Reference: **FAQs** Amazon EC2 Auto Scaling helps you ensure that you have the correct number of Amazon EC2 instances available to handle the What? load for your application. You create collections of EC2 instances, called Auto Scaling groups (ASG) by specifying minimum, **Category:** maximum and desired number of instances. Compute You can use the fleet management features of EC2 Auto Scaling to maintain the health and availability of your fleet. An Why? application running on Amazon EC2 instances, is referred as a 'fleet'. It helps you in better fault tolerance, Better availability Better cost management for your application. • You want to automatically launch or terminate EC2 instances based on user-defined scaling policies, scheduled actions, When? and health checks. It supports manual scaling, scheduled scaling, dynamic scaling and predictive scaling. • You want to reduce the need to manually provision Amazon EC2 capacity in advance. Amazon EC2 Auto • EC2 Auto Scaling groups are regional constructs. They can span Availability Zones, but not AWS regions. Scaling Where? • It lets you provision and automatically scale instances across purchase options, AZs, and instance families. It scales dynamically based on your Amazon CloudWatch metrics, or predictably according to a schedule that you define. Who? EC2 Auto Scaling monitors the health of running instances, automatically replaces impaired instances, and balances **Complete book:** capacity across Availability Zones. Click Here When you create an EC2 Auto Scaling group, you must specify a launch configuration. Using this launch configuration How? Amazon EC2 Auto Scaling always launches new instances such that they are balanced between availability zones. **Created by:** Predictive Scaling's machine learning algorithms detect changes in daily and weekly patterns, and adjust their forecasts. Ashish Prajapati How • Amazon EC2 Auto Scaling features have no additional fees beyond the service fees for Amazon EC2, CloudWatch (for scaling much? policies), and the other AWS resources that you use.