



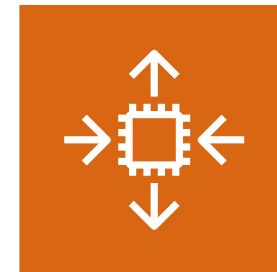
AWS Auto Scaling

Amazon Web Services

What is Auto Scaling

Auto Scaling enables you to ensure you have the correct number of EC2 instance to ensure application availability at any given point in time.

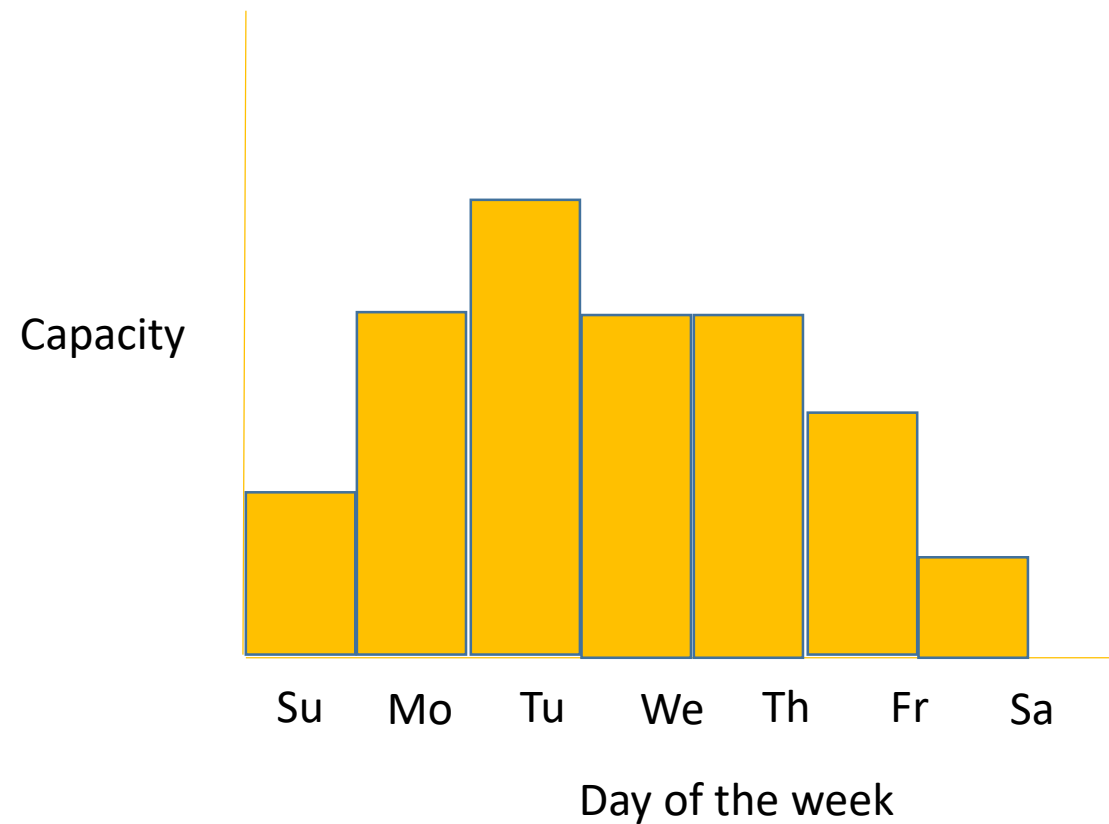
You can automatically add or remove EC2 instances based on predefined conditions you specify. Auto Scaling enables you to achieve automated scalability.



Amazon EC2 Auto Scaling

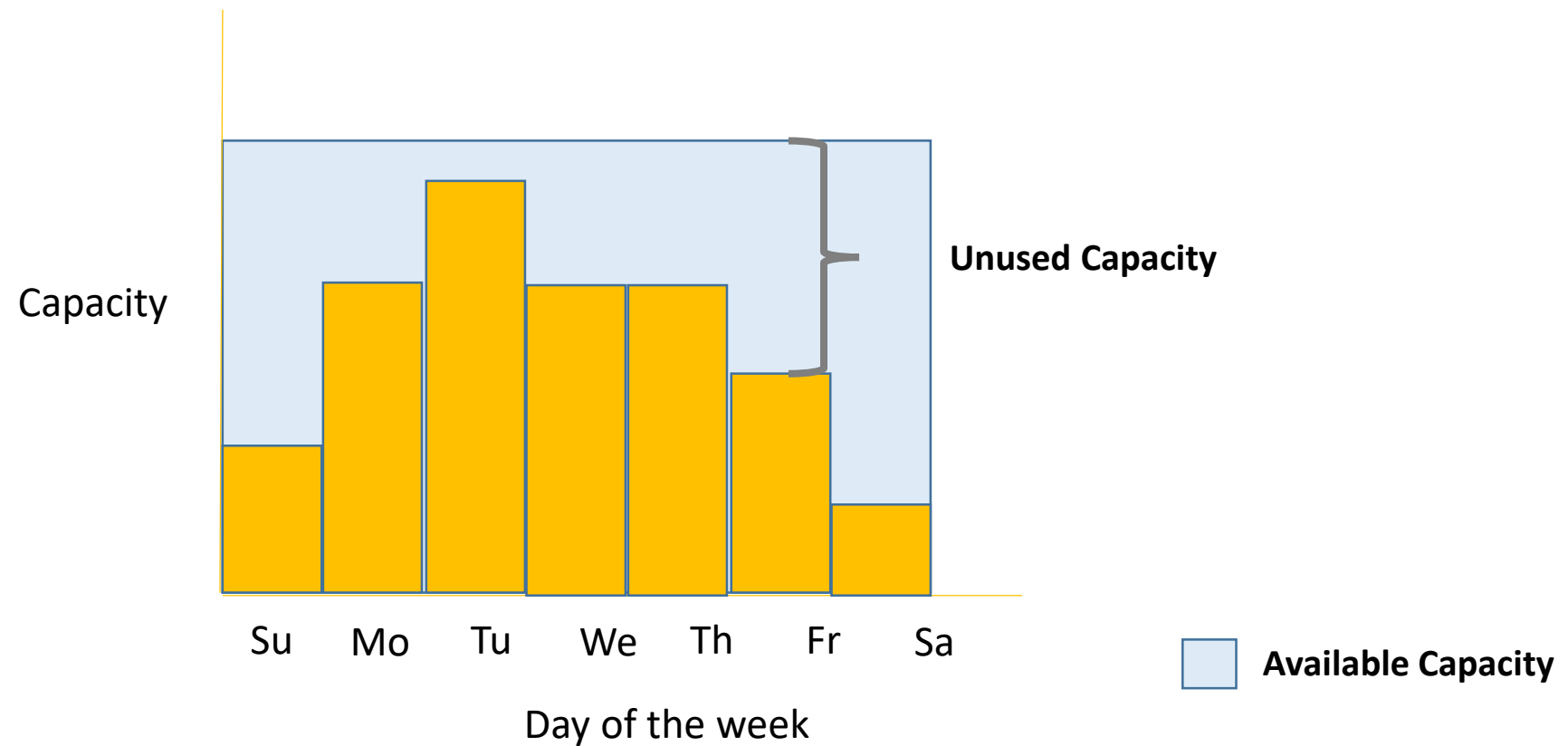
Capacity Management Example

Finance Department – Accounts Application



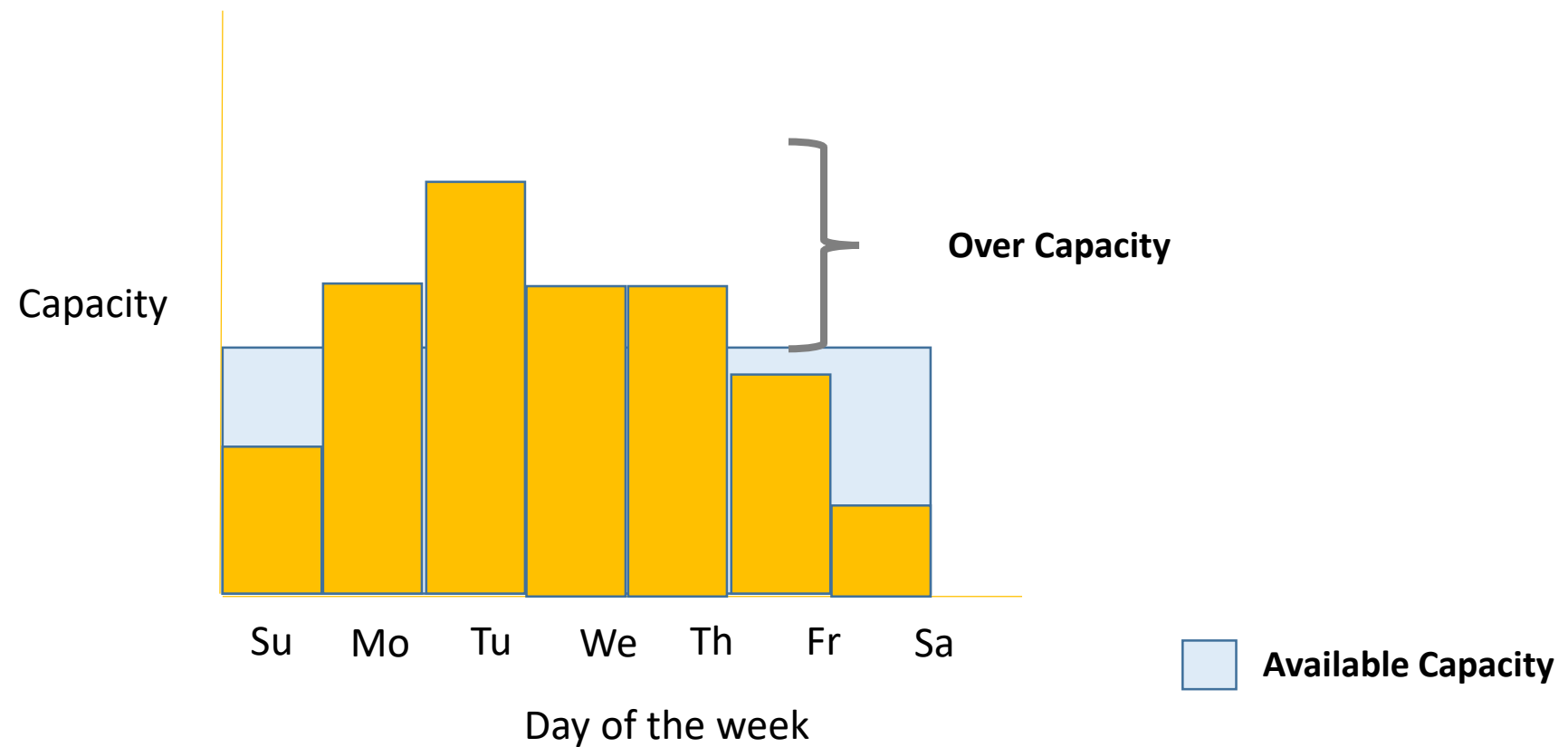
Capacity Management Example

Finance Department – Accounts Application



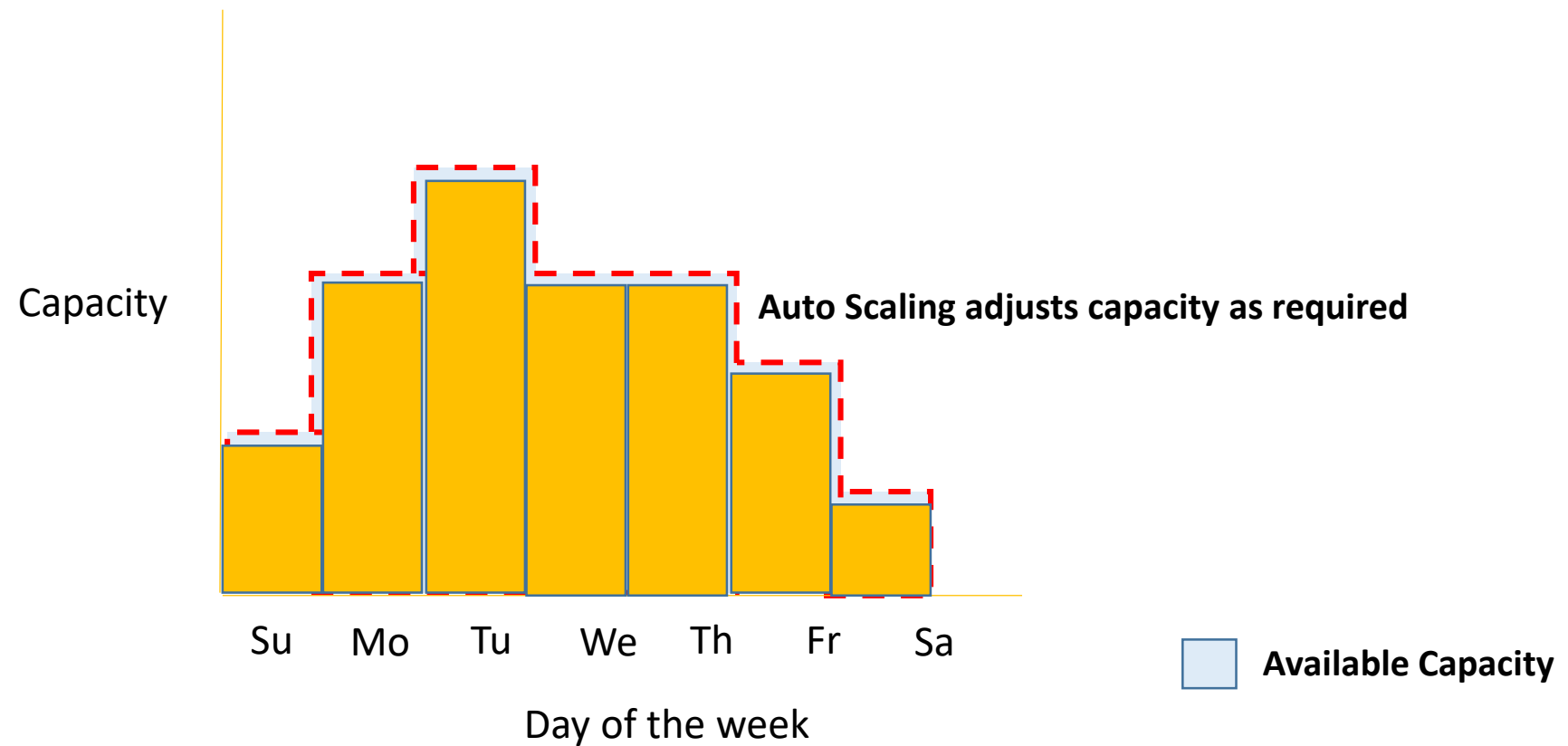
Capacity Management Example

Finance Department – Accounts Application



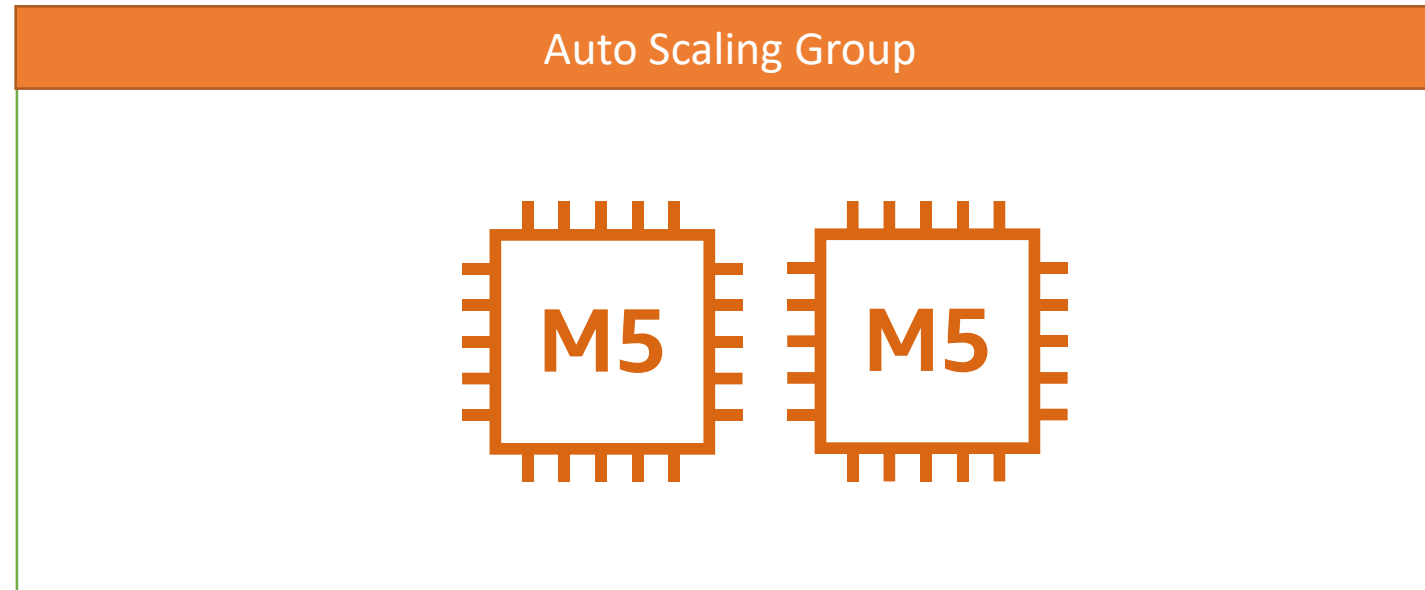
Capacity Management Example

Finance Department – Accounts Application



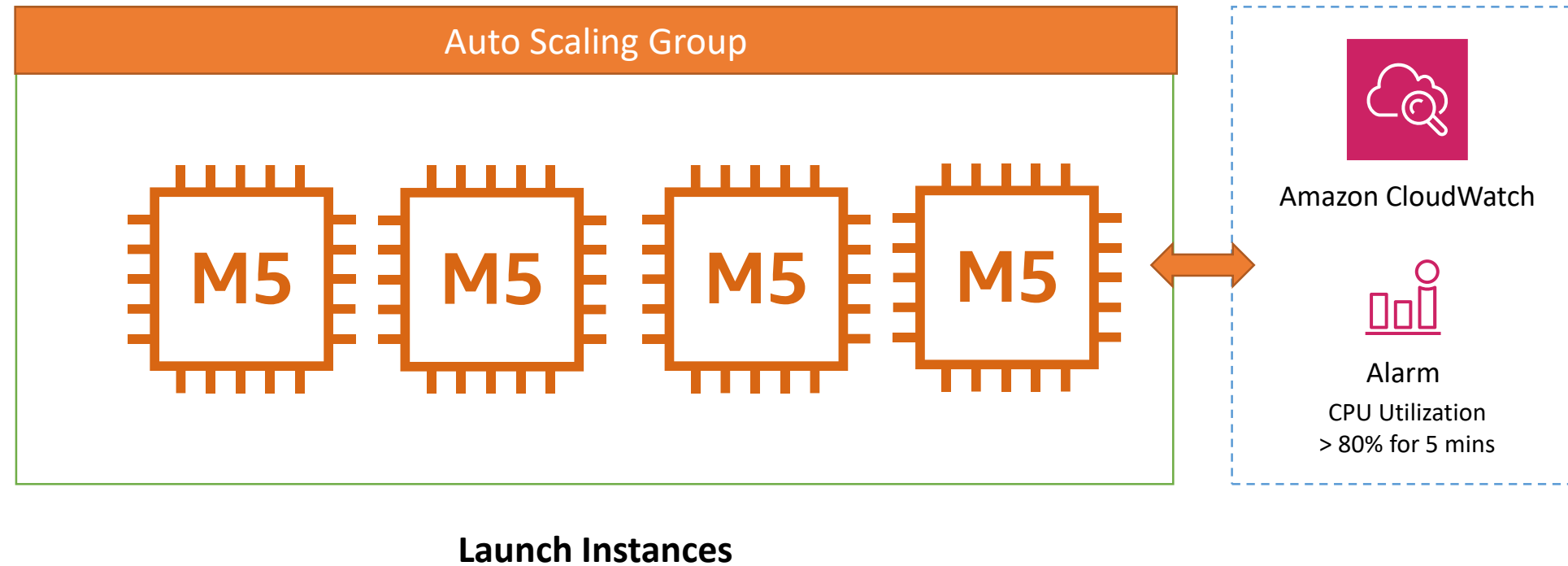
Auto Scaling Process

Initial Fleet Size



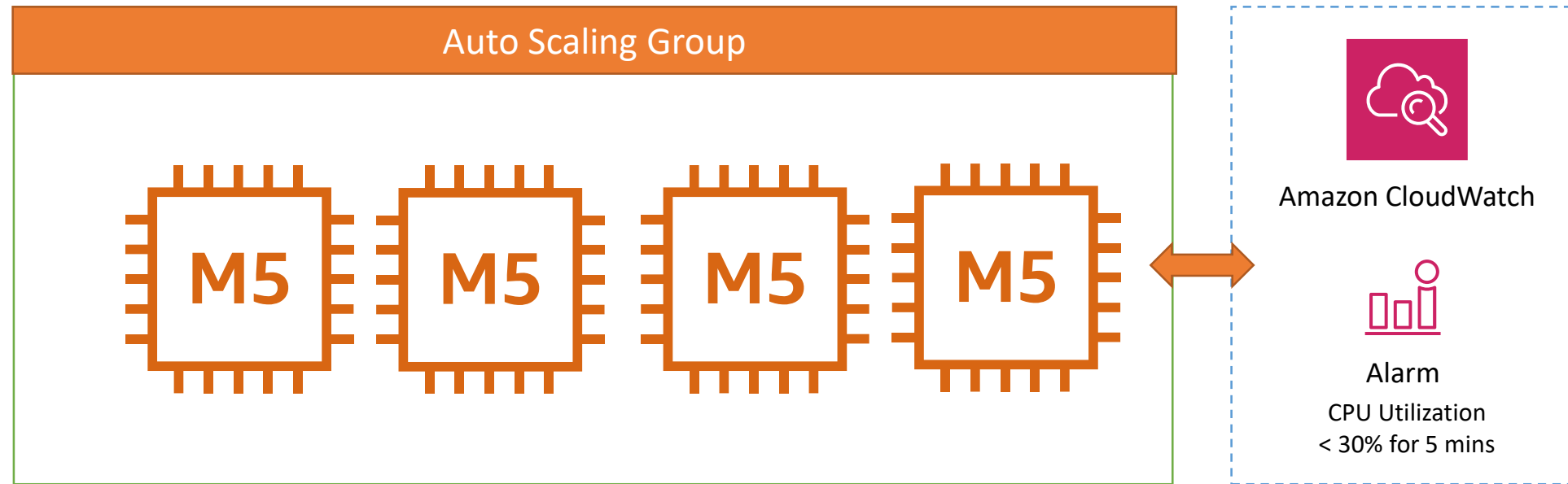
Auto Scaling Process

Scaling Out



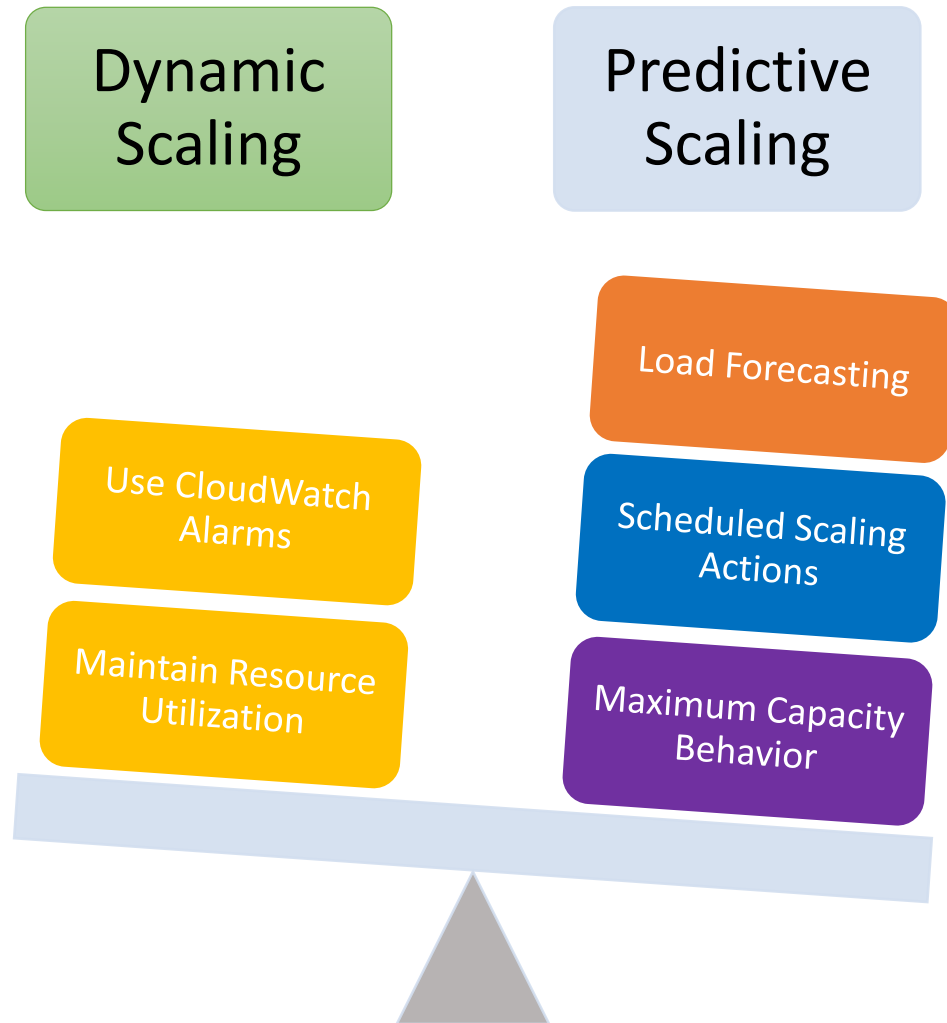
Auto Scaling Process

Scaling In



Terminate Instances

Dynamic & Predictive Scaling



Auto Scaling Components



Auto Scaling Components



AMI ID
Instance Type
Key Pairs
Security Group

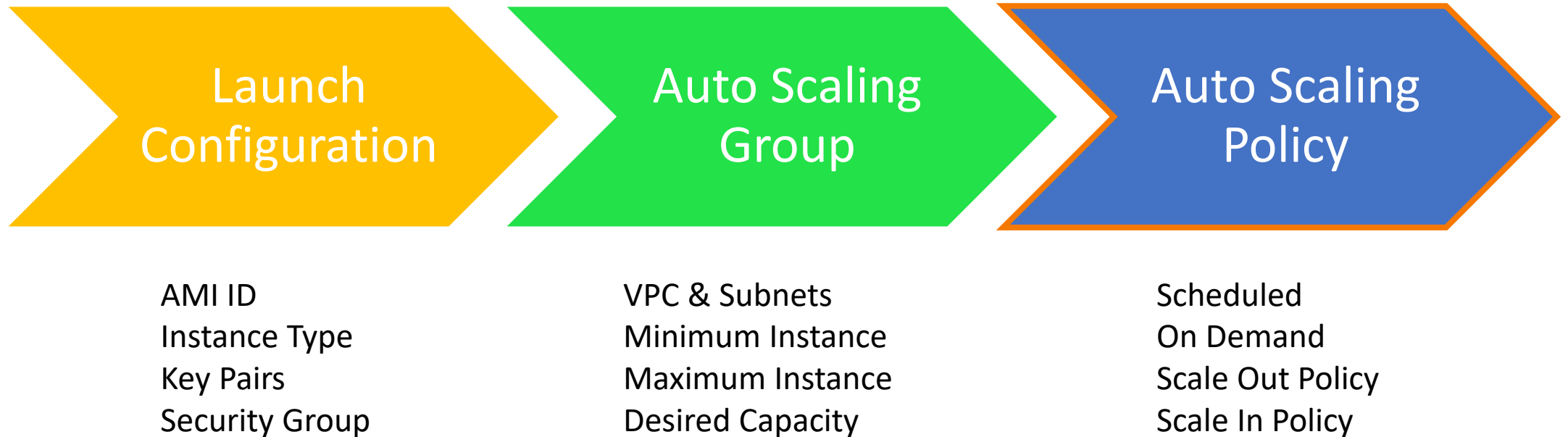
Auto Scaling Components



AMI ID
Instance Type
Key Pairs
Security Group

VPC & Subnets
Minimum Instance
Maximum Instance
Desired Capacity

Auto Scaling Components



Next Video

Configure Auto Scaling