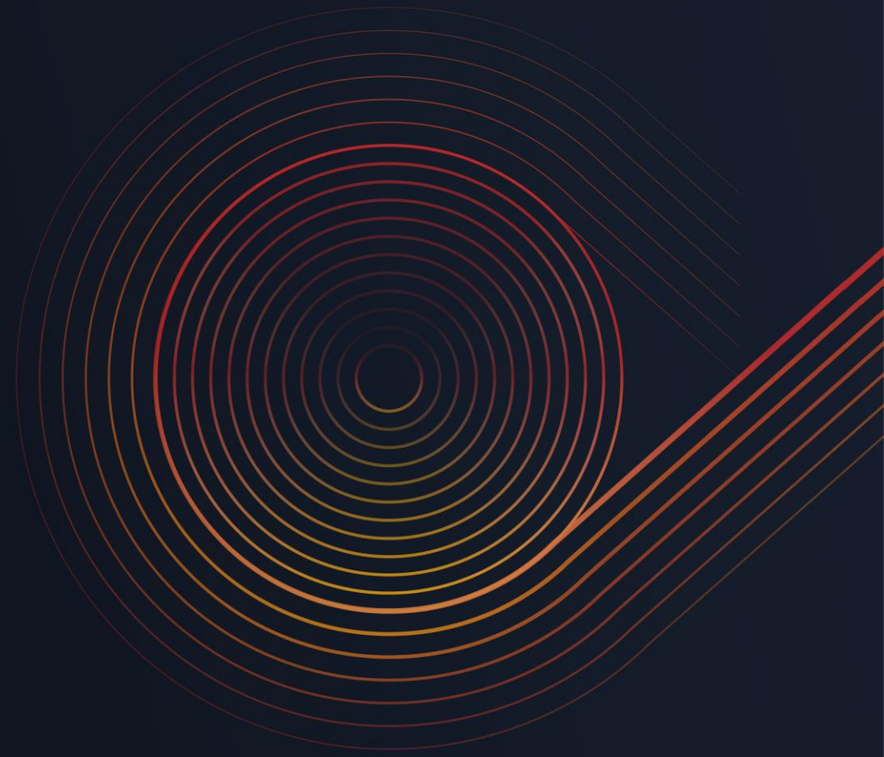
aws modern apps

Managing Costs for Serverless and Container Workloads

Michael Fischer - Sr. Specialist Solutions Architect, Containers
Dan Fox - Principal Specialist Solutions Architect, Serverless



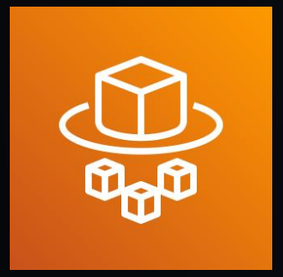
Agenda

Right sizing

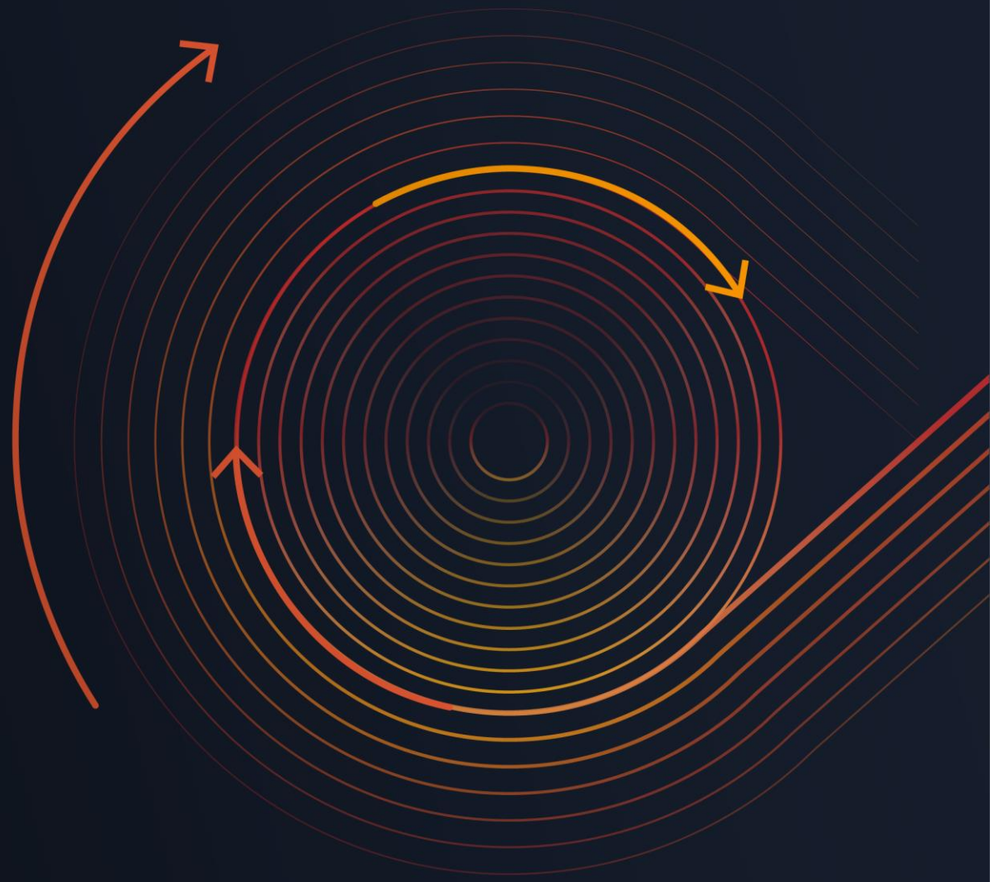
- AWS Fargate
- AWS Lambda
- Serverless Architectures

Cost optimization

- Spot capacity
- AWS Graviton2
- Compute Savings Plans



AWS Fargate



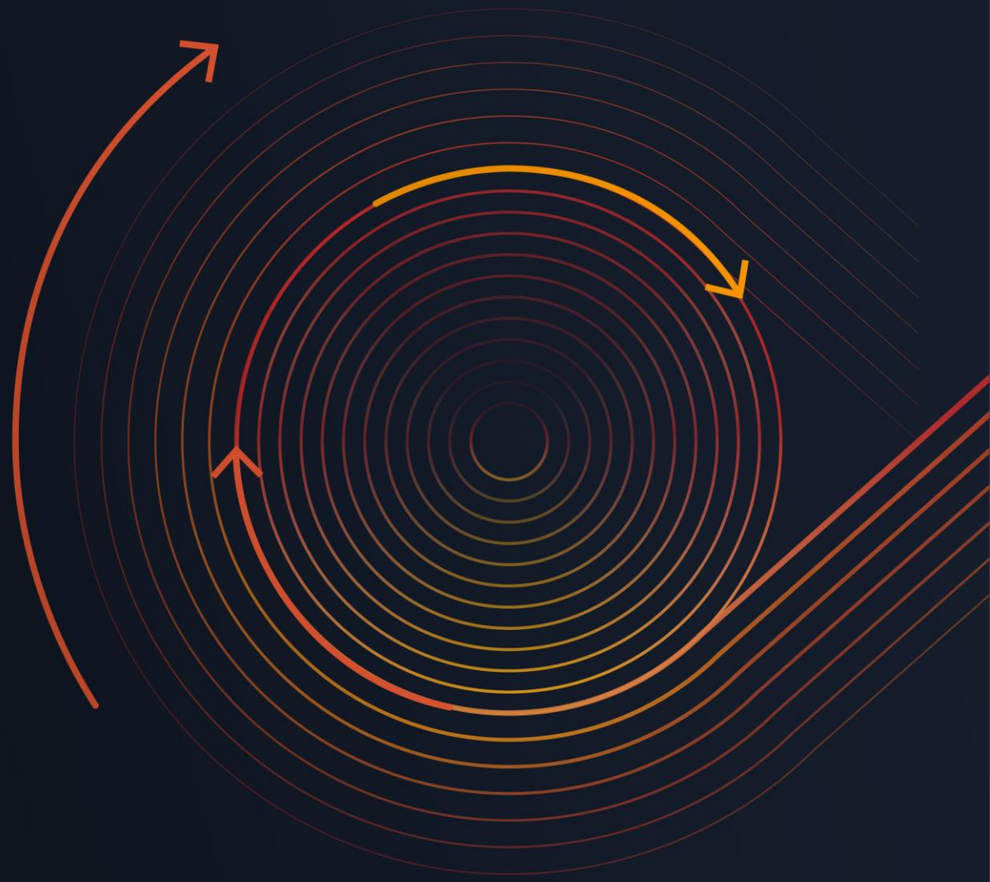
aws modern apps

Fargate

- Managed compute for containers
- Eliminates wasted capacity
- Lowers TCO

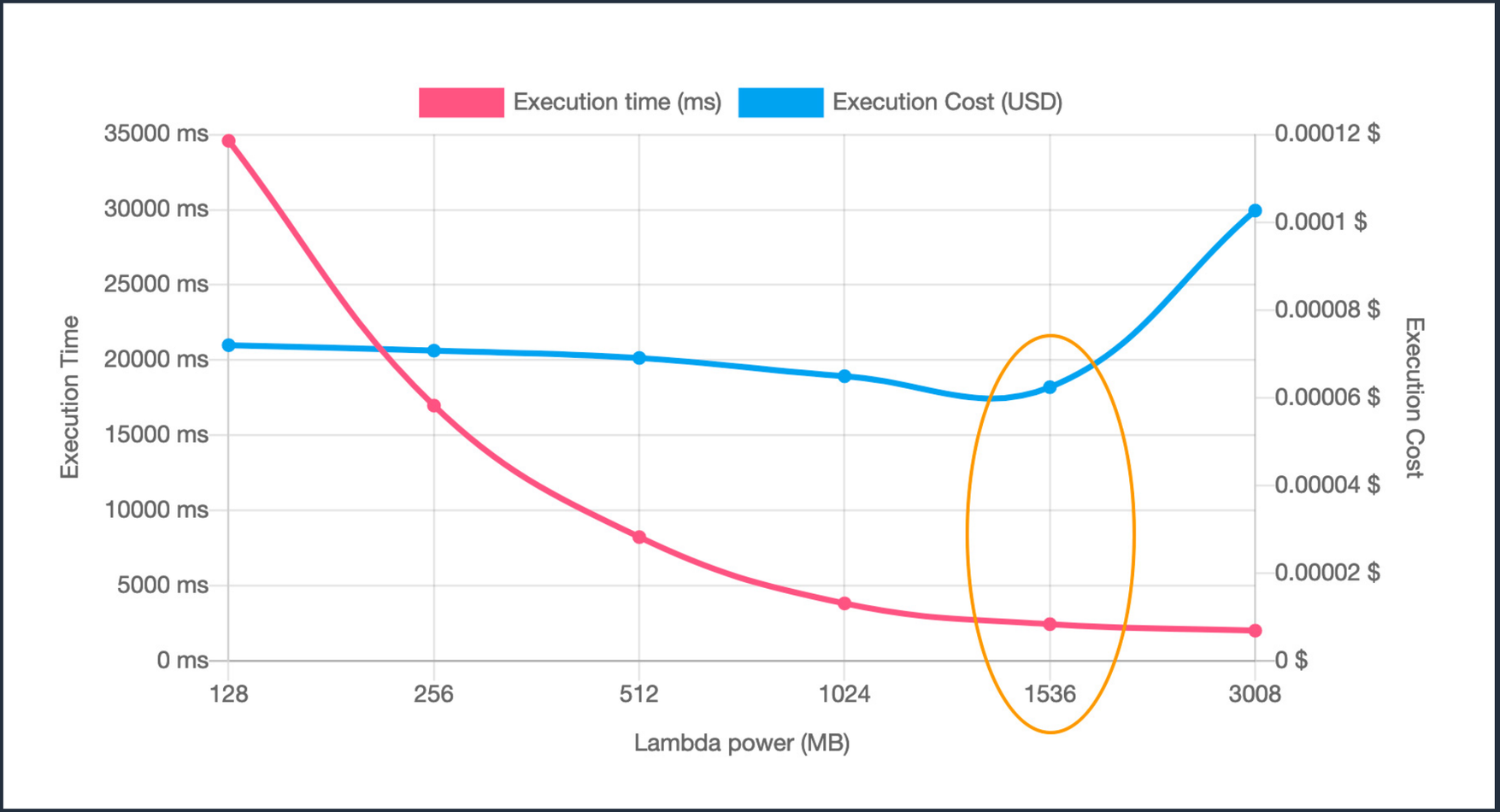
aws modern apps

AWS Lambda



aws modern apps

Lambda Power Tuning

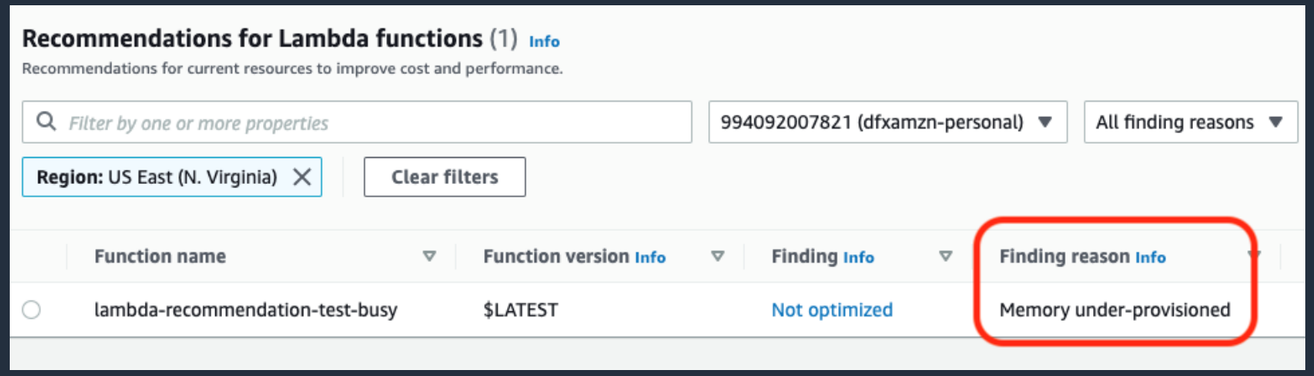


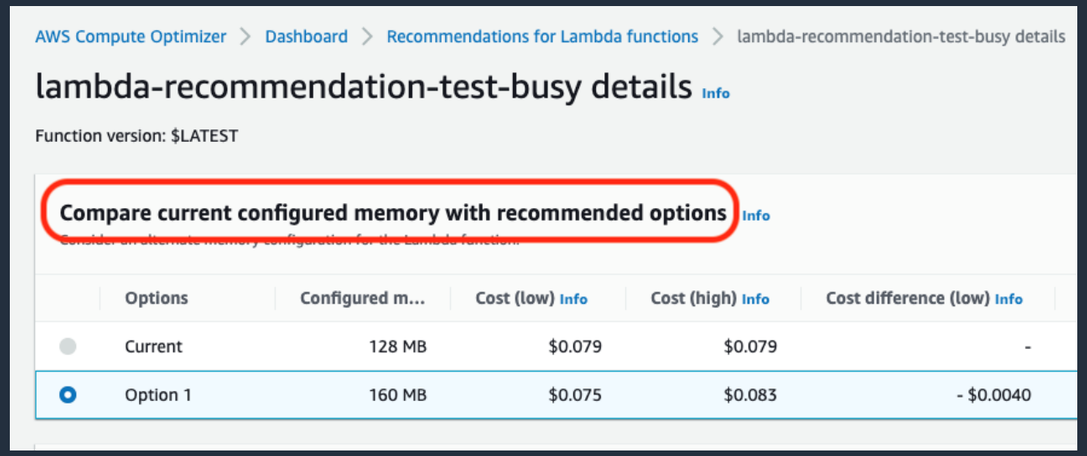
https://github.com/alexcasalboni/aws-lambda-power-tuning

aws modern apps

Lambda - Compute Optimizer

```
➜  ~ aws compute-optimizer update-enrollment-status --status Active
```

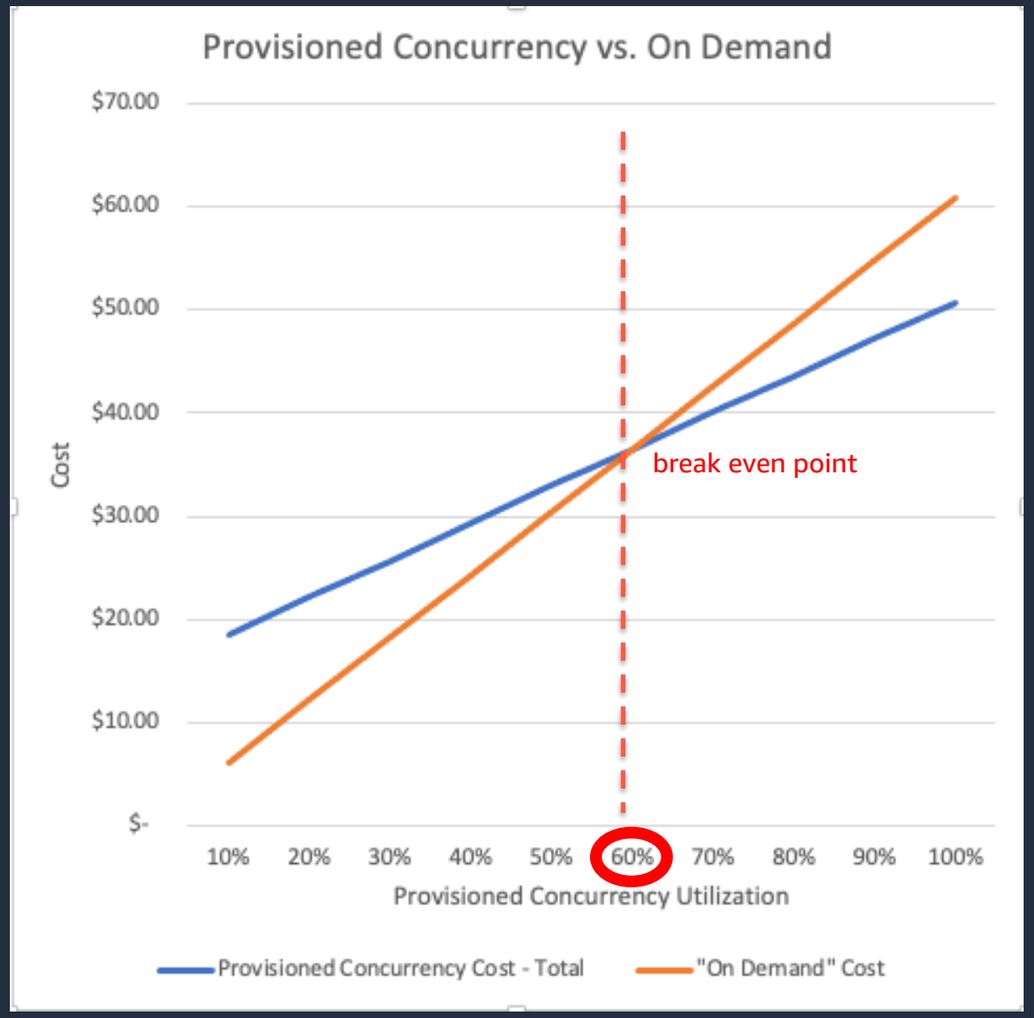




aws modern apps

Lambda - Provisioned concurrency

- PC limits the impact of cold starts
- Effective cost of compute is lower than on-demand if well utilized
- Break even for utilization is >60%



aws modern apps

Lambda - Log tuning

- Lambda uses CloudWatch to store execution output logs
- CloudWatch costs are based on ingestion and storage
- Set appropriate logging levels and remove unnecessary logging information to optimize log ingestion
- Set log retention periods for new and existing CloudWatch Logs groups
- Archive logs with cost-effective storage classes

Lambda – AVX2

Spot Capacity

aws modern apps

Spot options

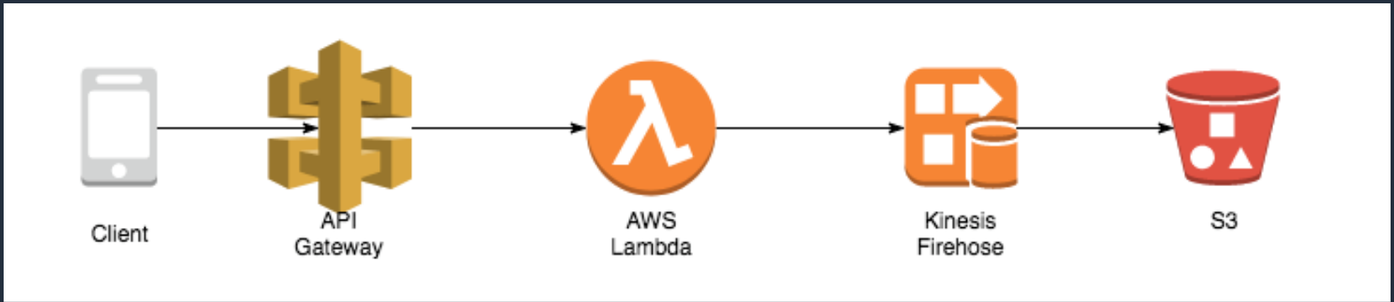
- Fargate Spot for ECS: Save up to 70% over Fargate
    - Blend with Fargate on-demand easily using Capacity Providers
- EC2 Spot Capacity for EKS and ECS
    - Choice of price- or capacity-optimized
    - Instance type and Availability Zone flexibility improves availability
- Outstanding choice for:
    - Overprovisioned (“burst”) capacity
    - Batch processing/ML workloads
    - Lower environments (staging, dev, etc.)
    - Latency-tolerant services

Serverless Architectures

aws modern apps

Lambda-free Serverless – Direct integrations

Sending data to Amazon S3 using Kinesis Data Firehose



Reducing cost by implementing AWS service proxy

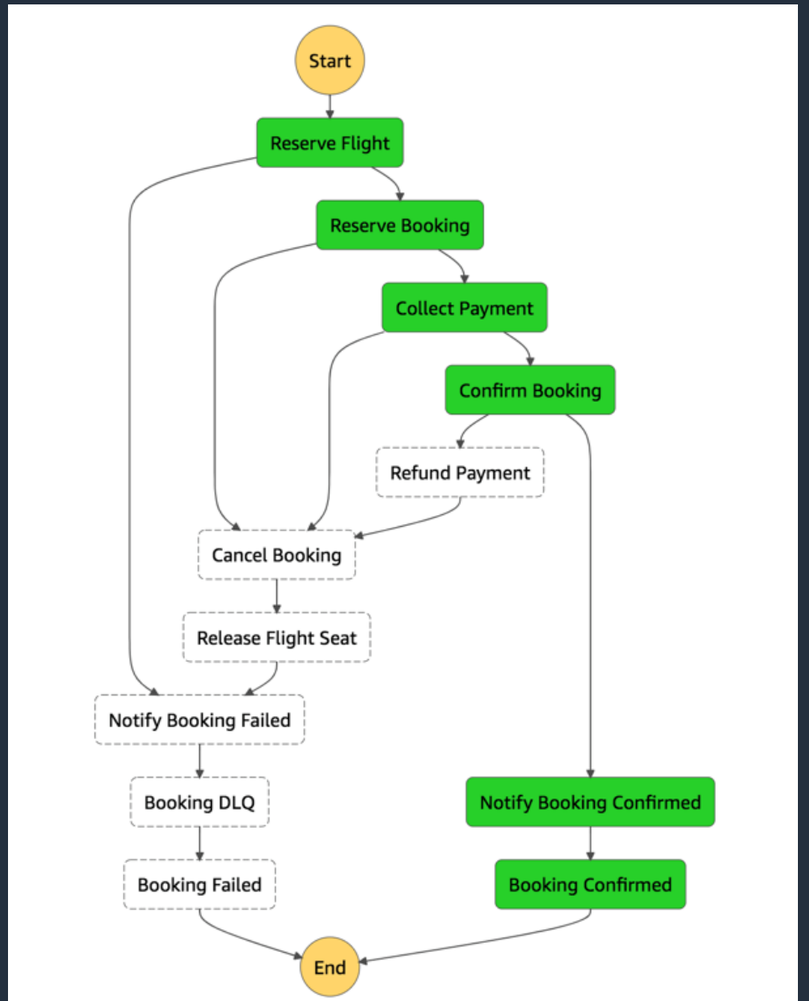
aws modern apps

Lambda-free Serverless – Step functions

Use case considerations:

- Long running processing tasks
- Lambda functions as orchestrators
- Lambda functions calling other Lambda functions

Consider Step Functions



Booking service Step Functions state machine

aws modern apps

AWS Graviton2

aws modern apps

Graviton2

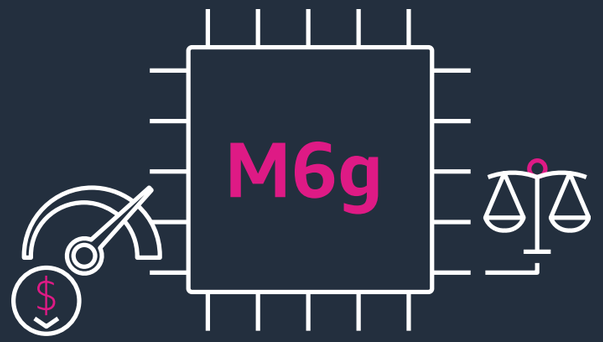
- Available on EC2, ECS, EKS
- Up to 40% **better performance** than comparable Intel-based instances
- 20% **less expensive** than comparable Intel instances
- 2-3.5x better performance per watt

Graviton2

aws modern apps

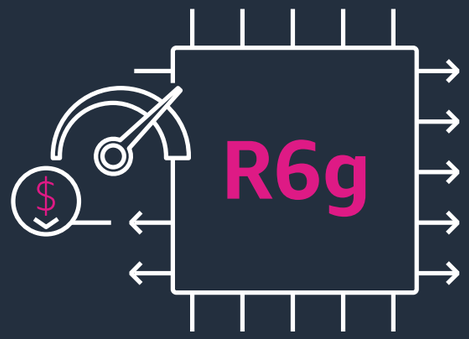
Graviton2

Instance types for every need



General purpose:
M6g, M6gd, T4g

Compute optimized:
C6g, C6gd, C6gn



Memory optimized:
R6g, R6gd

aws modern apps

Graviton2

Can your workload run on Graviton2?

- Most likely, yes!
- Many workloads can be run “as is”
    - Scripting languages: Python, Javascript/TypeScript, Ruby, PHP
    - Bytecode runtimes: Java 8+, .NET Core 3.1, .NET 5
    - Watch out for architecture-specific dependencies: shared libraries, JNI
- Most C/C++ apps and libraries can be recompiled easily
- Go: cross-compile from anywhere by setting GOARCH=arm64 and GOOS=linux

Compute Savings Plans

aws modern apps

Compute savings plans

Save up to 66% on compute costs by committing to future usage

Amazon EC2

AWS Fargate

AWS Lambda

- 1 or 3 year term plans
- Single plan can apply to all three compute types
- Migrate from EC2 to Fargate or Lambda, continue to pay plan price

aws modern apps