Auto scaling

Monday, May 27, 2024   4:26 PM

→ Autoscaling can be triggered

• Identify autoscaling factors.

It enables you to specify the conditions under which a web app should be scaled out, and back in again.

※ Auto scale Conditions.

if Scale based on metrics
    a) CPU Percentage
    b) Memory Percentage.
    c) Disk queue length.
    d) HTTP queue length.
    e) Data In
    f) Data Out

or Scale based on a time for a specific day of the week.

Monitor auto scaling activity.
        ↓
    (use Run History)

Considerations for scaling when multiple rules are configured in a profile.

On scale out → autoscale runs of any rule is met.

On scale in → autoscale requires all rules to be met.