricValue )]