INNOVATE

MODERN APPLICATIONS EDITION

20 October, 2022

Increase fault-tolerance of high-scale applications with AWS Fault Injection Simulator (FIS)

Ashwini Kumar
Senior Amazon EC2 Spot Specialist Solutions Architect,
AISPL

Somnath Chatterjee
Technical Account Manager
AISPL

Agenda

- A customer’s wish list
- Our roadmap to the solution
- Solution components
    - Amazon EC2 Auto Scaling groups
    - Amazon EC2 Spot
    - AWS Fault Injection Simulator (FIS)
- Solution architecture
- Demo

A customer’s wish list

- Deploy and run a music streaming web application
- Automate application deployment and scaling
- Maintain health and capacity of required compute at scale
- Load balance application traffic for performance and availability
- Provision and scale compute with optimized cost
- Increase application fault-tolerance and availability during server failures/ instance terminations
- Data resiliency during server failures/ instance terminations

Our Roadmap to the Solution

Customer requirement	Solution components on AWS
Automate deployment and scaling Maintain health and capacity of compute	 AWS CloudFormation  AWS CodeDeploy  Amazon EC2 Auto Scaling
Load balance application traffic for performance and high availability Increase fault-tolerance and availability during server failures/ instance terminations	 Application Load Balancer  Amazon EC2 Auto Scaling  AWS Fault Injection Simulator
Provision and scale compute with optimized cost	 Amazon EC2 Auto Scaling  Amazon EC2 Spot instances
Data resiliency during server failures	 Primary Standby Amazon RDS  Amazon Elastic File System

Why Amazon EC2 Auto Scaling ?

a. Support for Mixed instance types & Purchase options

b. Replacement of unhealthy instances

c. Balanced capacity across availability zones

d. Elastic Load Balancer integration

e. Proactive replacement of Spot instances, which are at higher risk of interruptions

f. Lifecycle hooks, termination policies, instance termination protection, scaling policies…

Let EC2 Auto Scaling do the Heavy Lifting

```
AutoScalingGroupName: my-asg
CapacityRebalance: true
DesiredCapacity: 4
MinSize: 4
MaxSize: 100
MixedInstancesPolicy:
  InstancesDistribution:
    OnDemandBaseCapacity: 3
    OnDemandPercentageAboveBaseCapacity: 50
    SpotAllocationStrategy: capacity-optimized
  LaunchTemplate:
    LaunchTemplateSpecification:
      LaunchTemplateName: my-launch-template
      Version: $Default
    Overrides:
    - InstanceRequirements:
        VCpuCount:
          Min: 2
          Max: 4
        MemoryMiB:
          Min: 2048
        CpuManufacturers:
        - intel
VPCZoneIdentifier: subnet-1,subnet-2,subnet-3
```

5. Let EC2 Auto Scaling proactively rebalance Spot Instances with higher risk of interruption to increase application availability

1. Let EC2 Auto Scaling provision a combination of On-Demand and Spot Instances

3. Set a base capacity with On-Demand Instances

4. Use the capacity-optimized allocation strategy to find most available Spot capacity

2. Specify Instance Attributes and Availability Zones and let EC2 Auto Scaling to pick the right instances, optimize for performance, cost, and availability

aws

Why Amazon EC2 Spot Instances?

Up to 90% off

EC2 Spot infrastructure

Is **same** as On-Demand and RIs - pools of spare unused capacity

EC2 Spot pricing

Smooth, infrequent changes, no spikes, more predictable

(no bidding)

Interruptions

Happen when EC2 needs to reclaim capacity or when max price threshold is crossed

(no bidding)

Diversification

Best practice - choose different instance types, sizes and AZs in a single fleet – to access more pools of spare capacity

EC2 Spot is more than cost savings……

More Compute  + Faster Results  = Accelerated Innovation 

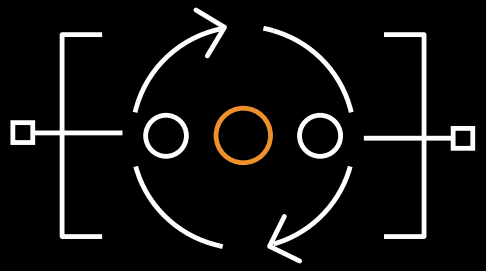
aws

What About Spot Interruptions?

Minimal interruptions

Over **95%** of the instances were not interrupted in the last 3 months

The work you do to make your applications fault-tolerant also make them Spot-ready



Use Spot instances for stateless, fault-tolerant, or flexible workloads.

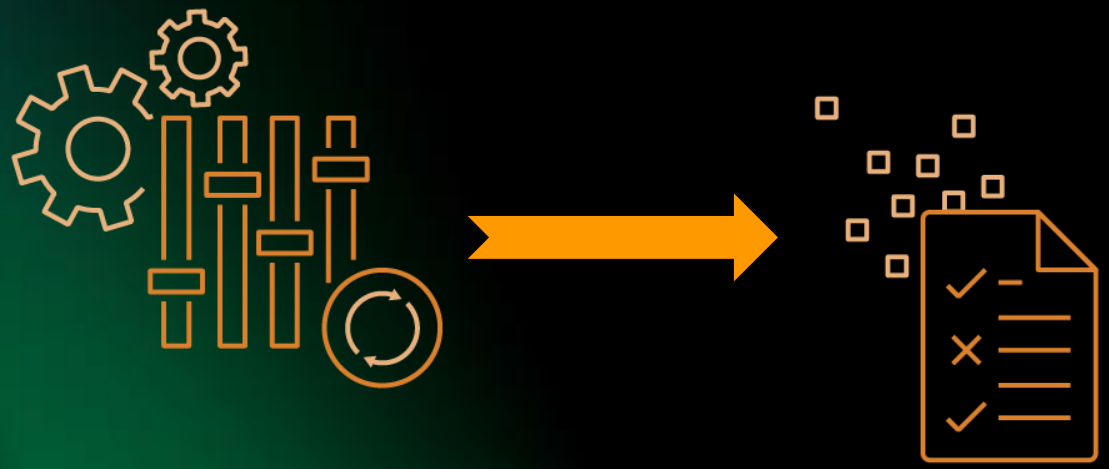
Any application that can have part or all of the work, paused and resumed or restarted, can use Spot

Leverage 2-minute instance termination notice via instance metadata or CloudWatch Events for:

- ☑ Checkpointing your work
- ☑ Draining gracefully from Elastic Load Balancer

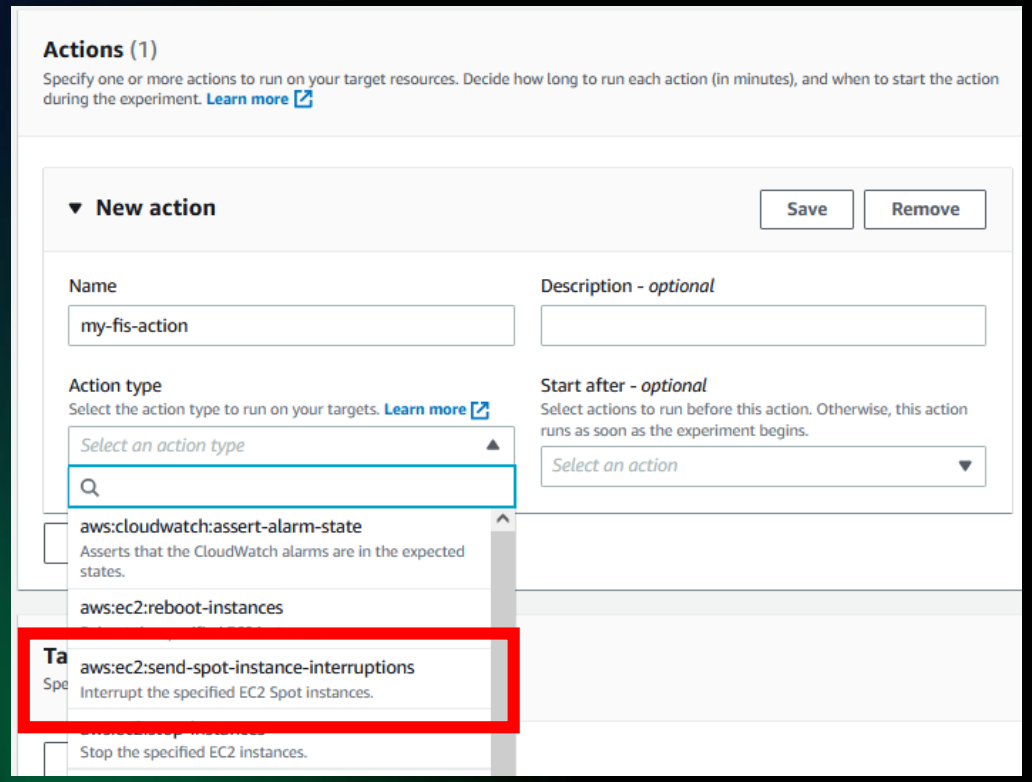
aws

Rebalance Recommendation Signal for Spot



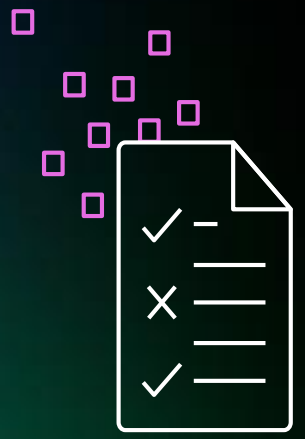
- A signal that notifies you when a Spot Instance is at elevated risk of interruption
- The signal can arrive sooner than the 2-minute Spot Instance interruption notice
- Gives an opportunity to proactively rebalance your workload to new or existing Spot Instances that are not at elevated risk of interruption
- Start checkpointing work early to save as much state as possible
- Prevent scheduling new work on instances at elevated risk of interruption, thus increasing chance of completing work

Why AWS Fault Injection Simulator ?



- AWS Fault Injection Simulator (FIS) is a managed service for chaos engineering
- AWS FIS can simulate Spot instance interruptions
- The actual Spot interruption is preceded by both Rebalance recommendation signal and 2-min Interruption notice

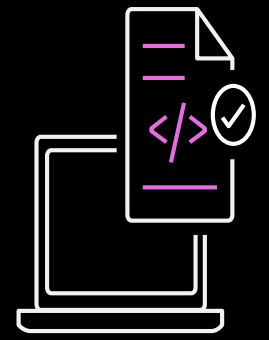
AWS FIS Components



Actions

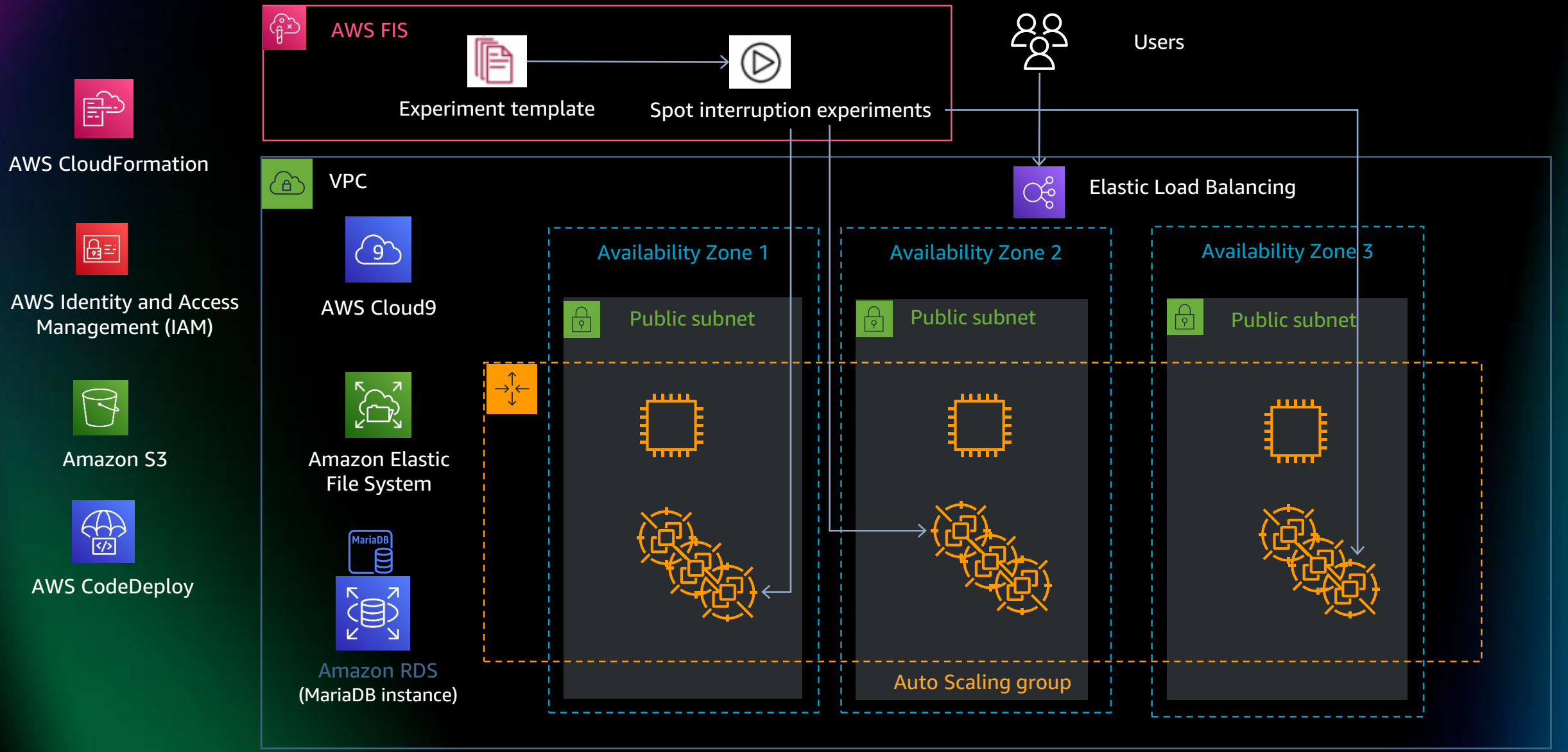
Targets

Experiment templates



Experiments

Solution Architecture



Demo

Additional resources

Blog post: Running high-scale web applications on Amazon EC2 Spot instances
https://aws.amazon.com/blogs/compute/running-high-scale-web-on-spot-instances/

Blog post: Implement interruption tolerance with Amazon EC2 Spot using AWS Fault Injection Simulator
https://aws.amazon.com/blogs/compute/implementing-interruption-tolerance-in-amazon-ec2-spot-with-aws-fault-injection-simulator/

Blog post: Proactively manage Spot instance lifecycle using Capacity Rebalancing for EC2 Auto Scaling
https://aws.amazon.com/blogs/compute/proactively-manage-spot-instance-lifecycle-using-the-new-capacity-rebalancing-feature-for-ec2-auto-scaling/

Workshop: Run EC2 workloads at scale with EC2 Auto Scaling
https://ec2spotworkshops.com/running-amazon-ec2-workloads-at-scale.html

aws

Thank you for attending AWS Innovate Modern Applications Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apj-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws

Thank you!