AWS re:Invent

SVS224-R

AWS Lambda function performance tuning

Alex Casalboni
Technical Evangelist
Amazon Web Services

AWS re:Invent

aws

Agenda

Fundamentals & news

Optimization best practices

AWS Lambda Power Tuning

Real-world examples

Whiteboard discussion

Fundamentals & news

"So what does the future look like?
All the code you ever write is business logic."

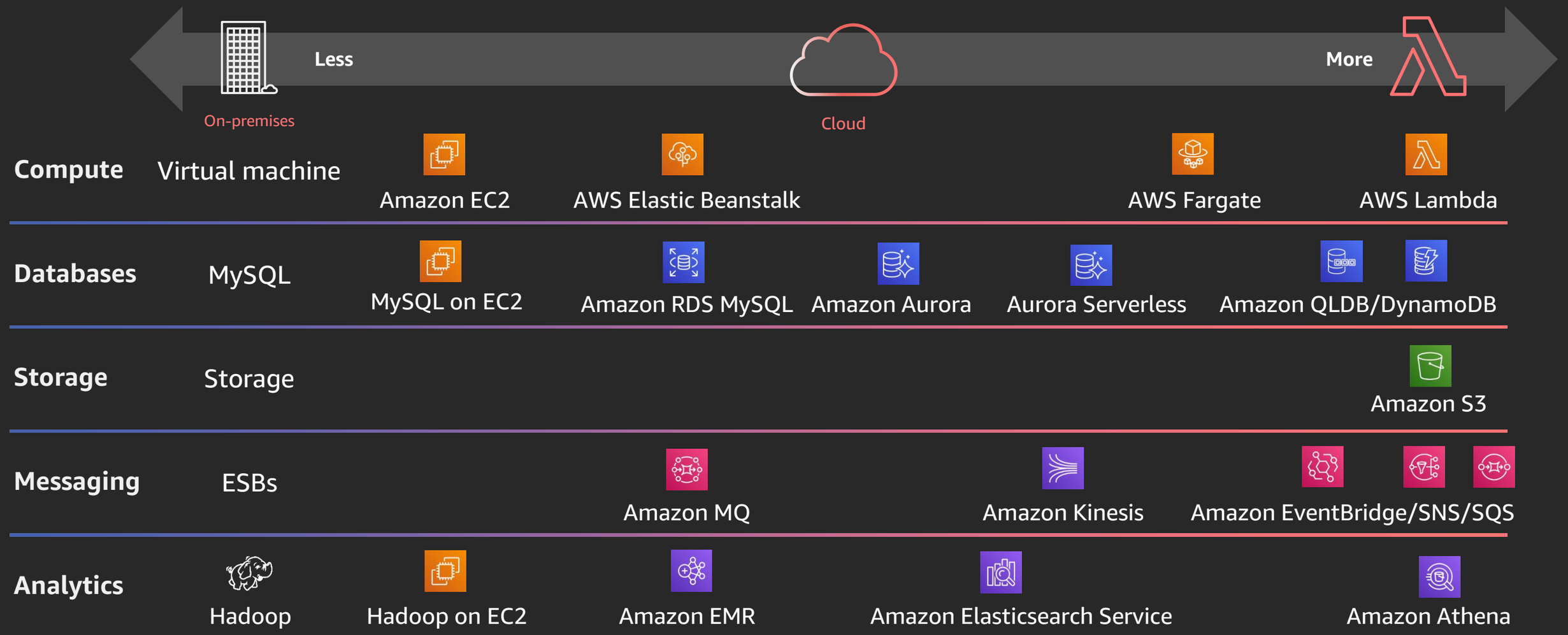
Dr. Werner Vogels

CTO of Amazon.com

AWS re:Invent

aws

AWS operational responsibility models



New!

Provisioned Concurrency for AWS Lambda

AWS re:Invent

aws

Provisioned Concurrency for AWS Lambda

Simple config to avoid cold starts
No more manual pre-warming
Predictable performance during spikes
Good fit for latency-sensitive apps
No code changes required
Bound to version or alias

AWS CloudFormation support
Scheduling with AWS Application Auto Scaling
From 1 to account concurrency limit
Two new metrics
Ramp-up time (500 per minute)
Initialization code is executed automatically

Provisioned Concurrency for AWS Lambda

Concurrency

Unreserved account concurrency **500**

- (●) Use unreserved account concurrency
- ( ) Reserve concurrency

Provisioned concurrency

To enable your function to scale without fluctuations in latency, use provisioned concurrency. Provisioned concurrency runs continually and is billed in addition to standard invocation costs. **Learn more**

Provisioned concurrency configurations (1)

Edit | Remove | Add

Find configuration

	Qualifier	Type	Provisioned concurrency	Status	Details
( )	1	version	0	In progress (0/500)	aliases: -

Provisioned Concurrency for AWS Lambda

Concurrency

Unreserved account concurrency **500**

- (●) Use unreserved account concurrency
- ( ) Reserve concurrency

Provisioned concurrency

To enable your function to scale without fluctuations in latency, use provisioned concurrency. Provisioned concurrency runs continually and is billed in addition to standard invocation costs. **Learn more**

Provisioned concurrency configurations (1)

Edit | Remove | Add

Find configuration

	Qualifier	Type	Provisioned concurrency	Status	Details
( )	1	version	500	Ready	aliases: -

Deployment frameworks

serverless framework

HashiCorp Terraform

Monitoring

DATADOG

epsagon

lumigo

New Relic

THUNDRA™

SignalFx

a Splunk® company

sumo logic

Optimization best practices

Cost & performance optimization best practices

Avoid «monolithic» functions
Optimize dependencies (and imports)
Minify/uglify production code
Lazy initialization of shared libs/objs
Externalize orchestration

Fine-tune resources allocation
Lambda Destinations
Discard uninteresting events asap
Keep in mind retry policies
Understand currency controls

Cost & performance optimization best practices

Avoid «monolithic» functions
Optimize dependencies (and imports)
Minify/uglify production code
Lazy initialization of shared libs/objs
Externalize orchestration

Fine-tune resources allocation
Lambda Destinations
Discard uninteresting events asap
Keep in mind retry policies
Understand currency controls

Resources allocation

Memory 👉 Power

CPU-bound example

"Compute **1,000 times** all prime numbers **<= 1M**"

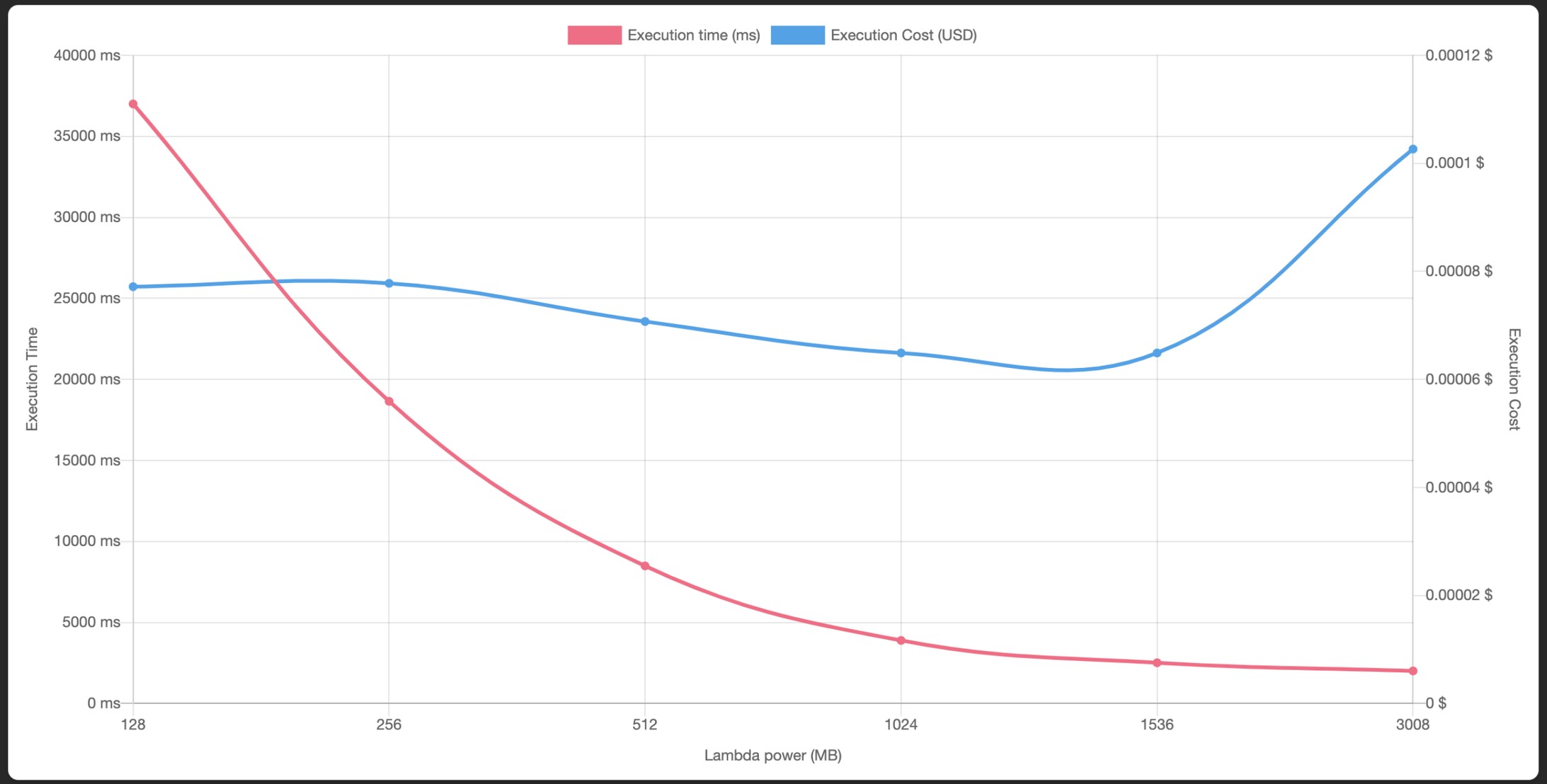
128 MB	11.722 sec	$0.024628
256 MB	6.678 sec	$0.028035
512 MB	3.194 sec	$0.026830
1024 MB	1.465 sec	$0.024638

CPU-bound example

"Compute **1,000 times** all prime numbers **<= 1M**"

128 MB	11.722 sec	$0.024628
256 MB	6.678 sec	$0.028035
512 MB	3.194 sec	$0.026830
1024 MB	1.465 sec	$0.024638

CPU-bound example



Cost-aware performance optimization

A

310ms 👉 400ms

5% performance optimization

294ms 👉 300ms

25% cost optimization

B

480ms 👉 500ms

15% performance optimization

408ms 👉 500ms

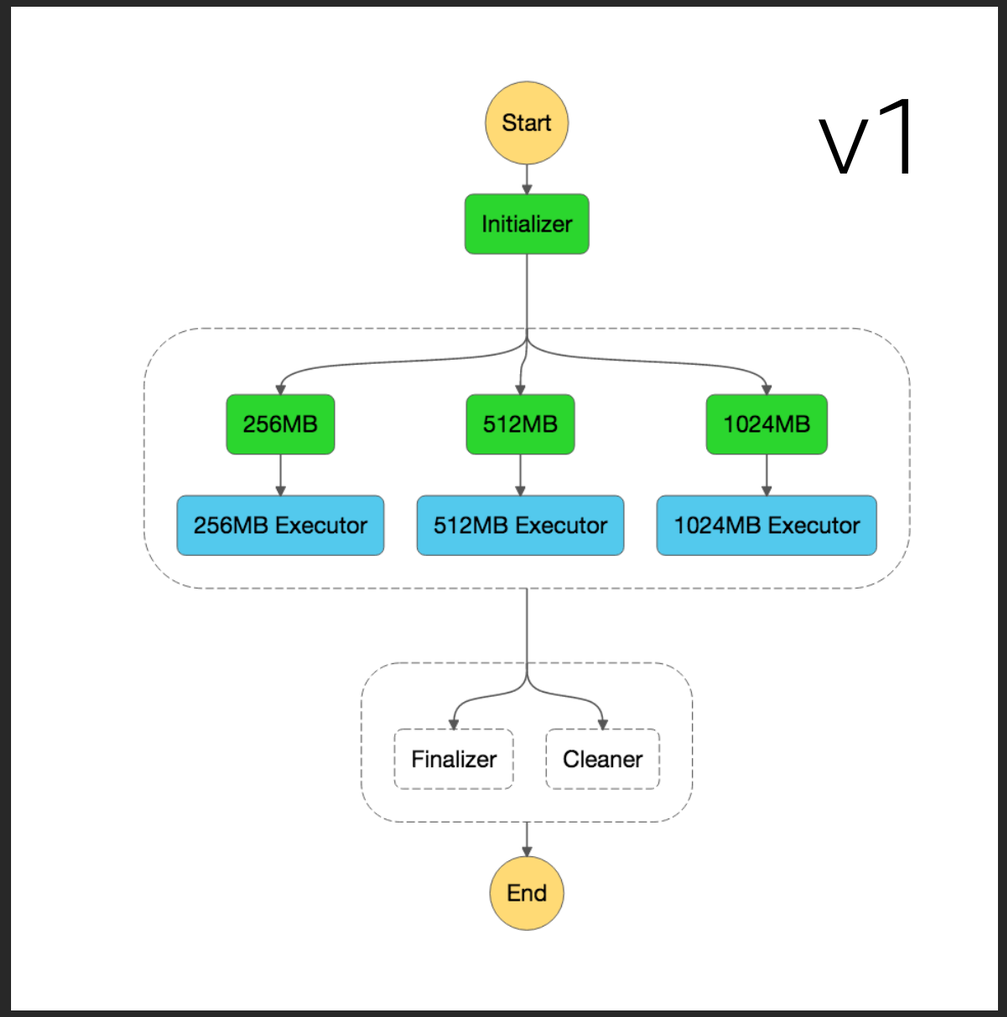
0% cost optimization

AWS Lambda Power Tuning

AWS re:Invent

aws

Don’t guesstimate!



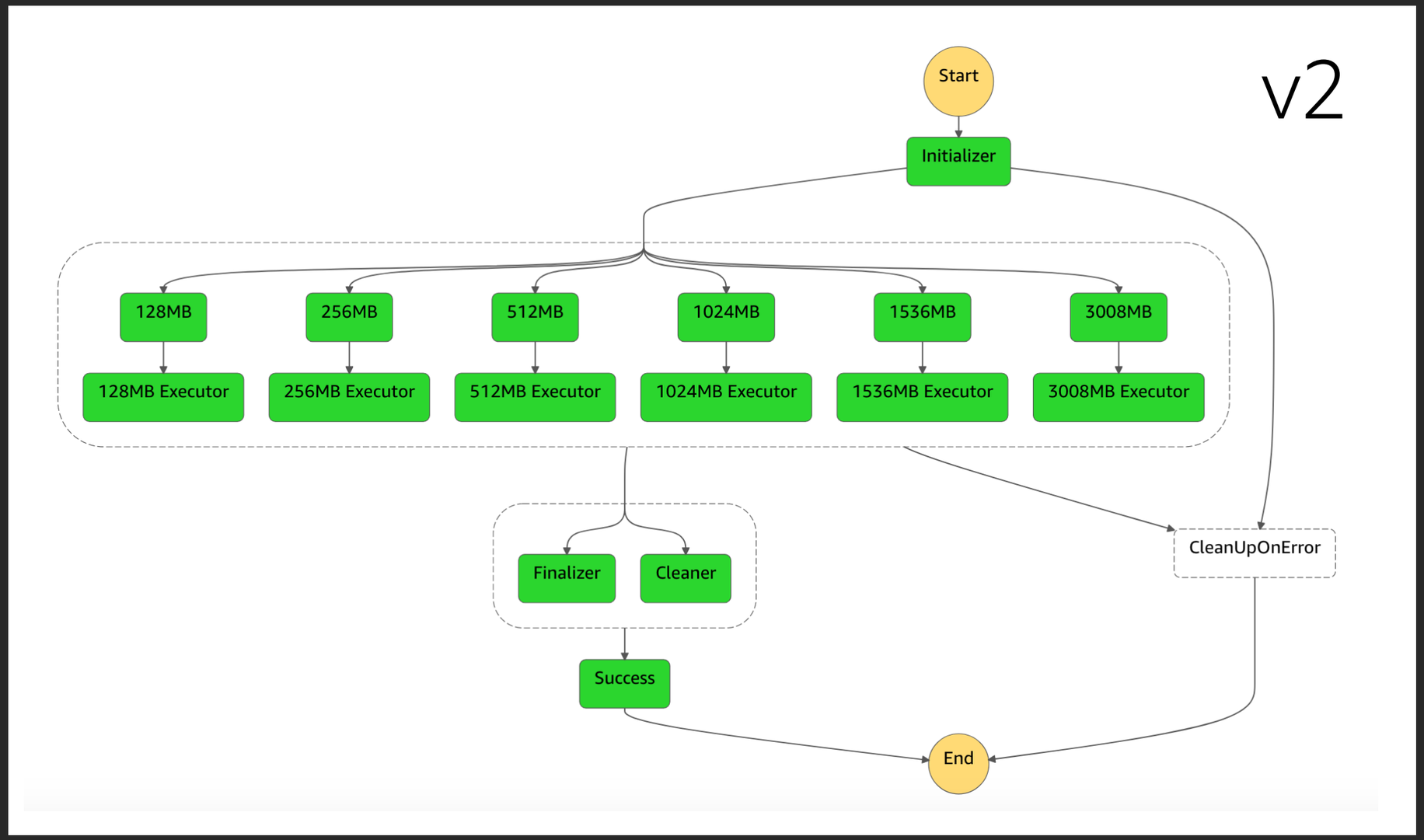
“AWS Lambda Power Tuning”

Data-driven cost & performance optimization for AWS Lambda

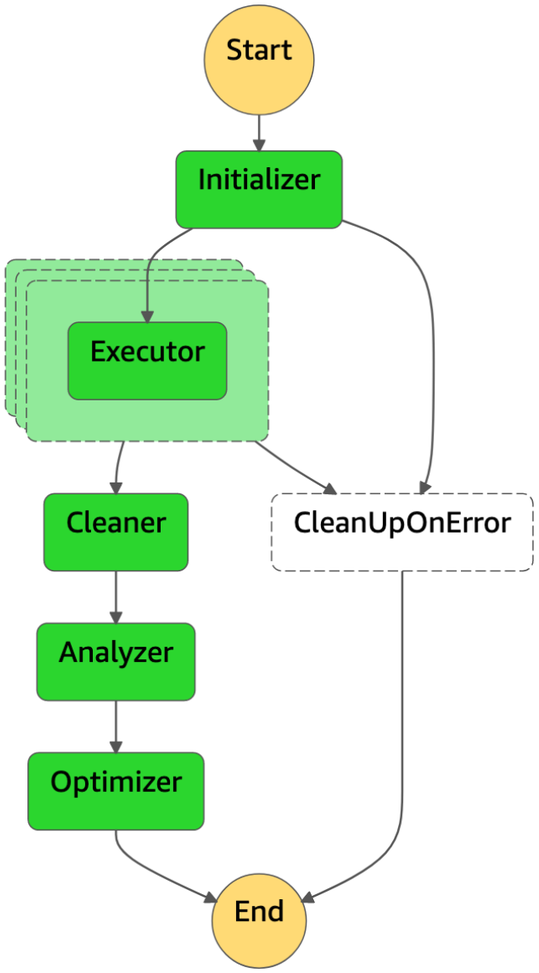
Available as a SAR app

Easy to integrate with CI/CD

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

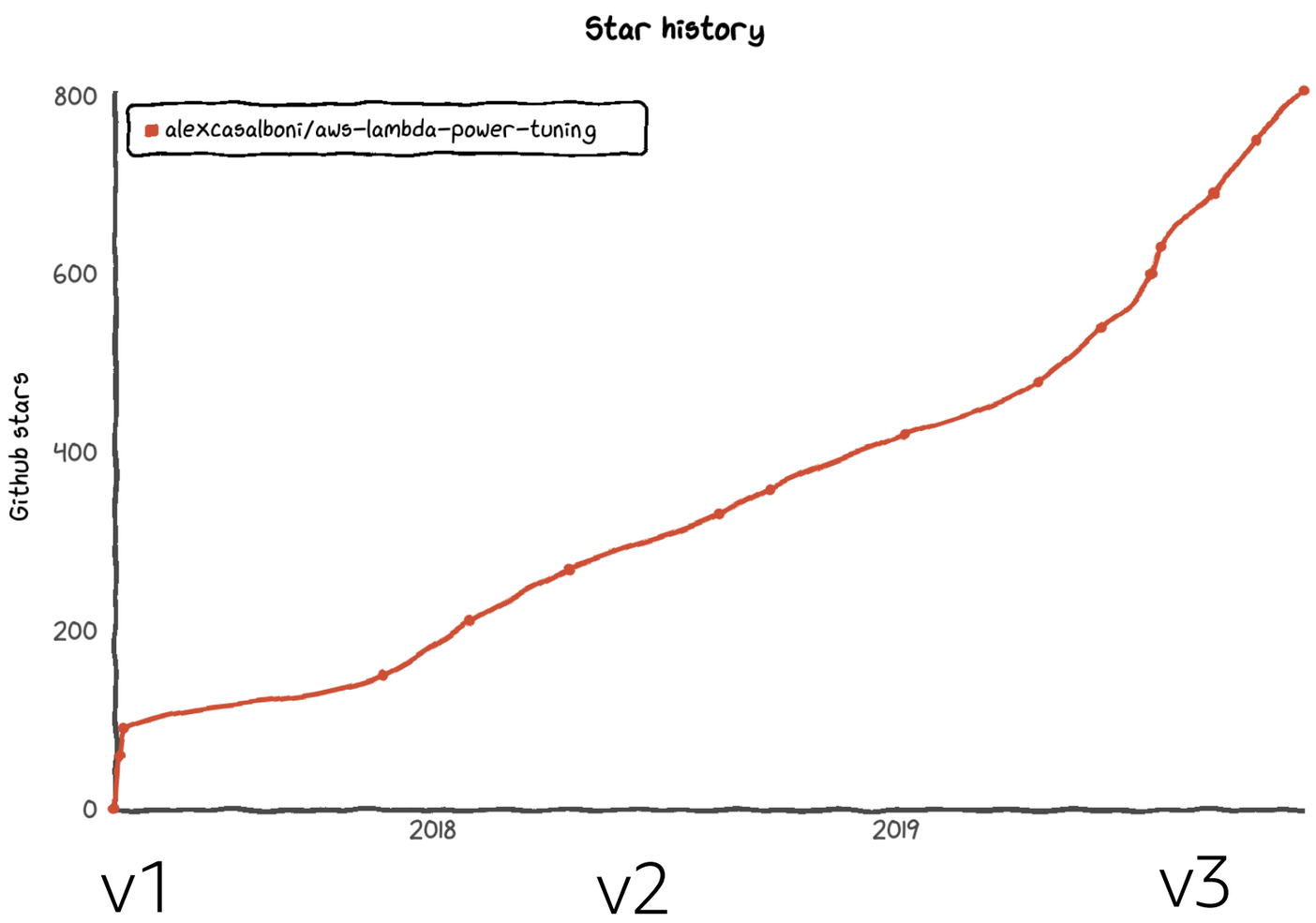


github.com/alexcasalboni/aws-lambda-power-tuning



v3

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



v1 v2 v3

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

AWS Lambda Power Tuning (input)

```
{
    "lambdaARN": "your-lambda-function-arn",
    "powerValues": [128, 256, 512, 1024, 2048, 3008],
    "num": 100,
    "payload": {"data": "abc"}
}
```

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

AWS Lambda Power Tuning (input)

```
{
    "lambdaARN": "your-lambda-function-arn",
    "powerValues": 'ALL',
    "num": 100,
    "payload": {"data": "abc"}
}
```

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

AWS Lambda Power Tuning (input)

```
{
    "lambdaARN": "your-lambda-function-arn",
    "powerValues": [128, 256, 512, 1024, 2048, 3008],
    "num": 100,
    "payload": {"data": "abc"},
    "parallelInvocation": true
}
```

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

AWS Lambda Power Tuning (input)

```
{
    "lambdaARN": "your-lambda-function-arn",
    "powerValues": [128, 256, 512, 1024, 2048, 3008],
    "num": 100,
    "payload": {"data": "abc"},
    "strategy": "speed|cost"
}
```

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

AWS Lambda Power Tuning (input)

```
{
    "lambdaARN": "your-lambda-function-arn",
    "powerValues": [128, 256, 512, 1024, 2048, 3008],
    "num": 100,
    "payload": {"data": "abc"},
    "strategy": "balanced",
    "balancedWeight": 0.5
}
```

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

AWS Lambda Power Tuning (input)

```
{
    "lambdaARN": "your-lambda-function-arn",
    "powerValues": [128, 256, 512, 1024, 2048, 3008],
    "num": 100,
    "payload": {"data": "abc"},
    "autoOptimize": true,
    "autoOptimizeAlias": "prod"
}
```

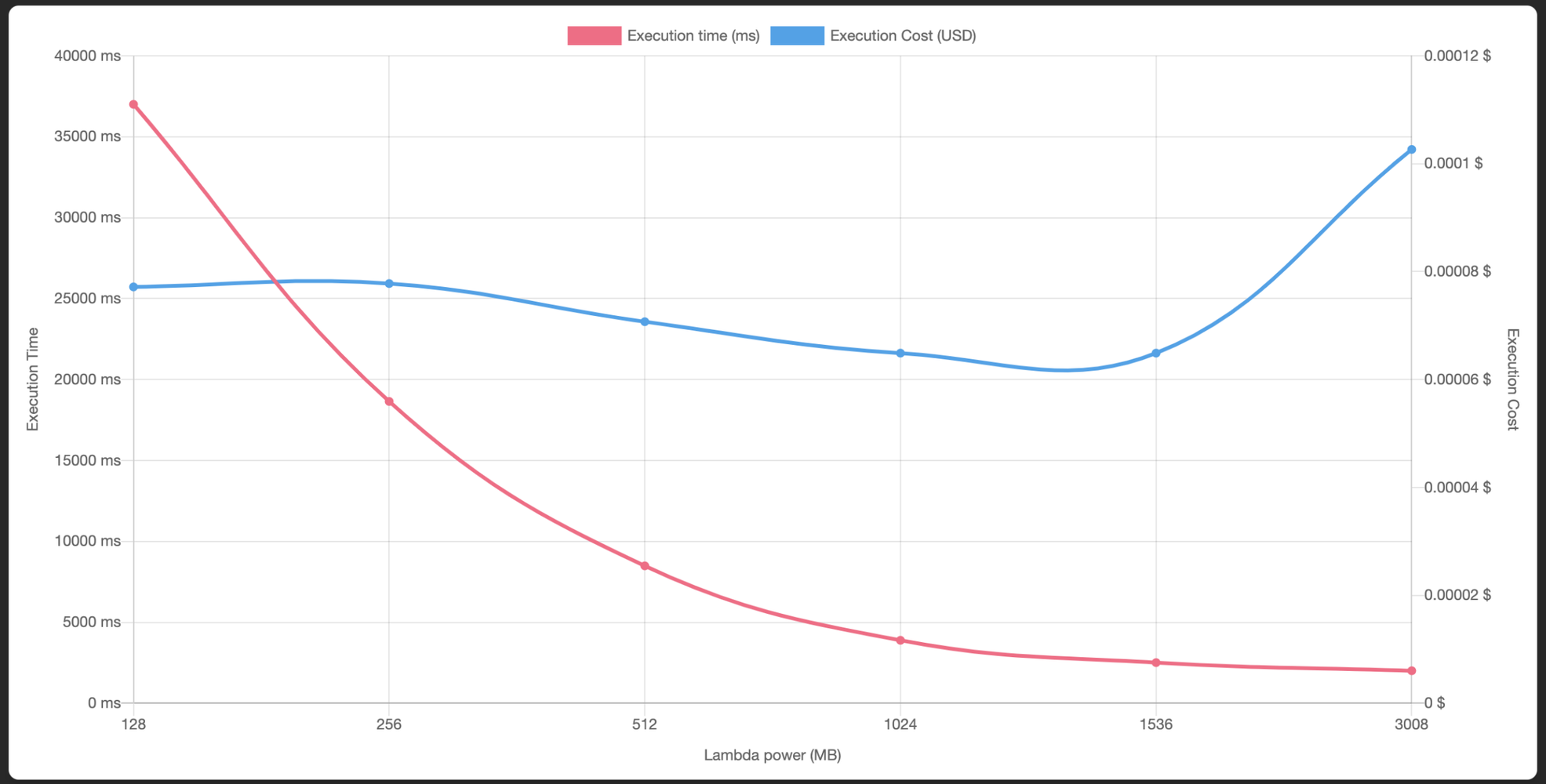
github.com/alexcasalboni/aws-lambda-power-tuning

AWS Lambda Power Tuning (output)

```
{
  "results": {
    "power": "128",
    "cost": 2.08e-7,
    "duration": 2.906,
    "stateMachine": {
      "executionCost": 0.00045,
      "lambdaCost": 0.0005252,
      "visualization": "https://lambda-power-tuning.show/ ... "
    }
  }
}
```

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

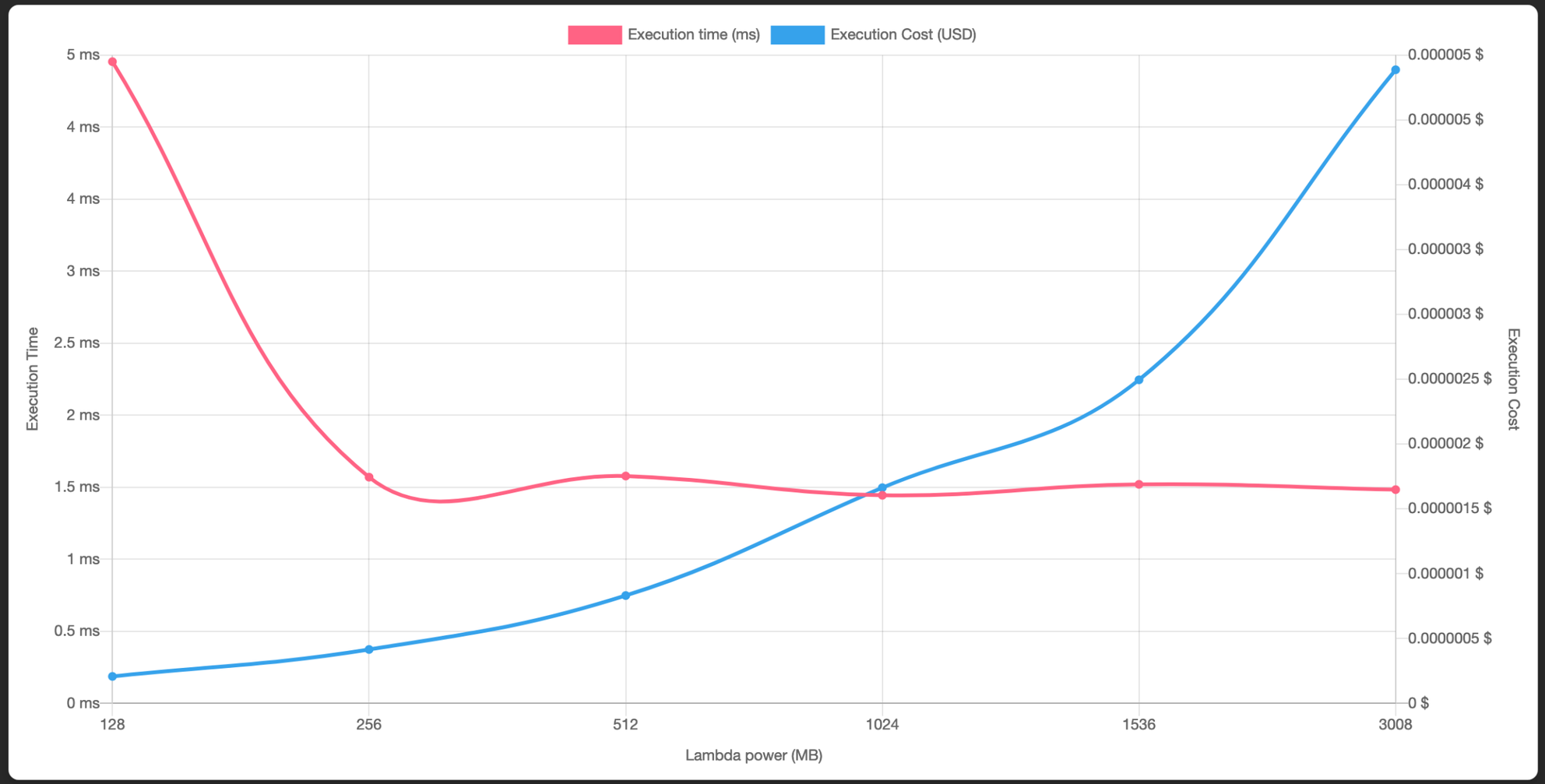
AWS Lambda Power Tuning (visualization)



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

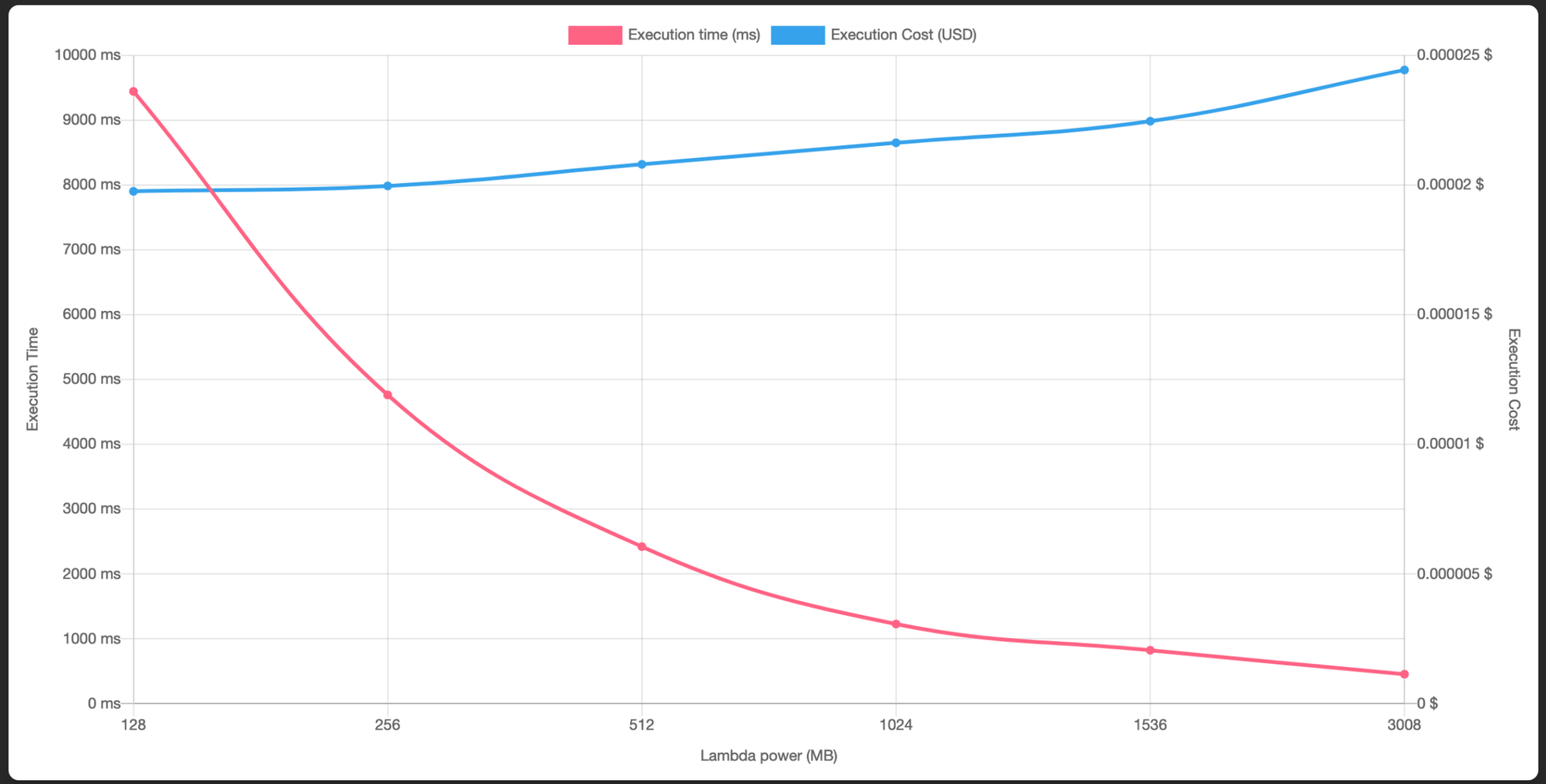
Real-world examples

No-Op (trivial data manipulation <100ms)



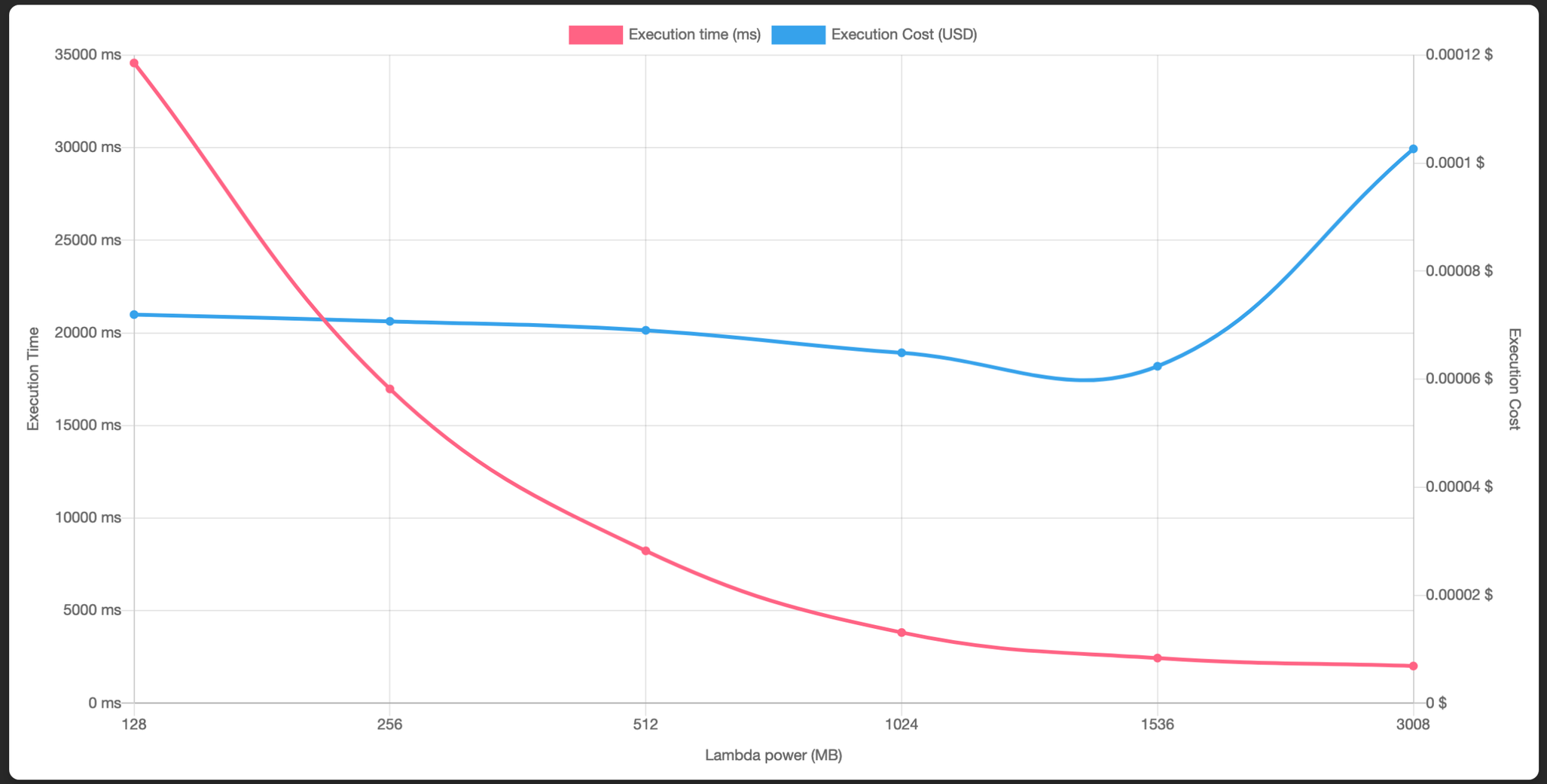
github.com/alexcasalboni/aws-lambda-power-tuning

CPU-bound (numpy: inverting 1500x1500 matrix)



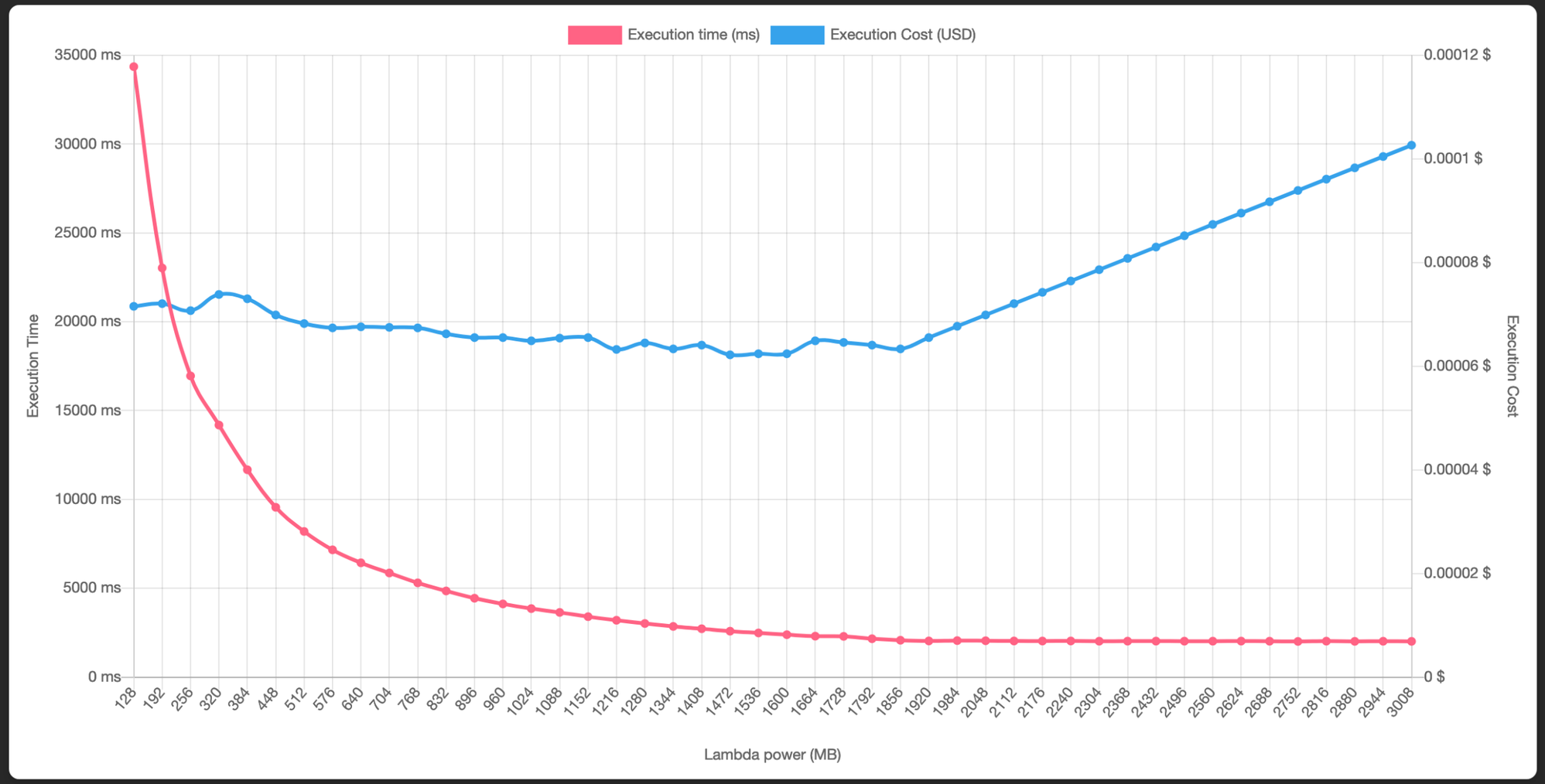
github.com/alexcasalboni/aws-lambda-power-tuning

CPU-bound (prime numbers)



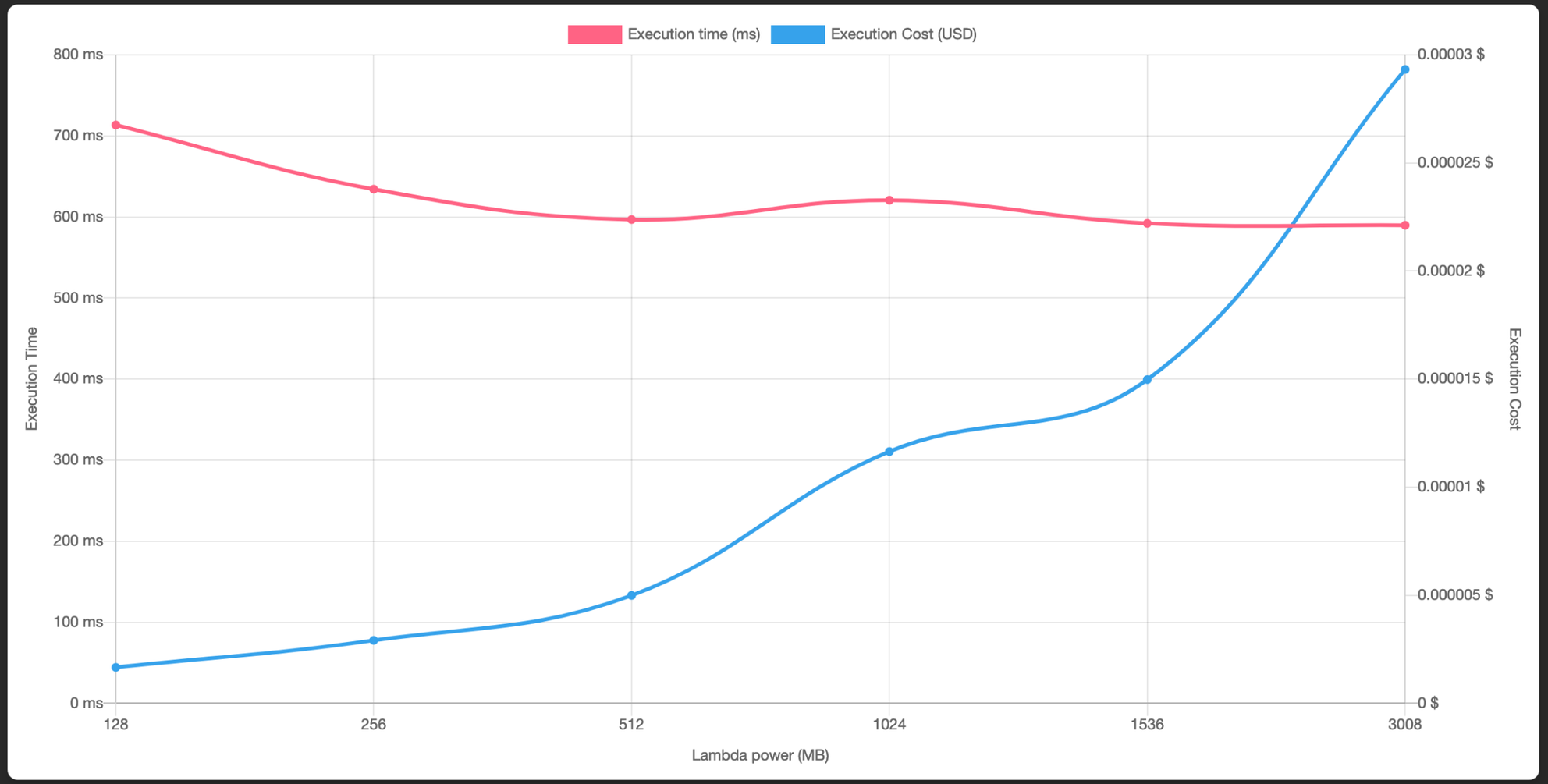
github.com/alexcasalboni/aws-lambda-power-tuning

CPU-bound (prime numbers – more granularity)



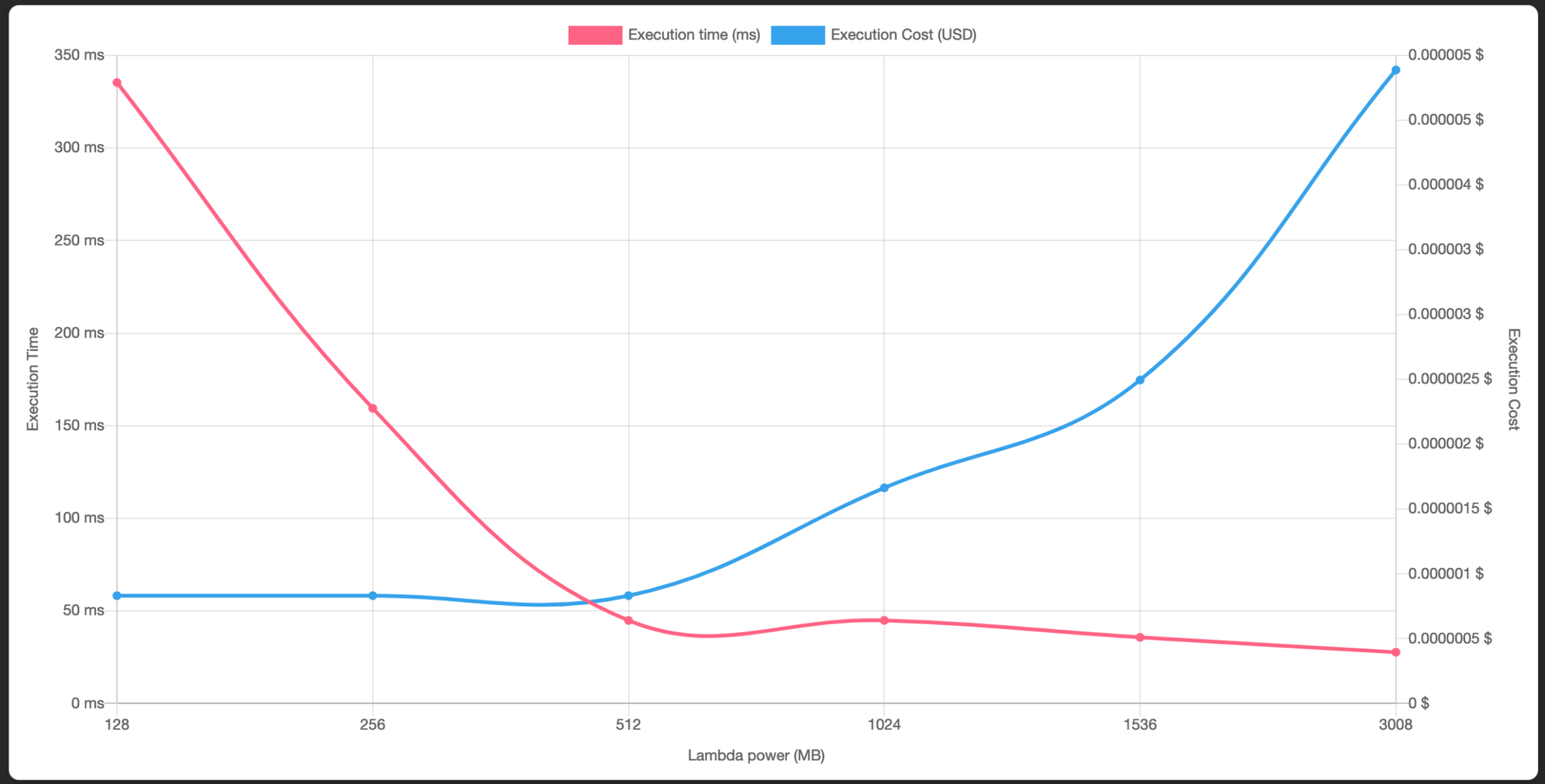
github.com/alexcasalboni/aws-lambda-power-tuning

Network-bound (third-party API call)



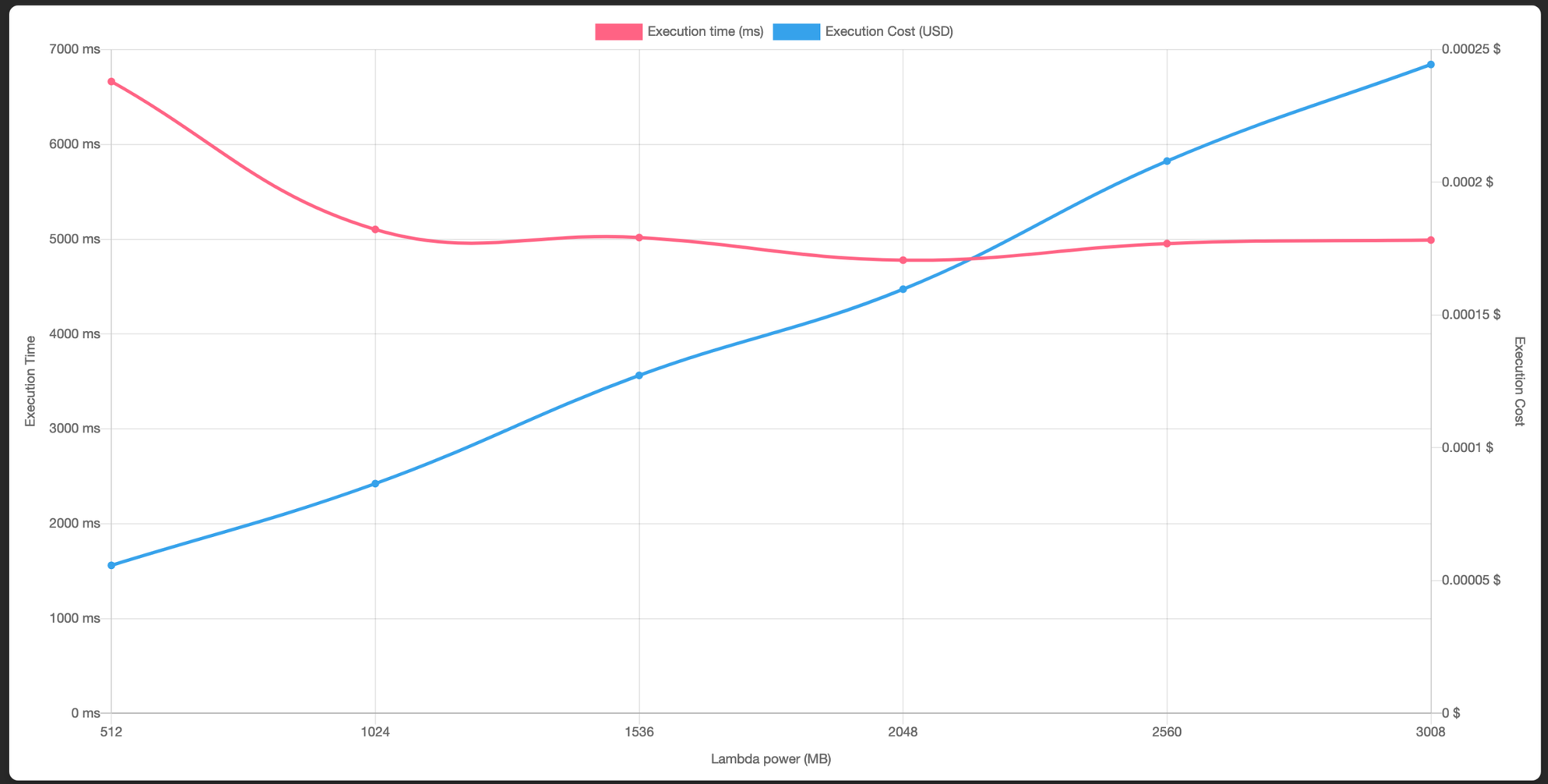
github.com/alexcasalboni/aws-lambda-power-tuning

Network-bound (3x DDB queries)



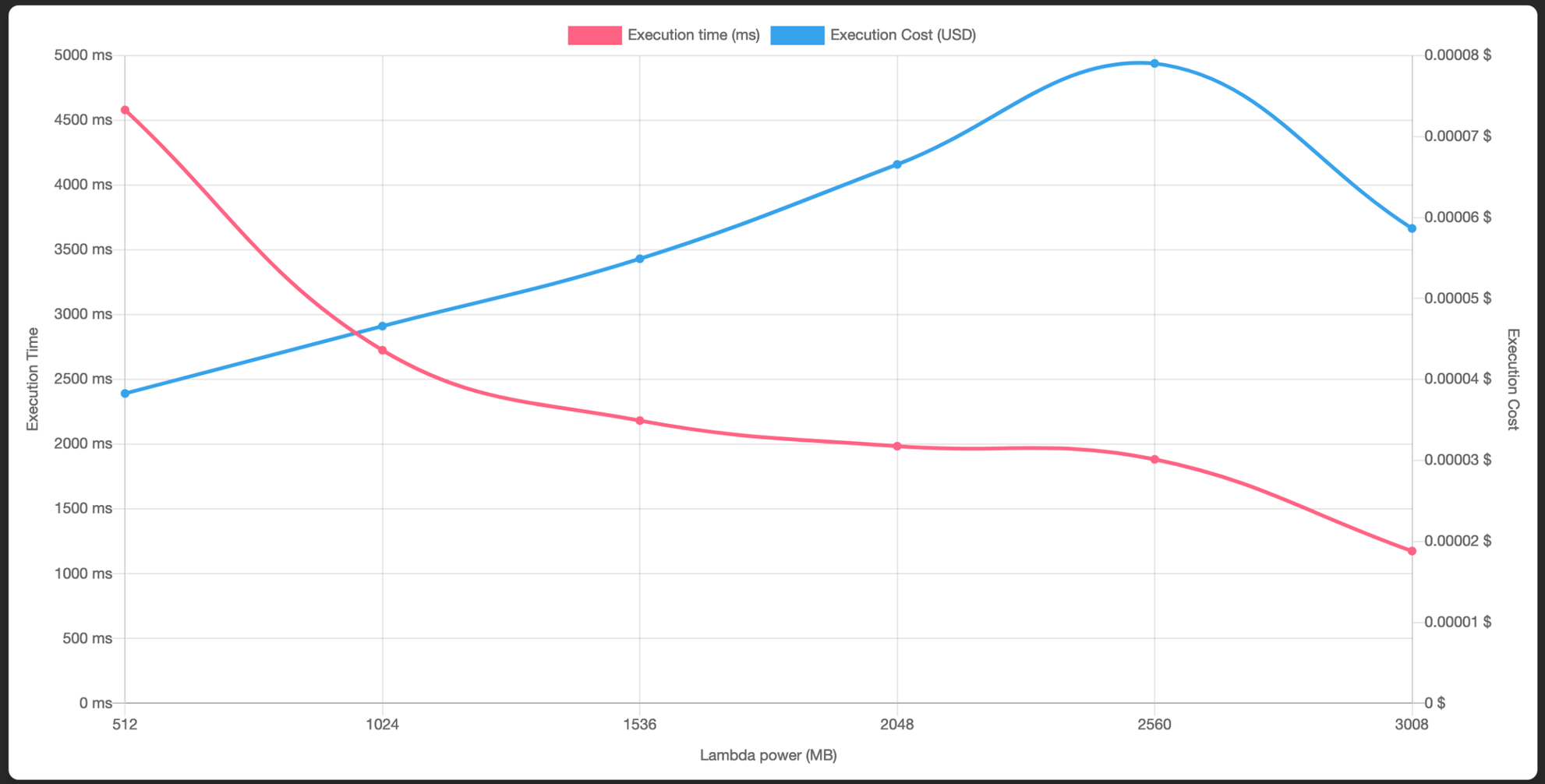
github.com/alexcasalboni/aws-lambda-power-tuning

Network-bound (S3 download – 150MB)



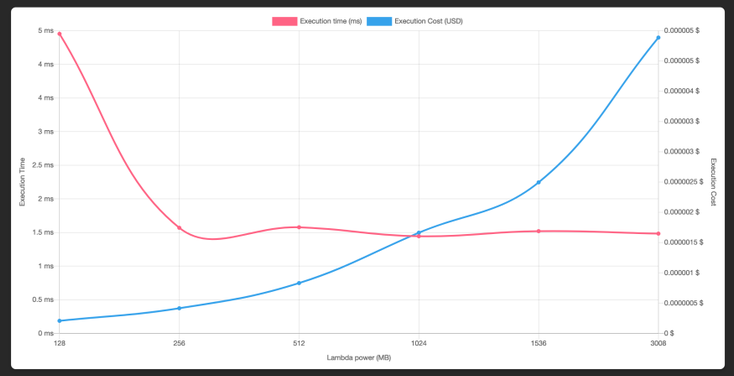
github.com/alexcasalboni/aws-lambda-power-tuning

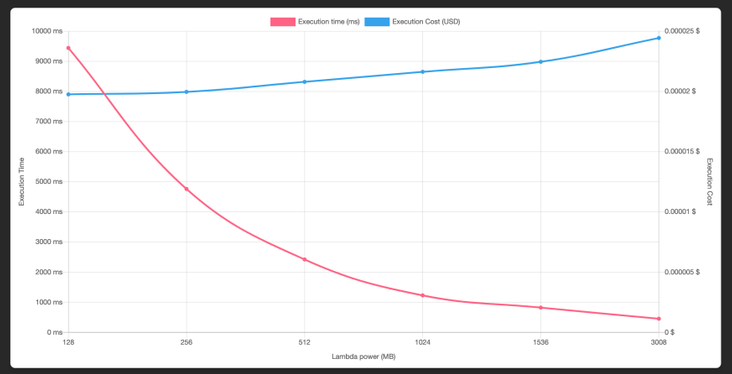
Network-bound (S3 download multithread – 150MB)

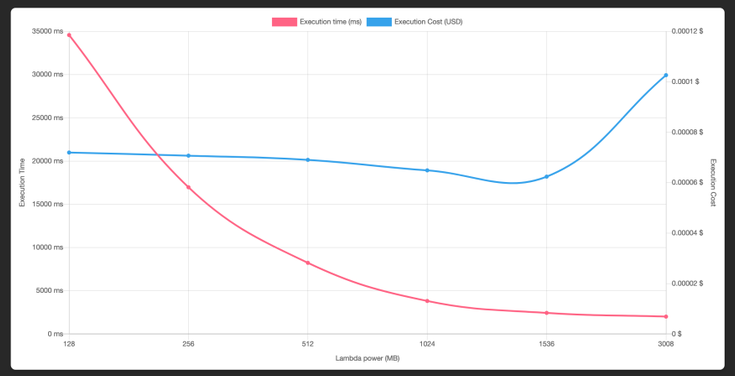


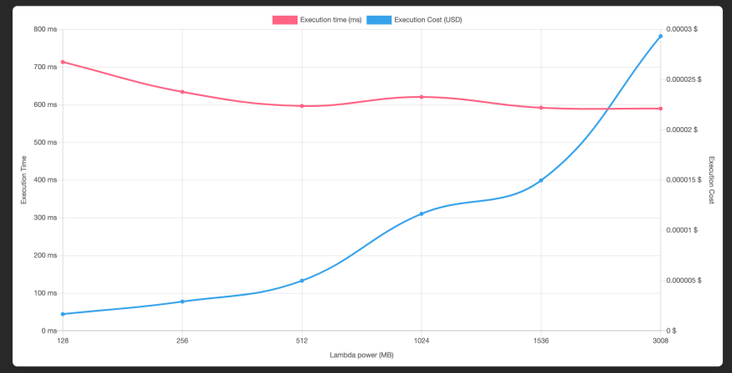
github.com/alexcasalboni/aws-lambda-power-tuning

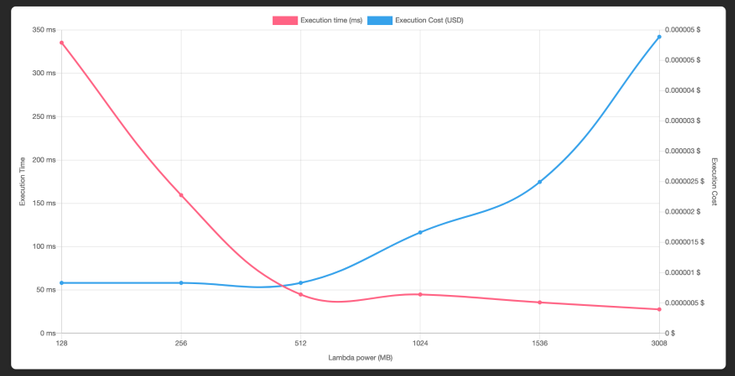
Cost/Performance patterns

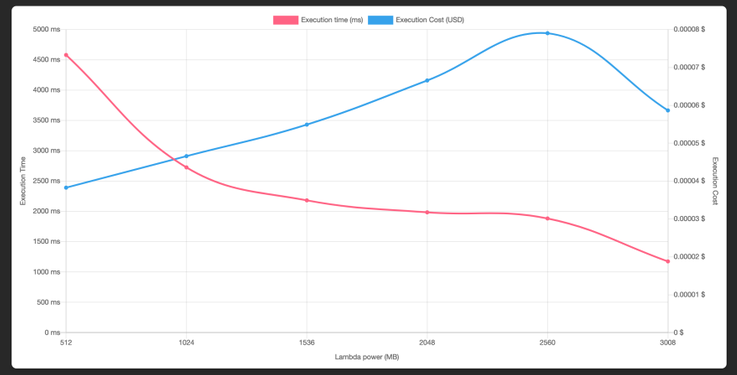












github.com/alexcasalboni/aws-lambda-power-tuning

Takeaways

Memory 👉 Power

Avoid cold starts with Provisioned Concurrency

Optimal resources allocation can be automated (CI/CD)

Think in terms of workload categories and cost/performance patterns

Visualize optimal trade-offs with /alexcasalboni/aws-lambda-power-tuning

Whiteboard discussion

Chalk talk repeats and related breakouts

SVS224-R1 - [REPEAT 1]
Dec 4, 10:45 a.m. –11:45 a.m. – Venetian, Lando 4206

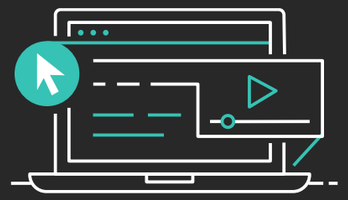
SVS224-R2 - [REPEAT 2]
Dec 5, 1:00 p.m. – 2:00 p.m. – Aria, Mariposa 1

SVS224-R3 - [REPEAT 3]
Dec 6, 10:00 a.m. – 11:00 a.m. – Venetian, Murano 3301B

CON213-L - Using containers and serverless to accelerate MAD
Dec 4, 9:15 a.m. – 10:15 a.m. – Venetian, Venetian Theatre

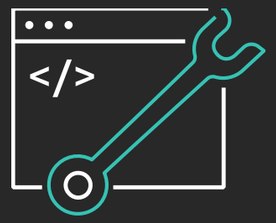
Learn serverless with AWS Training and Certification

Resources created by the experts at AWS to help you learn modern application development



Free, on-demand courses on serverless, including

- Introduction to Serverless Development
- Getting into the Serverless Mindset
- AWS Lambda Foundations
- Amazon API Gateway for Serverless Applications
- Amazon DynamoDB for Serverless Architectures



Additional digital and classroom trainings cover modern application development and computing

Visit the Learning Library at https://aws.training

aws training and certification

Thank you!

Alex Casalboni

acasal@amazon.com
@alex_casalboni

AWS re:Invent

aws

Please complete the session survey in the mobile app.

AWS re:Invent

aws