**AUTO SCALING GROUP**

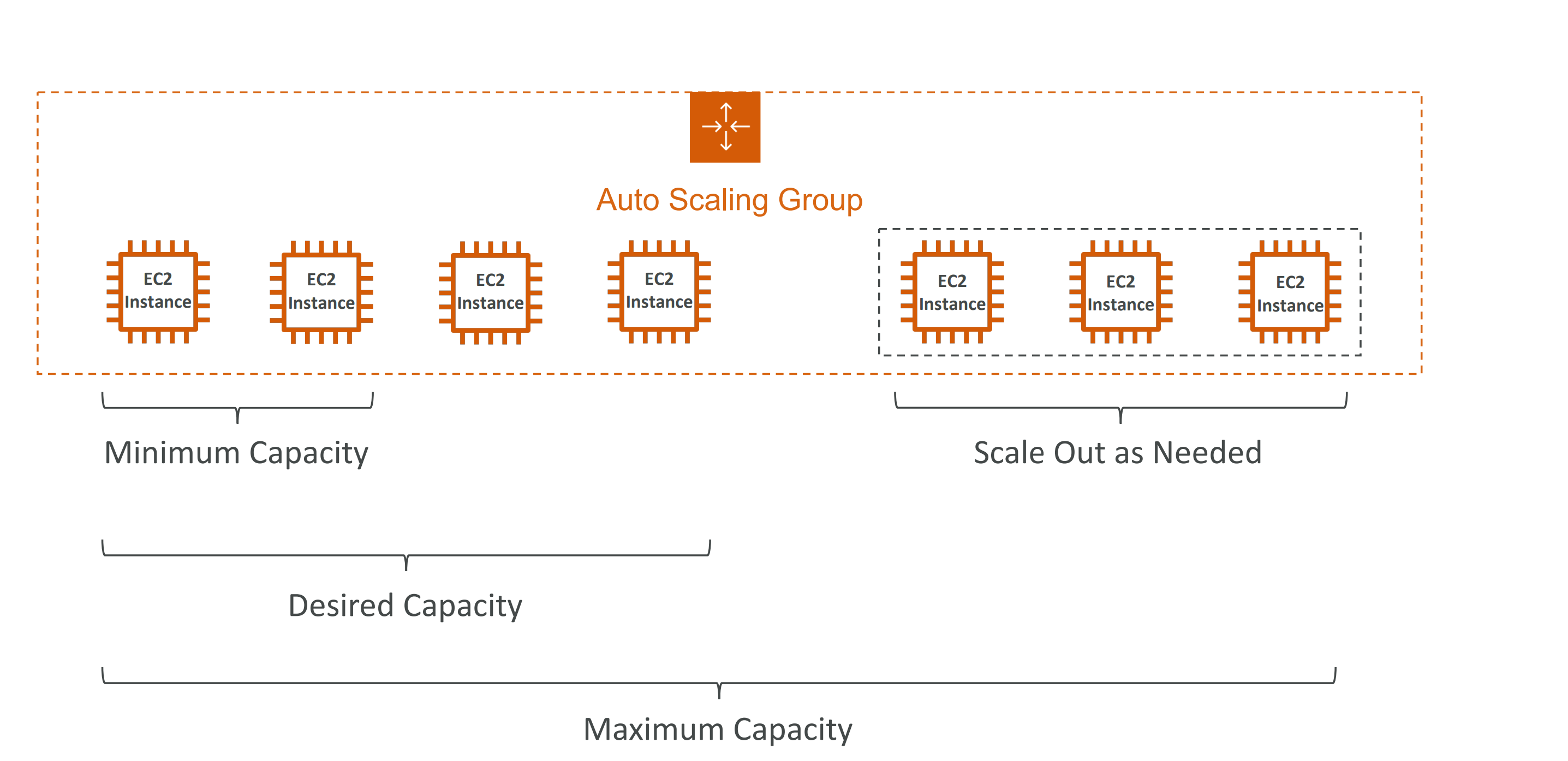
Auto scaling is a cloud computing feature that allows the automatic adjustment of computing resources in a system to handle changes in demand.

The goal of auto scaling is to ensure that your application or service has the right amount of resources to handle the current load, without manual intervention.

**2 types of Auto Scaling :**

**Vertical Scaling** : Vertical scaling involves adjusting the capacity of a single instance by increasing or decreasing its resources, such as CPU, memory, or storage, to meet changing performance needs.

**Horizontal Scaling :** Horizontal scaling involves adding or removing multiple instances to or from a system to distribute the workload, ensuring efficient resource utilization and improved performance as demand fluctuates.

****