

Autoscaling

Autoscaling in Kubernetes (KCNA)

- **Horizontal Pod Autoscaler (HPA):**
 - Scales Pods based on resource utilization (CPU, memory, custom metrics).
 - Monitors and adjusts Pod replicas dynamically.
- **Vertical Pod Autoscaler (VPA):**
 - Adjusts resource requests/limits (CPU, memory) for Pods.
 - Optimizes resource allocation for individual Pods.
- **Cluster Autoscaler:**
 - Scales nodes in a cluster up or down based on resource demands.
 - Adds/removes nodes when Pods can't be scheduled due to resource constraints.
- **Custom Metrics Autoscaler:**
 - Scales based on custom metrics (e.g., application performance metrics).
- **Scaling Policies:**
 - Define minimum and maximum replica limits.
 - Set scaling thresholds (e.g., CPU, memory utilization).

Revision #2

Created 18 November 2024 21:52:12 by Admin

Updated 21 November 2024 20:06:13 by Admin