

# Reducing Infrastructure Cost During Development and in Production

Sally Wahba  
@sallyky

# Annual Planning 🏠

- Do more with less
- Bring the cost down
- Reduce the cloud bill
- \$\$\$\$



<https://www.pexels.com/photo/people-discuss-about-graphs-and-rates-3184292/>  
<https://a.slack-edge.com/production-standard-emoji-assets/14.0/apple-large/1f4b8.png>

# Story Time - Your Dev Environment

---




 Engineers want their own environment

 Engineers go home for the day/weekend/vacation

 Cloud bill goes 

# What did we do?

---

-  Reservation system
-  Auto deletion of cloud resources
-  Regular individualized emails

# Other ideas

---



## Using cheaper resources

- Spot instances vs on-demand
- Slower storage
- Less replication
- Limit instance types

Instance name ▲	On-Demand hourly rate ▼	vCPU ▼	Memory ▼	Storage ▼	Network performance ▼
t4g.2xlarge	\$0.2688	8	32 GiB	EBS Only	Up to 5 Gigabit
t3.2xlarge	\$0.3328	8	32 GiB	EBS Only	Up to 5 Gigabit
t3a.2xlarge	\$0.3008	8	32 GiB	EBS Only	Up to 5 Gigabit
t2.2xlarge	\$0.3712	8	32 GiB	EBS Only	Moderate
m7a.2xlarge	\$0.46368	8	32 GiB	EBS Only	Up to 12500 Megabit
m7g.2xlarge	\$0.3264	8	32 GiB	EBS Only	Up to 15 Gigabit
m7gd.2xlarge	\$0.4271	8	32 GiB	1 x 475 NVMe SSD	Up to 15 Gigabit
m7i.2xlarge	\$0.4032	8	32 GiB	EBS Only	Up to 12500 Megabit

<https://aws.amazon.com/ec2/pricing/on-demand/>

# Other ideas

---



## Using cheaper resources

- Spot instances vs on-demand
- Slower storage
- Less replication
- Limit instance types



## How good are your tests?

- Do you need to be using cloud resources for each engineer?

# Story time - Your Prod Environment

---



You're a success story



More customers



Adding more features/products



Scaling your infra



Cloud bill goes 🚀



# What do we do?

---



Autoscaling



AMD vs ARM

Instance name	▲	On-Demand hourly rate
t4g.2xlarge		\$0.2688
t3.2xlarge		\$0.3328
t3a.2xlarge		\$0.3008

# What do we do?





---

- 💰 Autoscaling
- 💰 AMD vs ARM
- 💰 Batch processing
- 💰 The big rearchitecture

Instance name ▲	On-Demand hourly rate
t4g.2xlarge	\$0.2688
t3.2xlarge	\$0.3328
t3a.2xlarge	\$0.3008

# Takeaways

---

-  Tag your resources
-  Transparency matters
-  An effective solution may not be technical
-  Start with the low hanging fruit first

**Thank you!**

---