In chaos engineering, a "pod auto scaler" refers to a component or feature in Kubernetes that automatically adjusts the number of pod replicas based on observed metrics such as CPU utilization, memory usage, or custom-defined metrics.

The purpose of a pod auto scaler is to dynamically scale the number of pods up or down to meet the desired performance or resource utilization targets. For example, if the CPU usage of a service increases beyond a certain threshold, the auto scaler can provision additional pod replicas to handle the increased load. Conversely, if the CPU usage decreases, the auto scaler can scale down the number of replicas to conserve resources.

Important

For example, check if the pods are successfully rescheduled within a specified period in cases where the existing nodes are already running at the specified limits.

* The experiment aims to check the ability of nodes to accommodate the number of replicas a given application pod.

TOTAL\_CHAOS\_DURATION :- The timeout for the chaos experiment (in seconds).

REPLICA\_COUNT:- Number of replicas upto which we want to scale

Kubectl get deploy

Kubectl get po -w

RAMP\_TIME :- Period to wait before and after injection of chaos in sec.