aws
SUMMIT
LONDON | APRIL 27, 2022

AR-04

Scaling up to your first 10 million users

Dr Mike Rizzo
Senior Solutions Architect
Amazon Web Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Meet Ada, CEO Unicorn Rentals

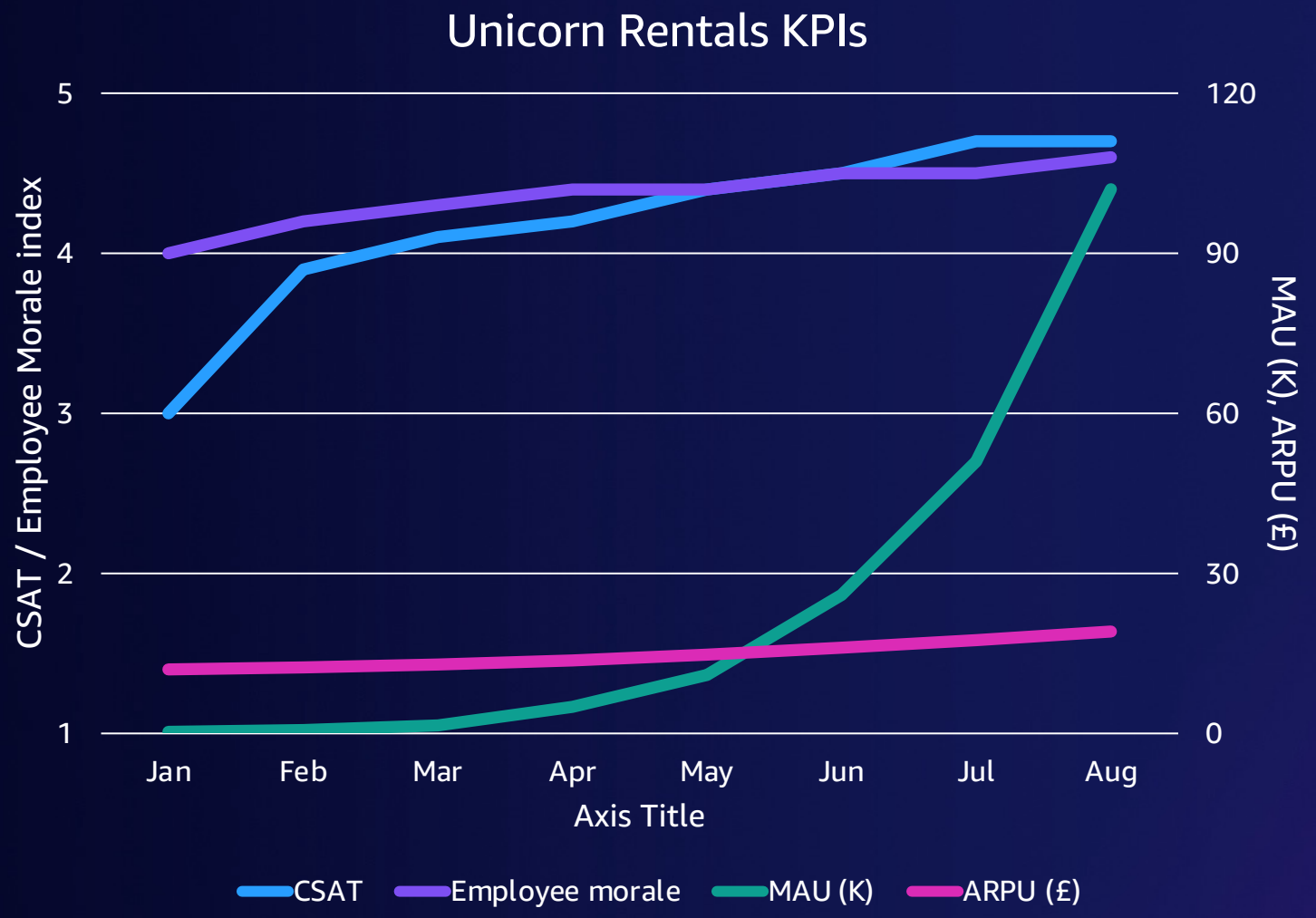
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Clean transport for a greener planet!

aws

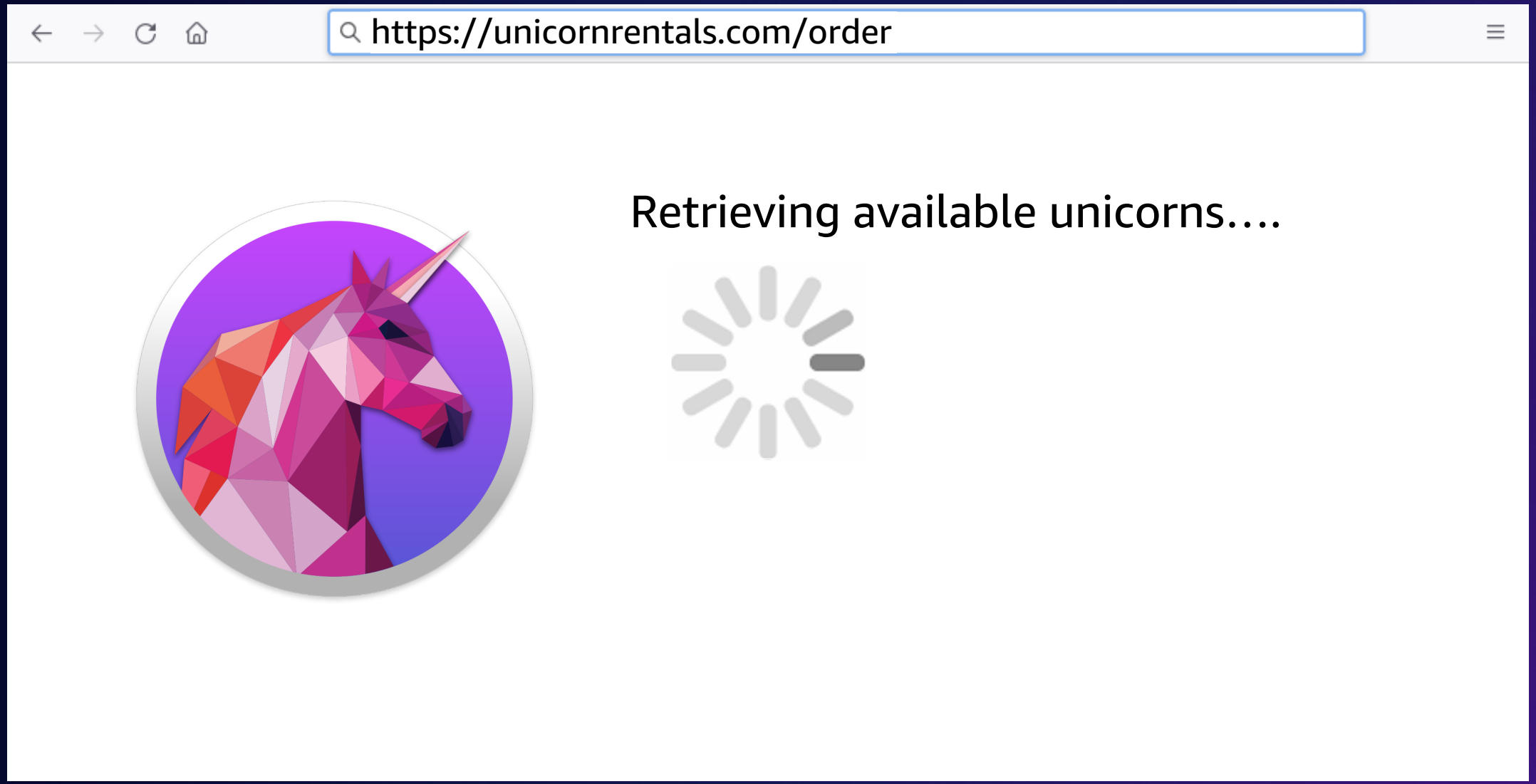
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Things were looking great ...



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

… until this started to happen,



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

… and then this,

https://unicornrentals.com/order

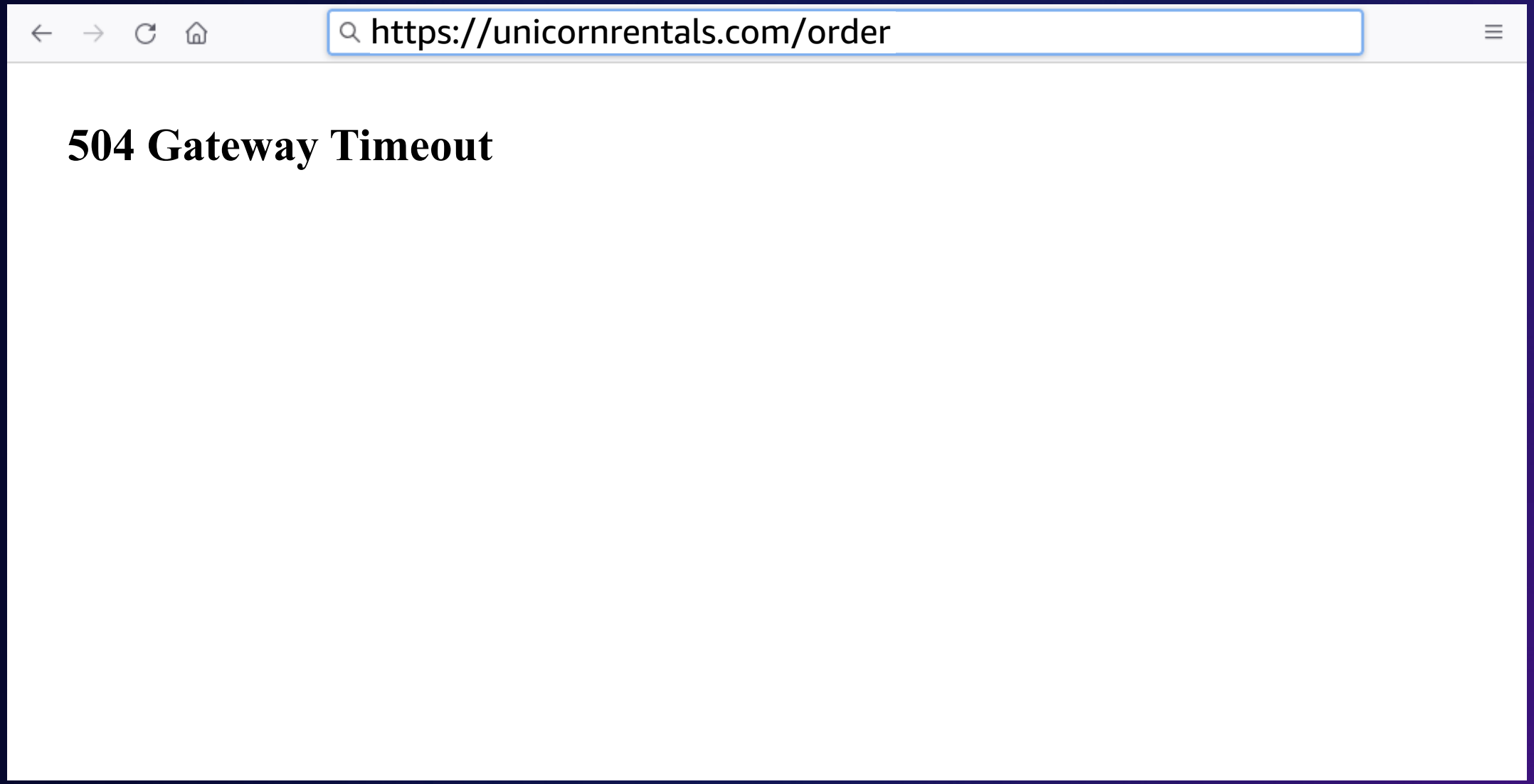
We are really busy right now!

Sorry about that!
Please try again later.

PS You could try walking…

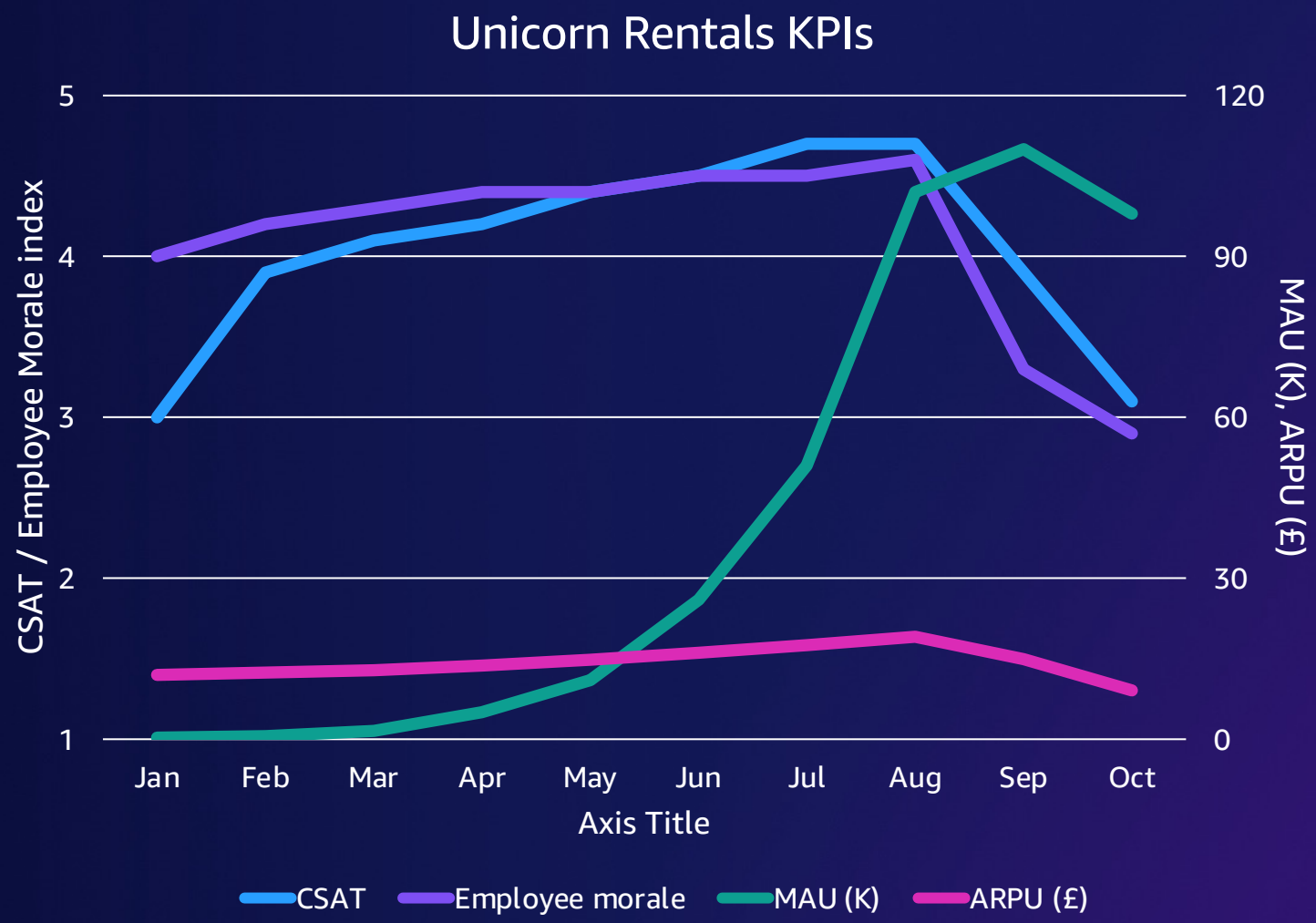
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

… and then things got really bad!



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Things took a turn for the worse !



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Scalability matters !

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How do we fix it?

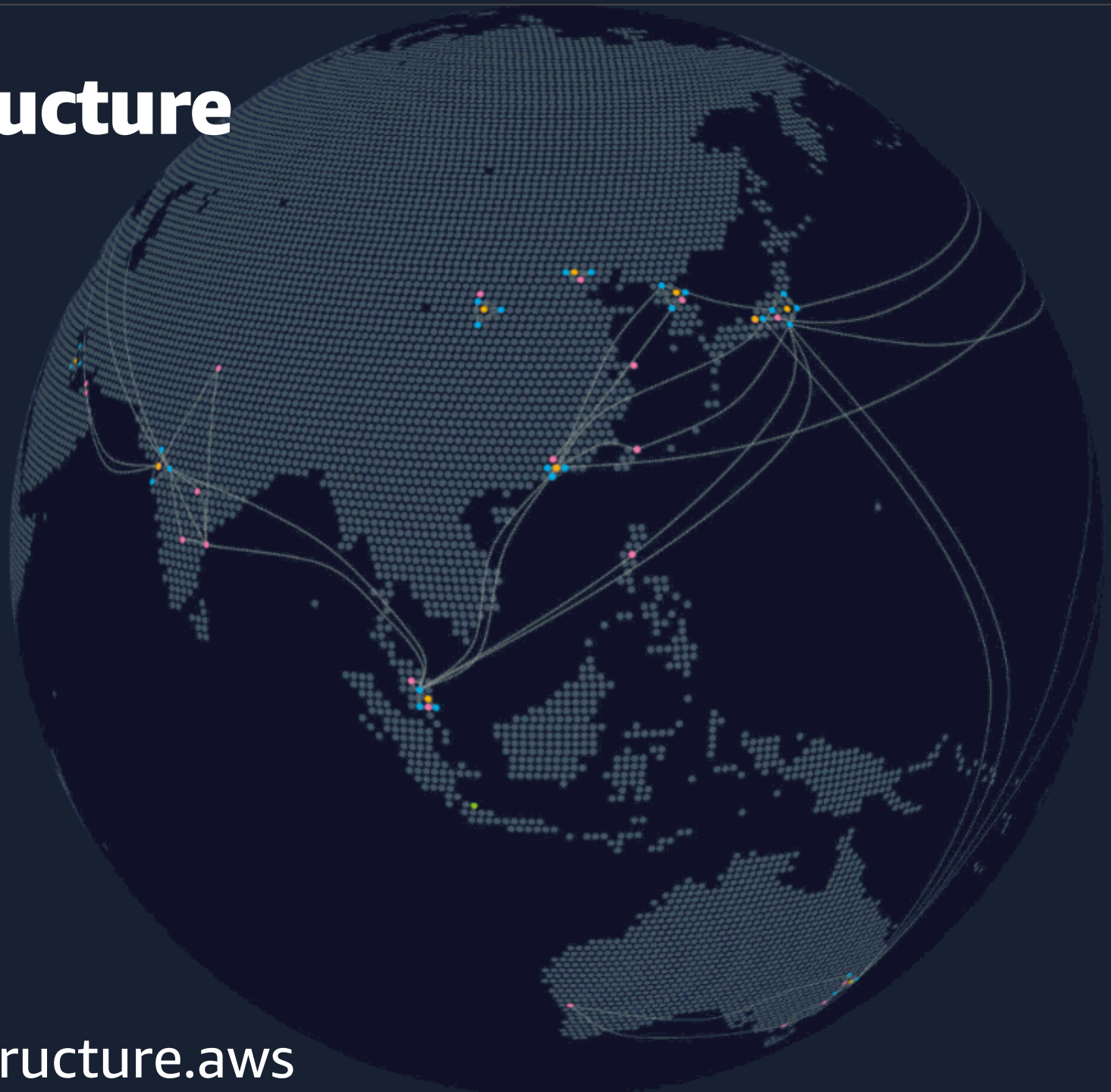
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Global Infrastructure

26 geographic Regions

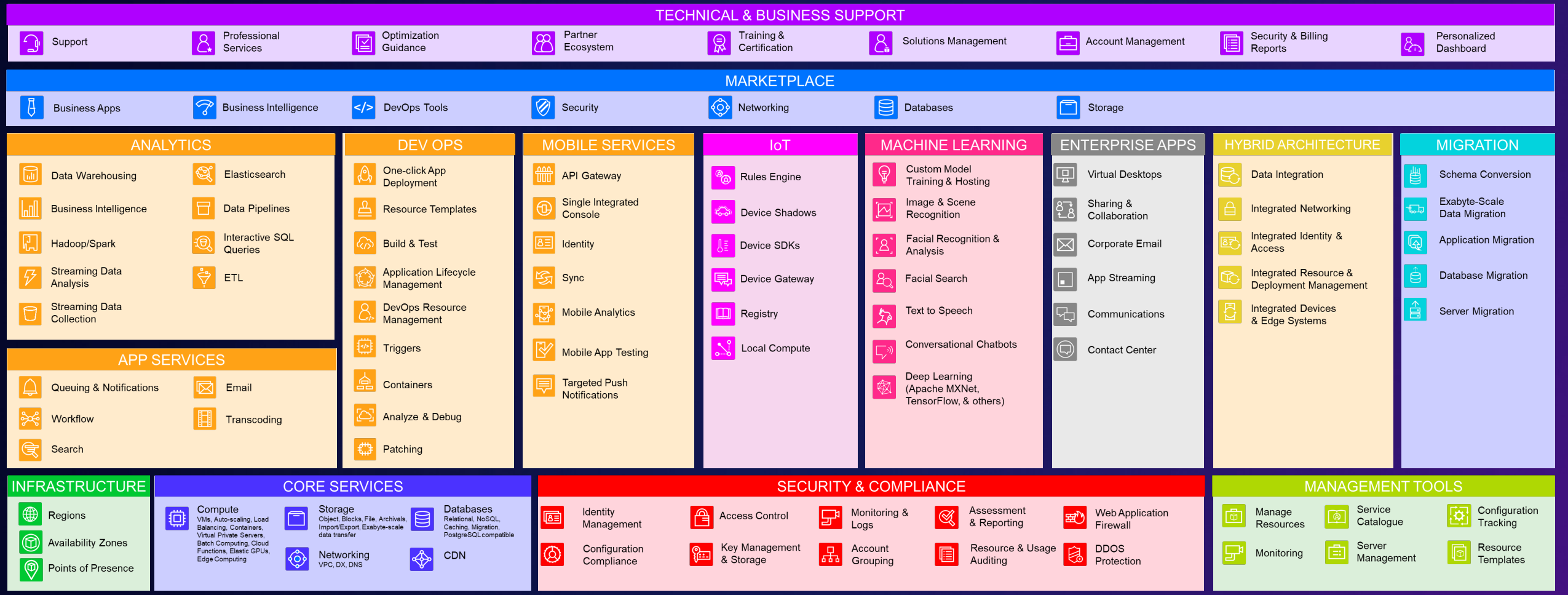
84 Availability Zones

310+ points of presence



https://www.infrastructure.aws

Breadth and depth of service platform



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Self-managed vs AWS-managed services

AWS-managed

Self-managed / DIY

Low abstraction / Servers visible

High abstraction / Serverless

Amazon EC2

Amazon EBS

Amazon RDS

Amazon EMR

Amazon SQS

Amazon EFS

More control
More responsibility
Need to build scalability and reliability

Less control
Less responsibility
Built-in scalability and reliability
But watch out for service limits!

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“Many decisions are reversible, two-way doors.”

Jeff Bezos

Founder and Executive Chair of Amazon

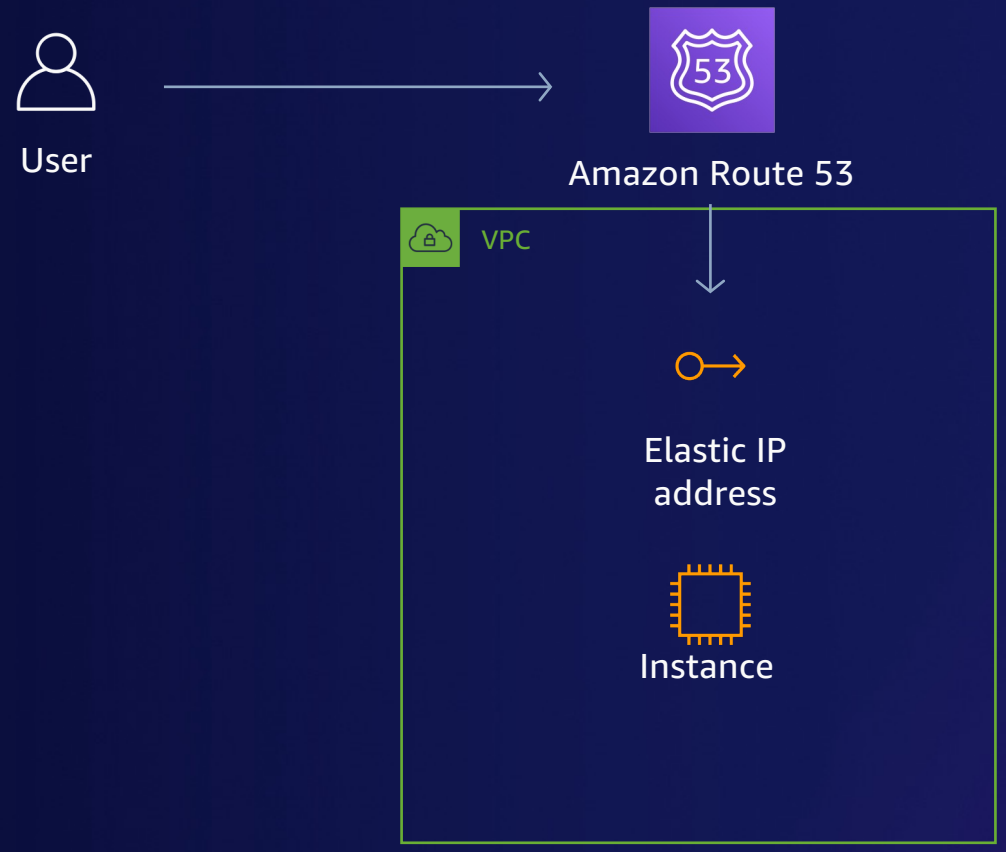
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Unicorn Rentals: the story so far ...

Day 1

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

This was only intended to support a PoC …



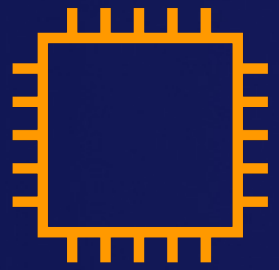
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“We’re gonna need a bigger box”

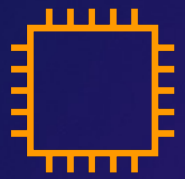
Simplest approach: vertical scaling

- Instance types:
  - General purpose
  - Memory optimized
  - Compute optimized
  - Storage optimized
  - Accelerated compute
- Easy to change instance sizes

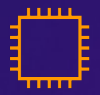
- Disk types
  - Magnetic vs SSD
  - Provisioned IOPS



c5.9xlarge



m5.2xlarge



t3.nano

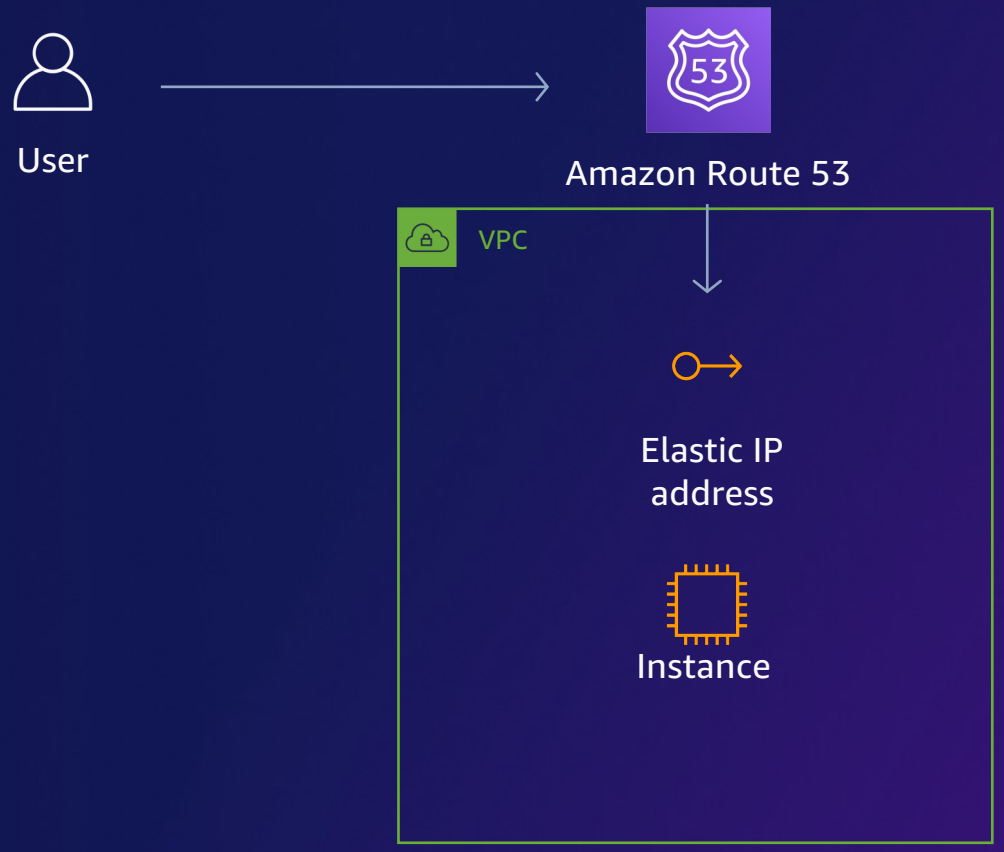
Will eventually hit a maximum limit!

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

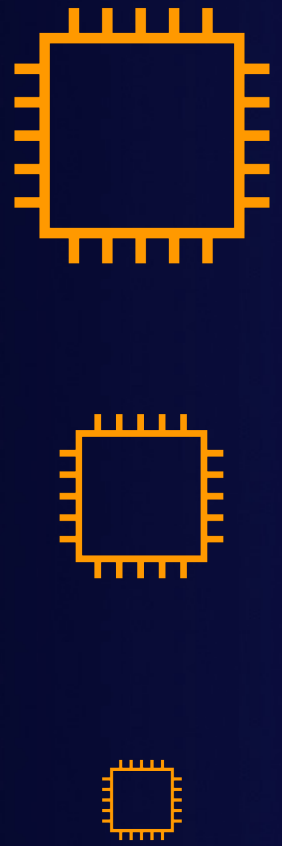
Considerations for real-world use

- No failover
- No redundancy
- Can’t scale individual components independently
- Can’t make technology choices for individual components



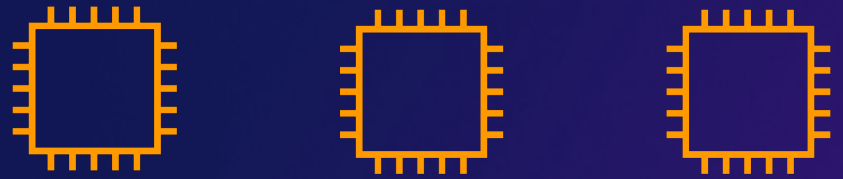
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Scaling options



Vertically

Horizontally



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step 1: Separate the database

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Aurora

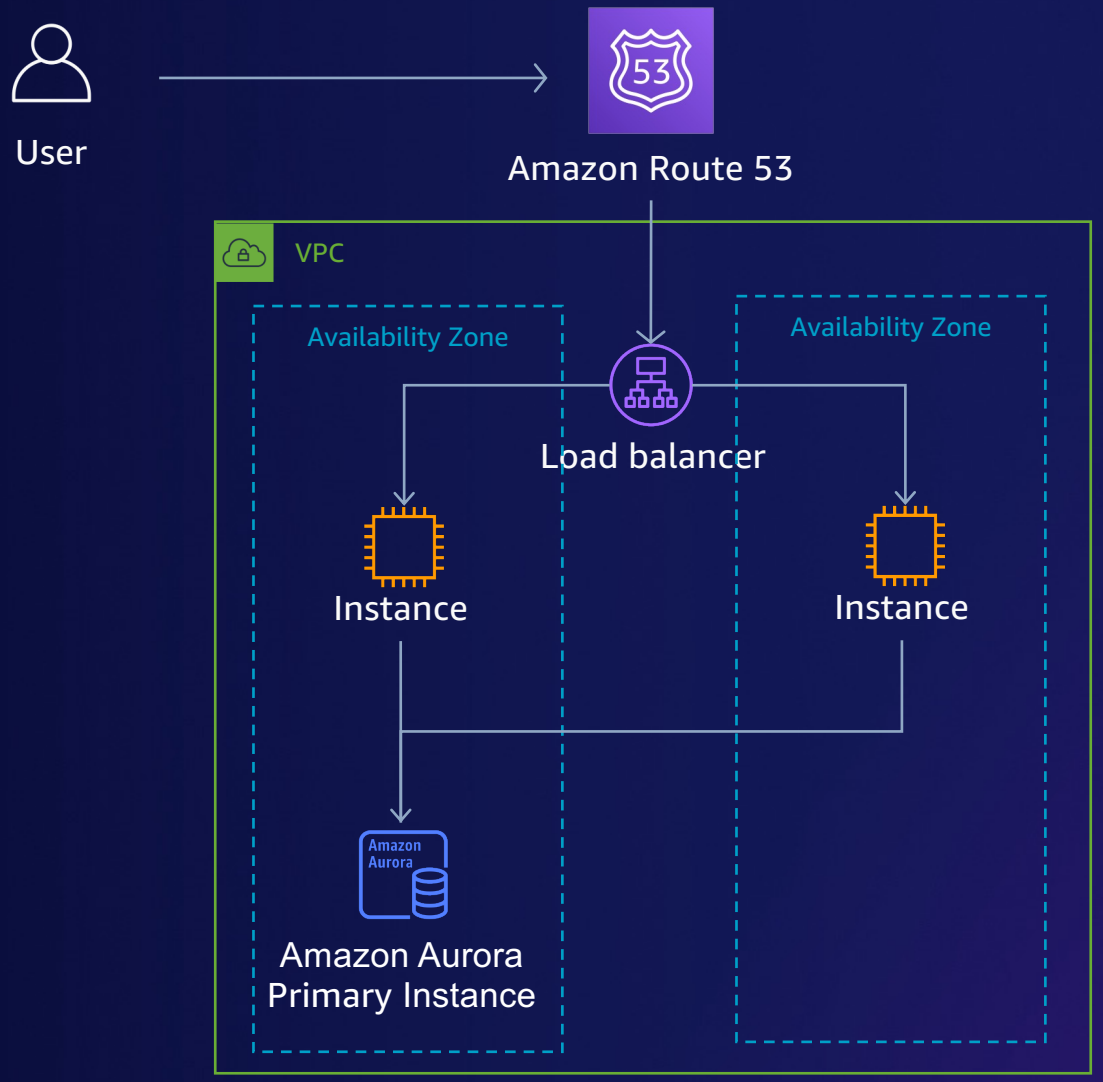
Amazon Aurora

- MySQL or PostgreSQL compatible
- Automatic storage scaling (up to 64 TB)
- Up to 15 read replicas
- Continuous (incremental) backups
- Six-way replication across three zones

aws

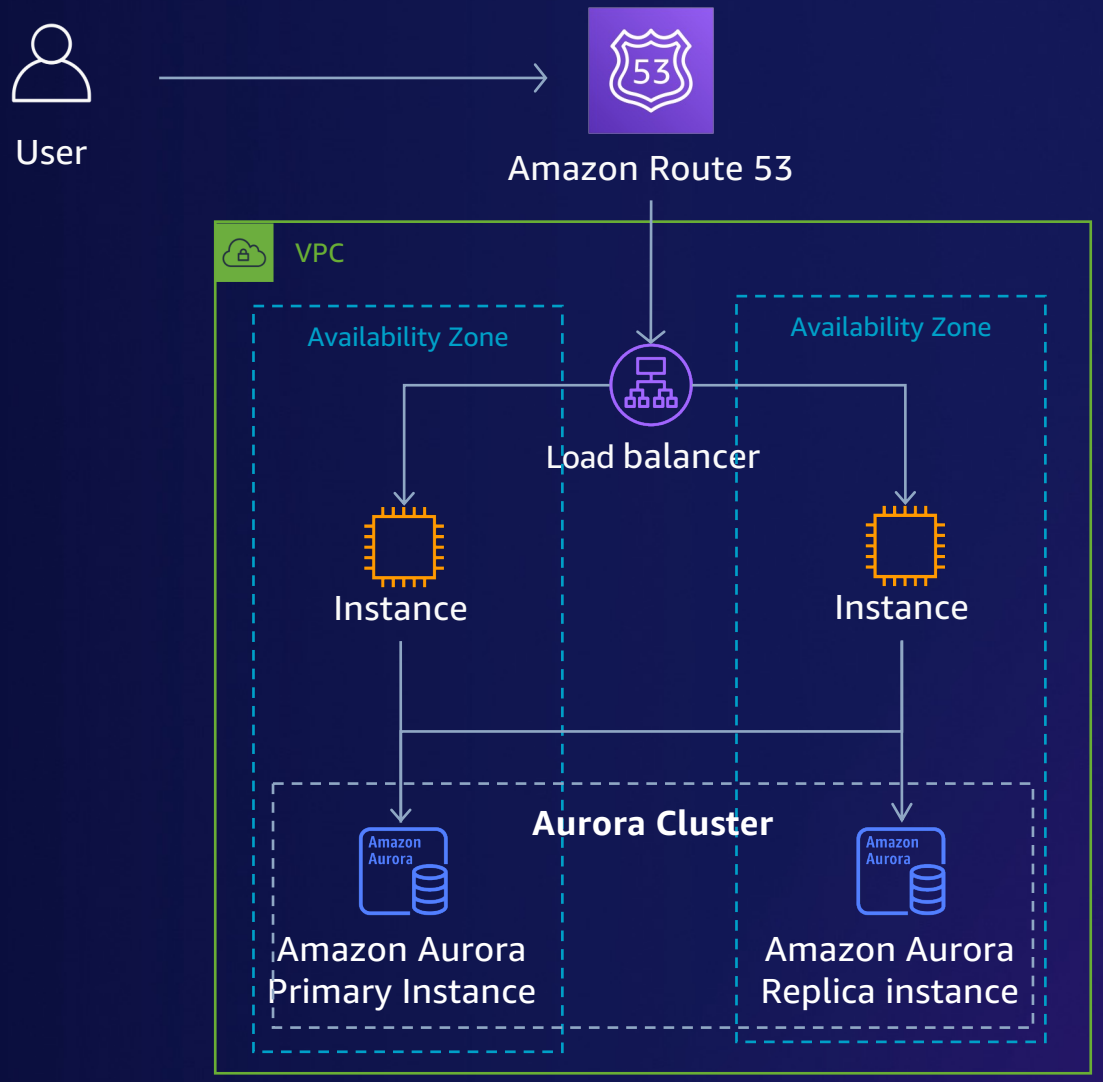
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step 2: Horizontally scale the application



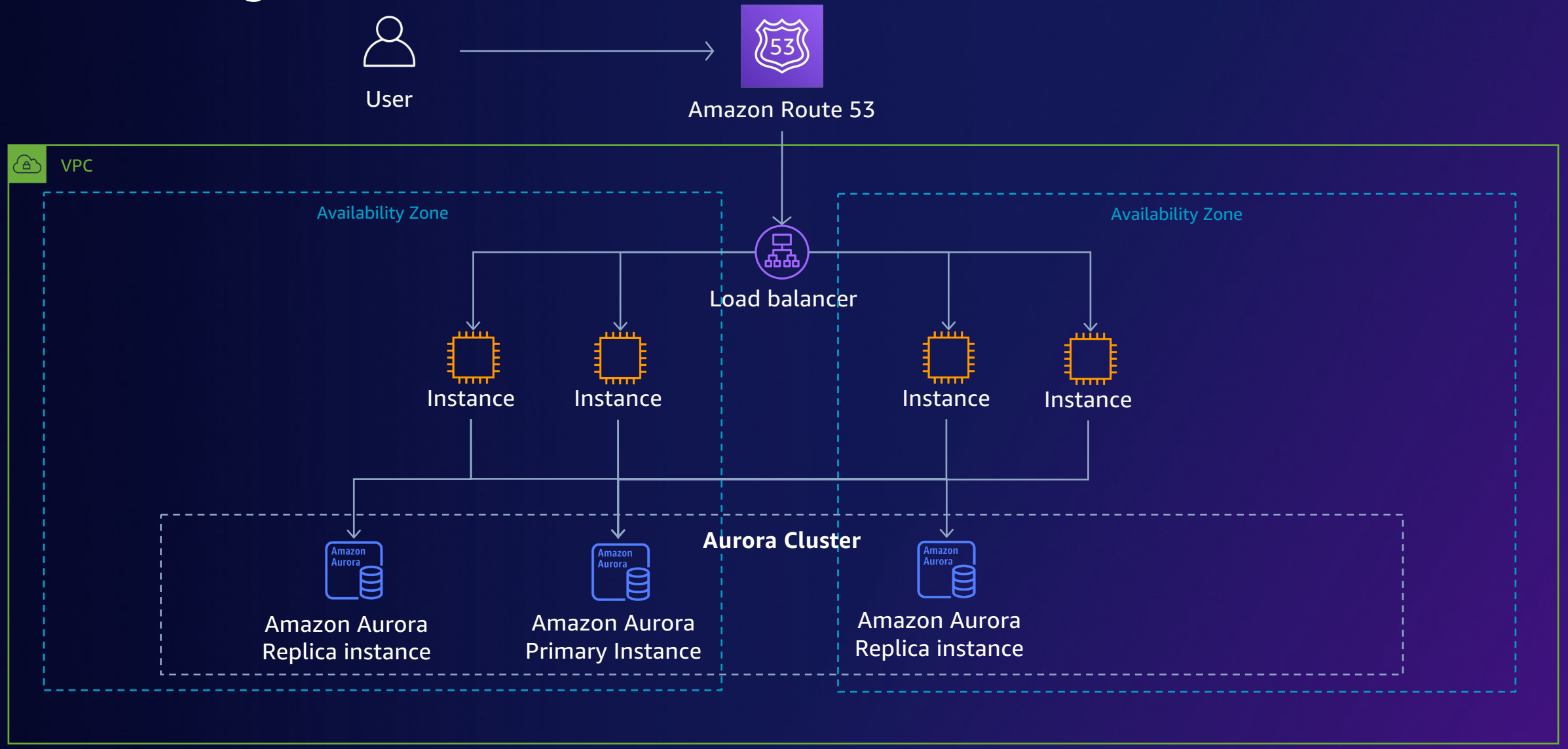
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Step 3: Add an Aurora DB replica



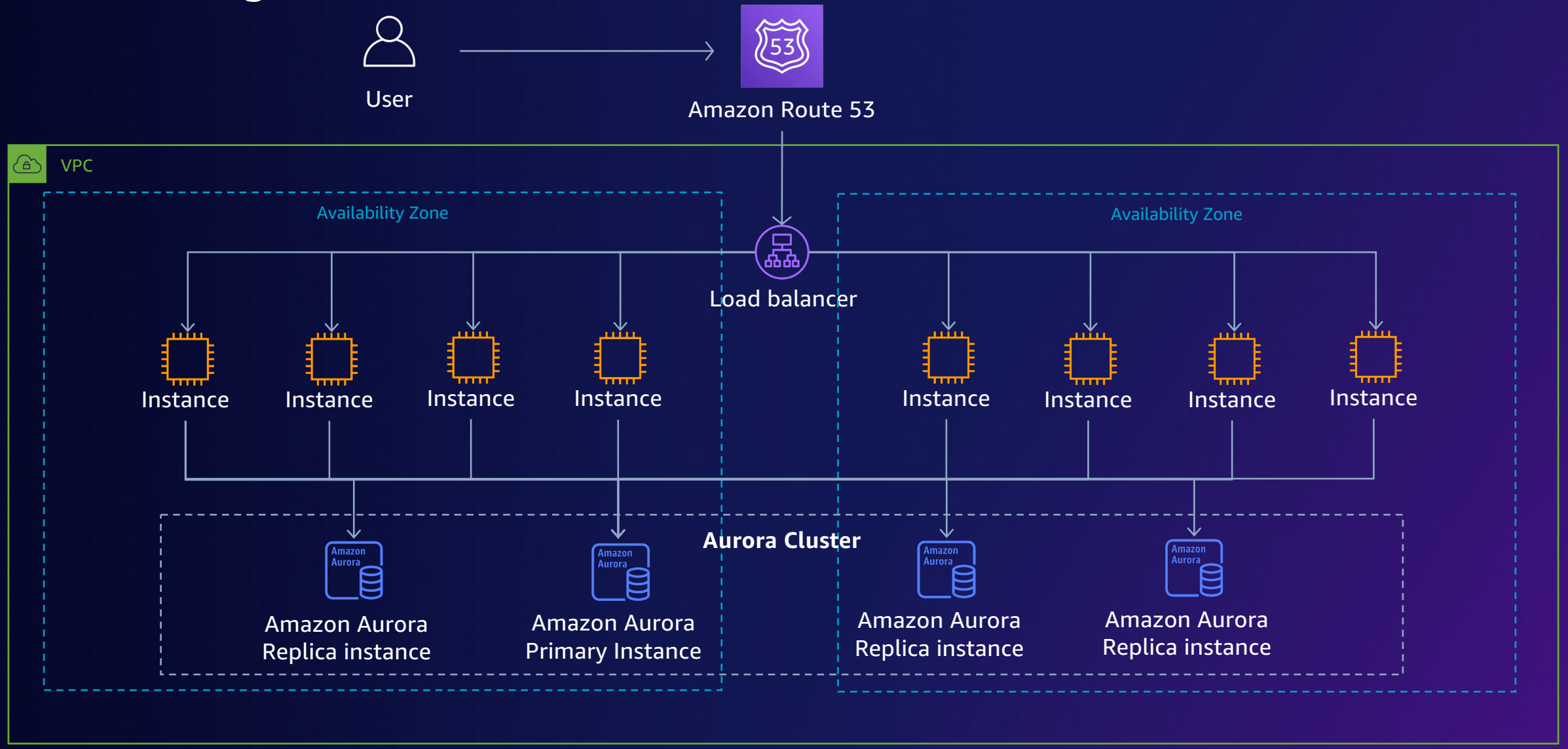
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monthly Active Users < 80,000



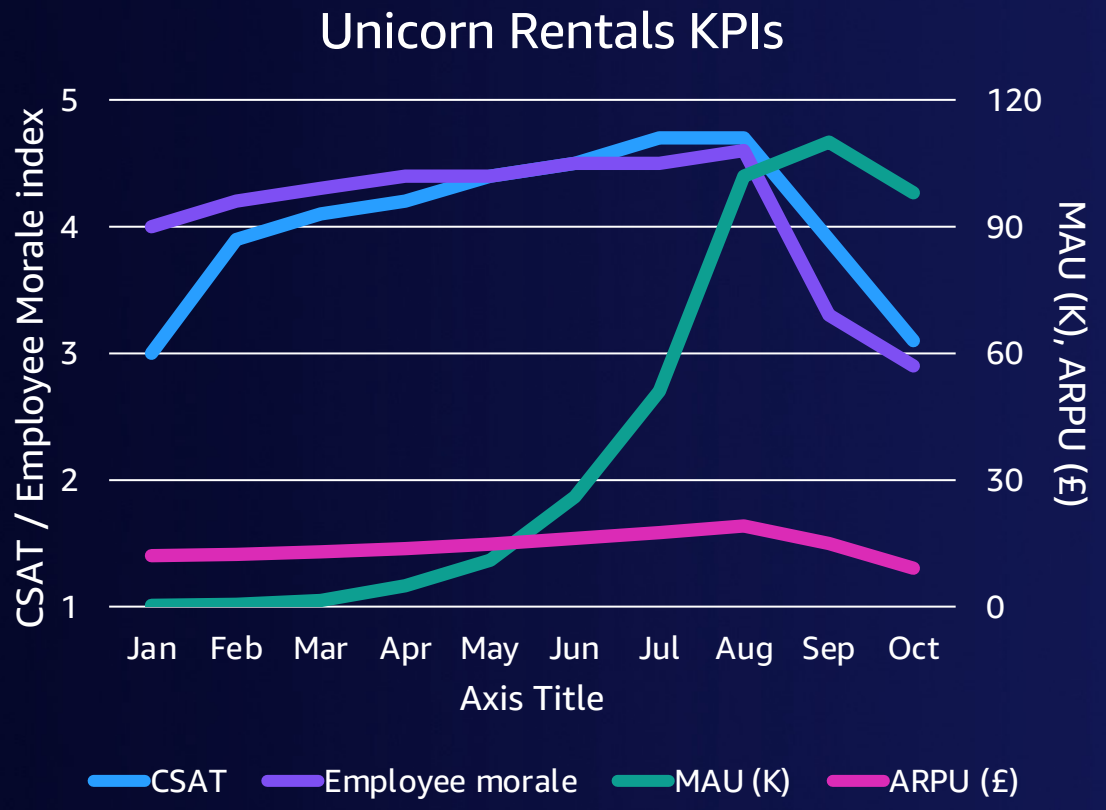
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monthly Active Users < 150,000



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Mounting problems for the CEO !



- KPIs heading south
- Tensions between developers and operations staff
- Security team want MFA
- Support desk want lost password reset
- Investor asking why cost per user not going down
- Press asking about carbon footprint of IT systems

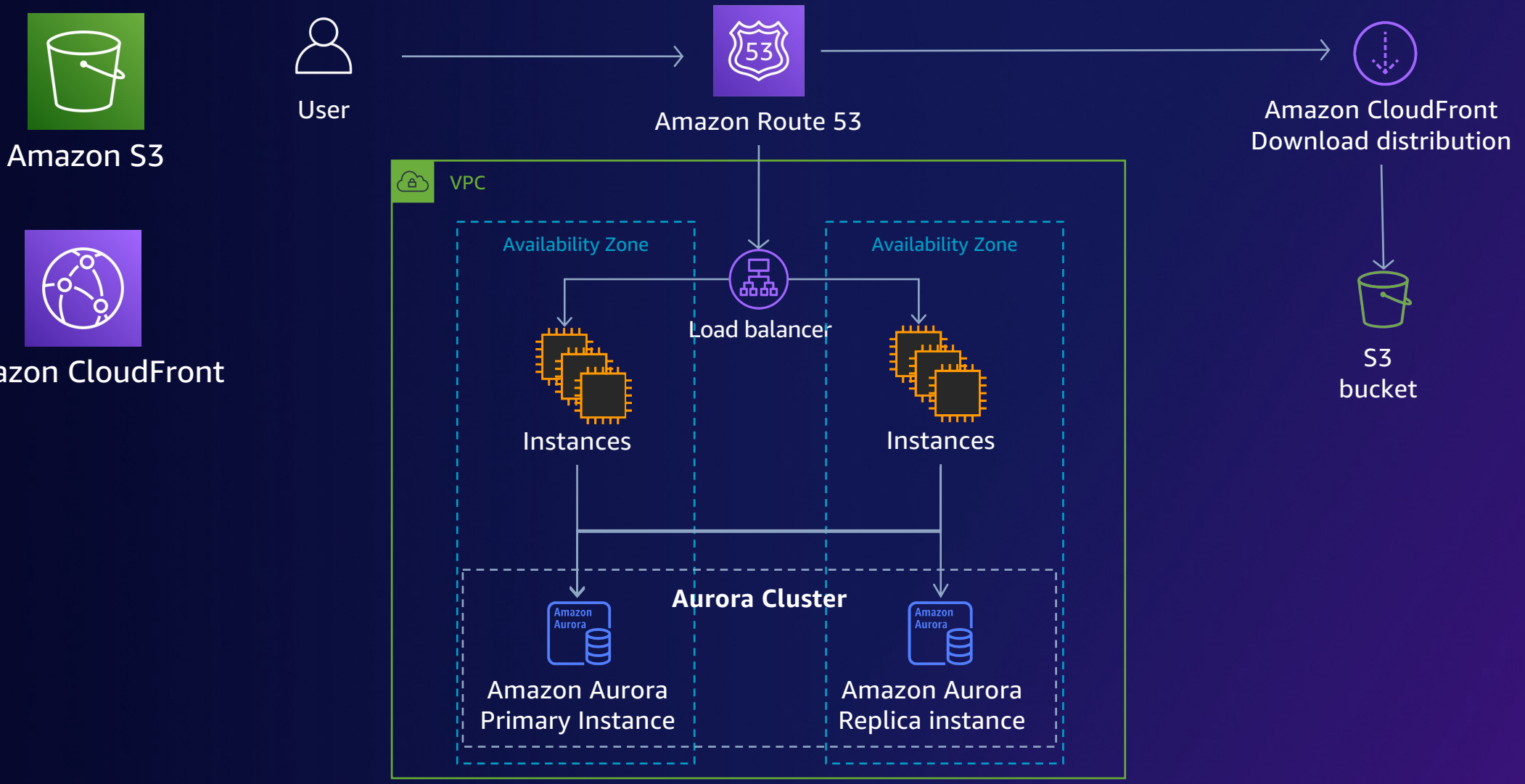
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CEO challenge

How can Unicorn Rentals maintain a great customer experience as the business grows exponentially, while also meeting targets related to employee well-being, cost-efficiency and sustainability?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

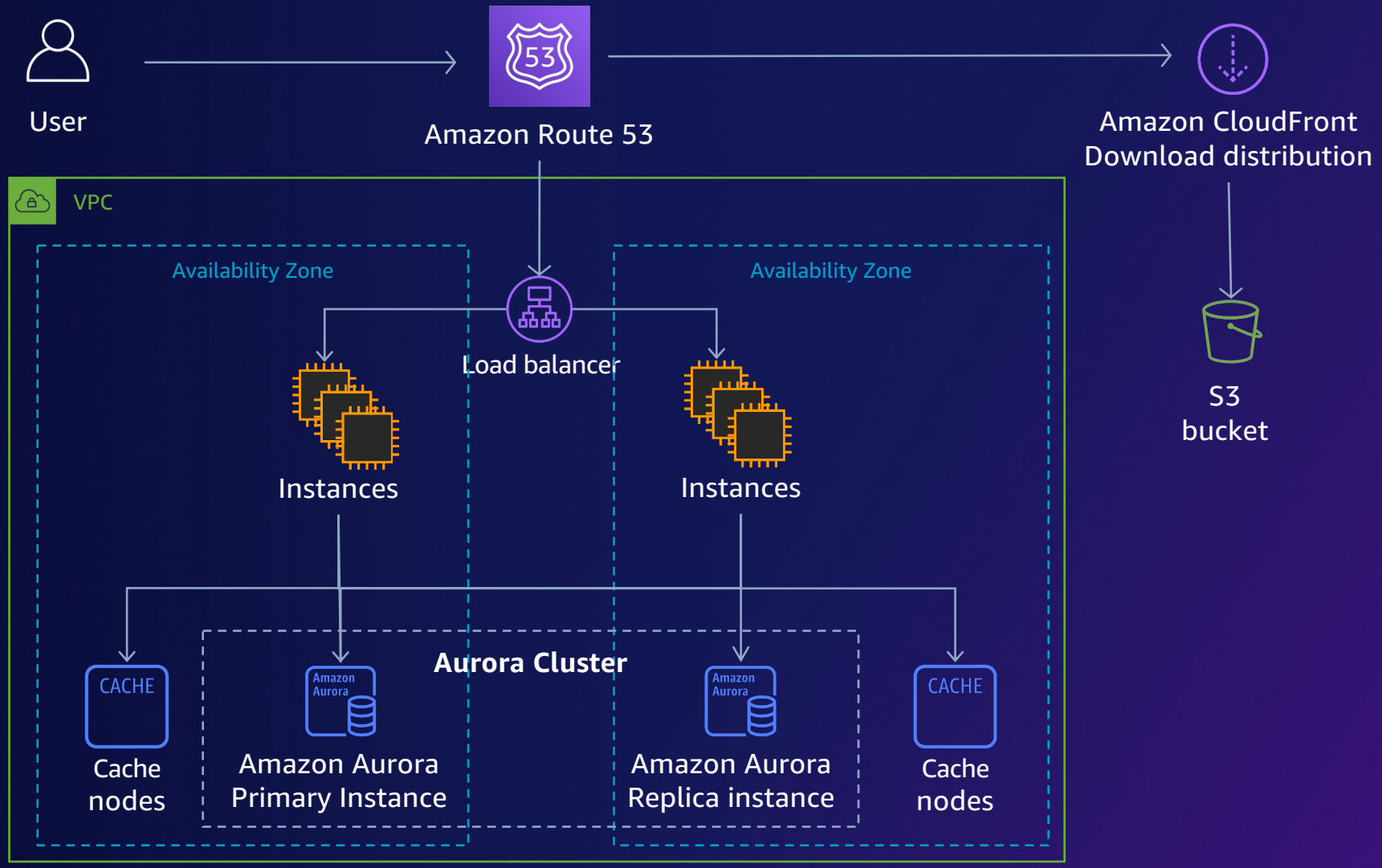
Offload static content



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use in-memory caching

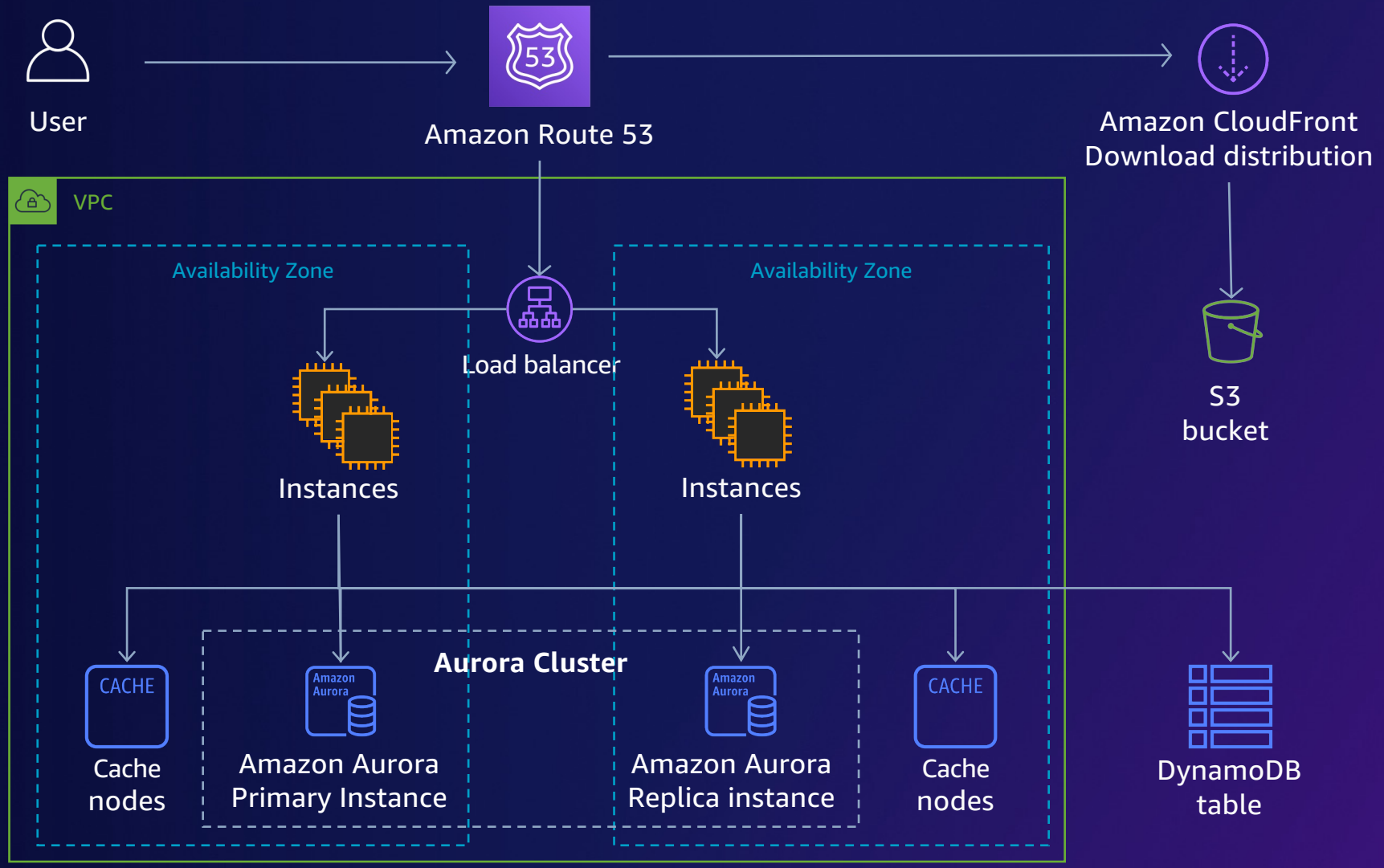
Amazon ElastiCache



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

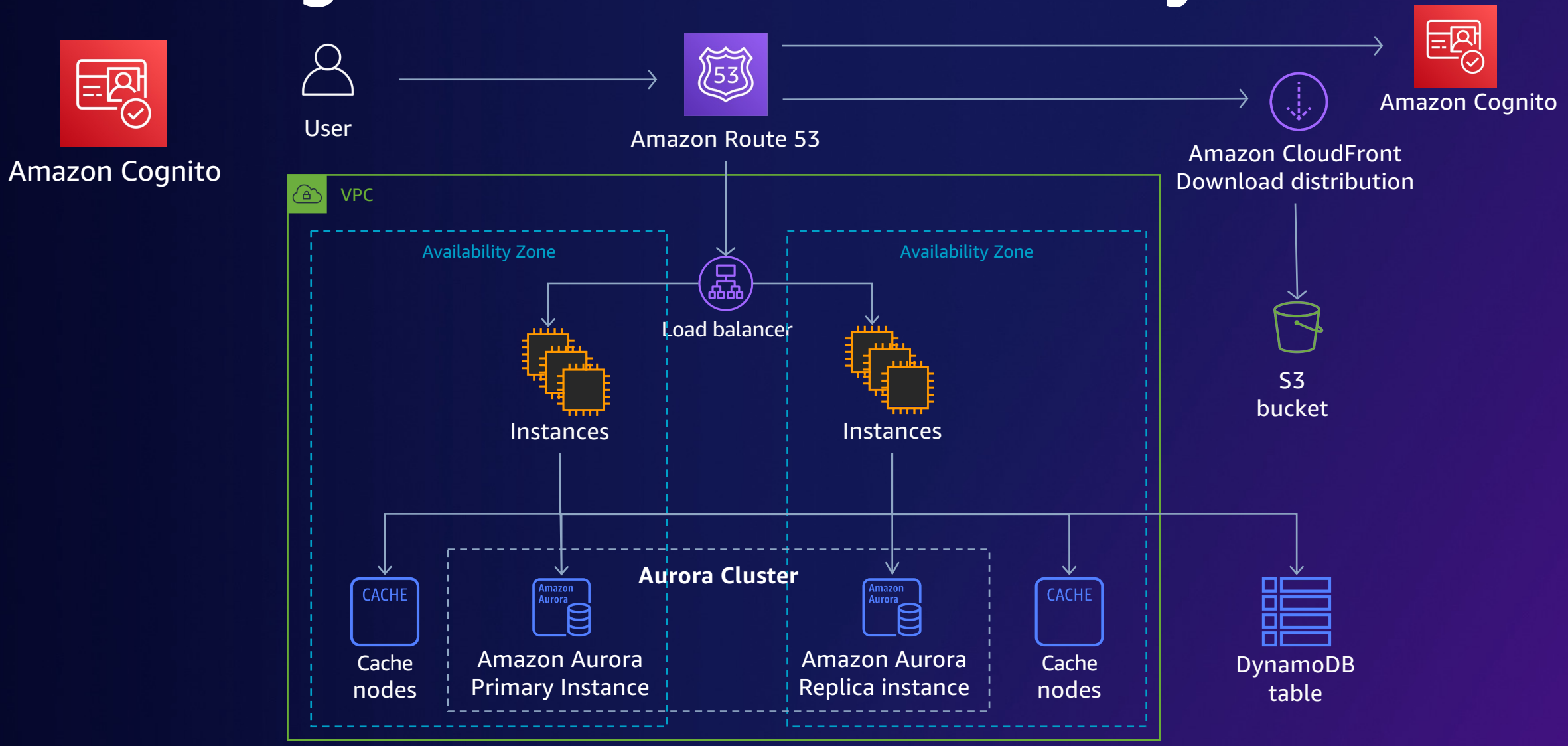
Use purpose-built managed databases

Amazon DynamoDB



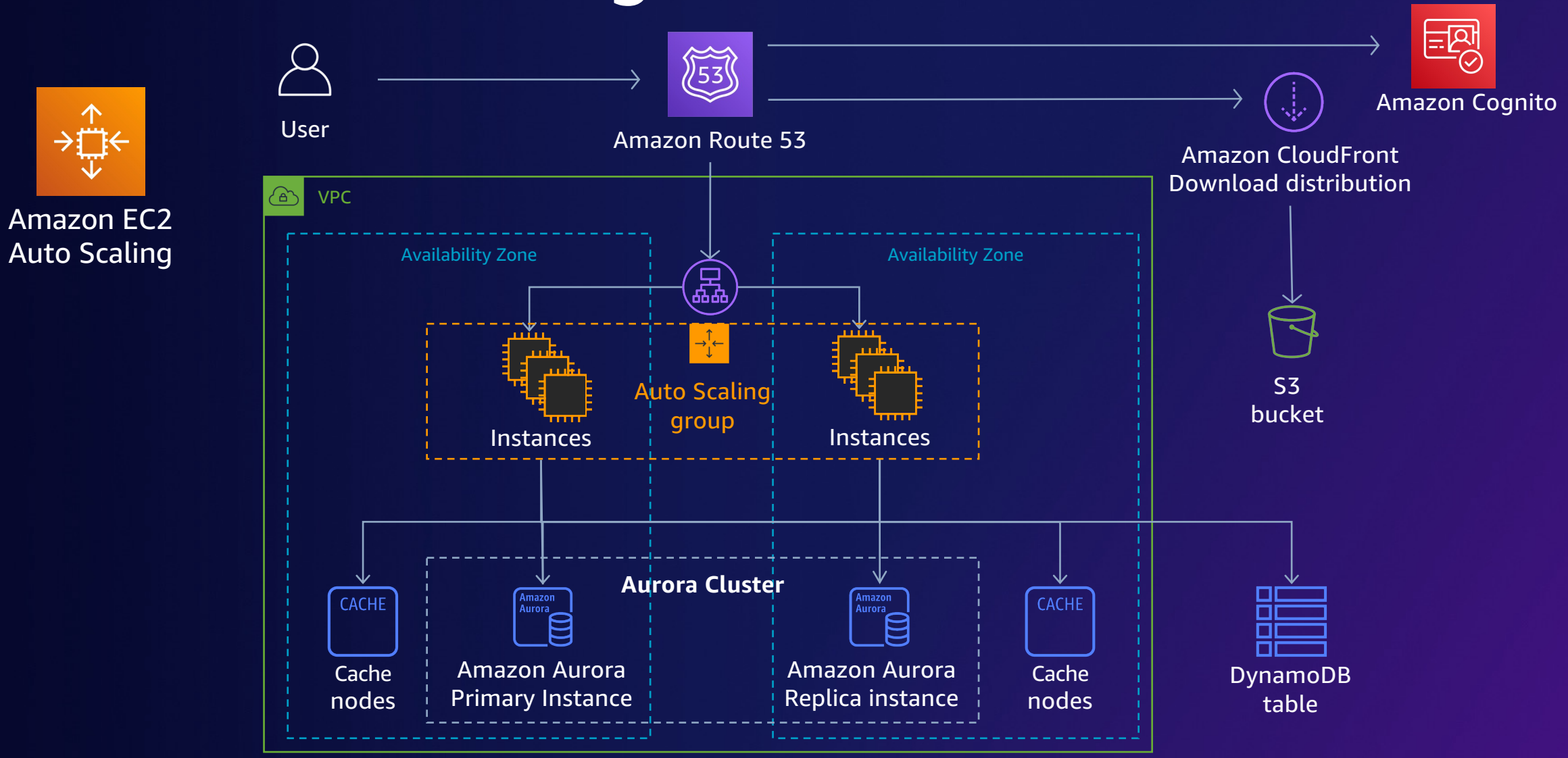
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use managed service for user identity



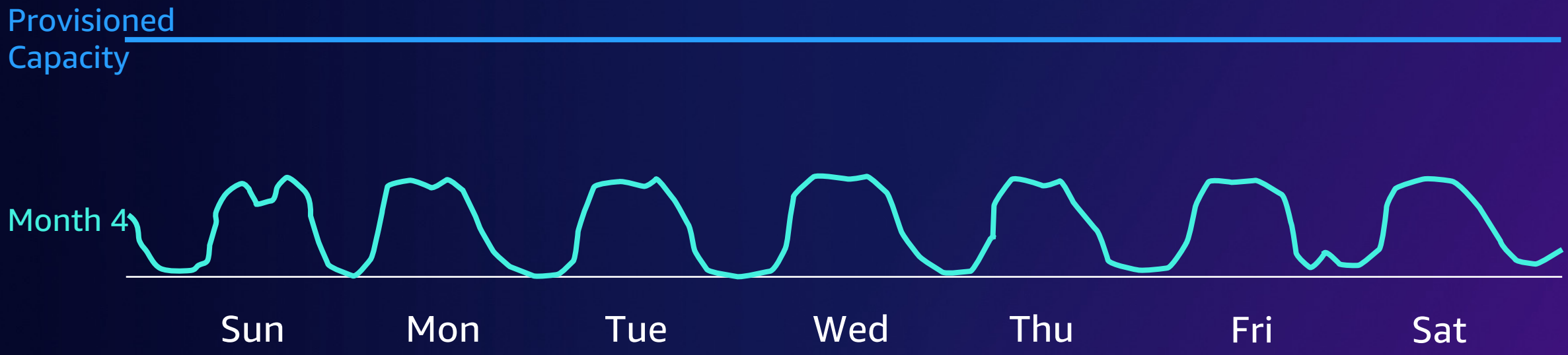
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use EC2 Auto Scaling



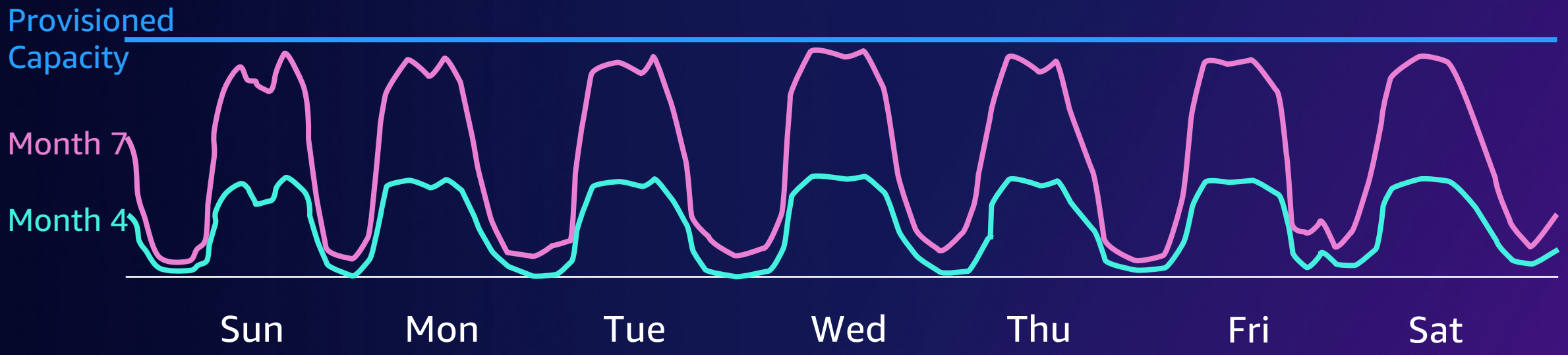
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Weekly traffic to UnicornRentals.com



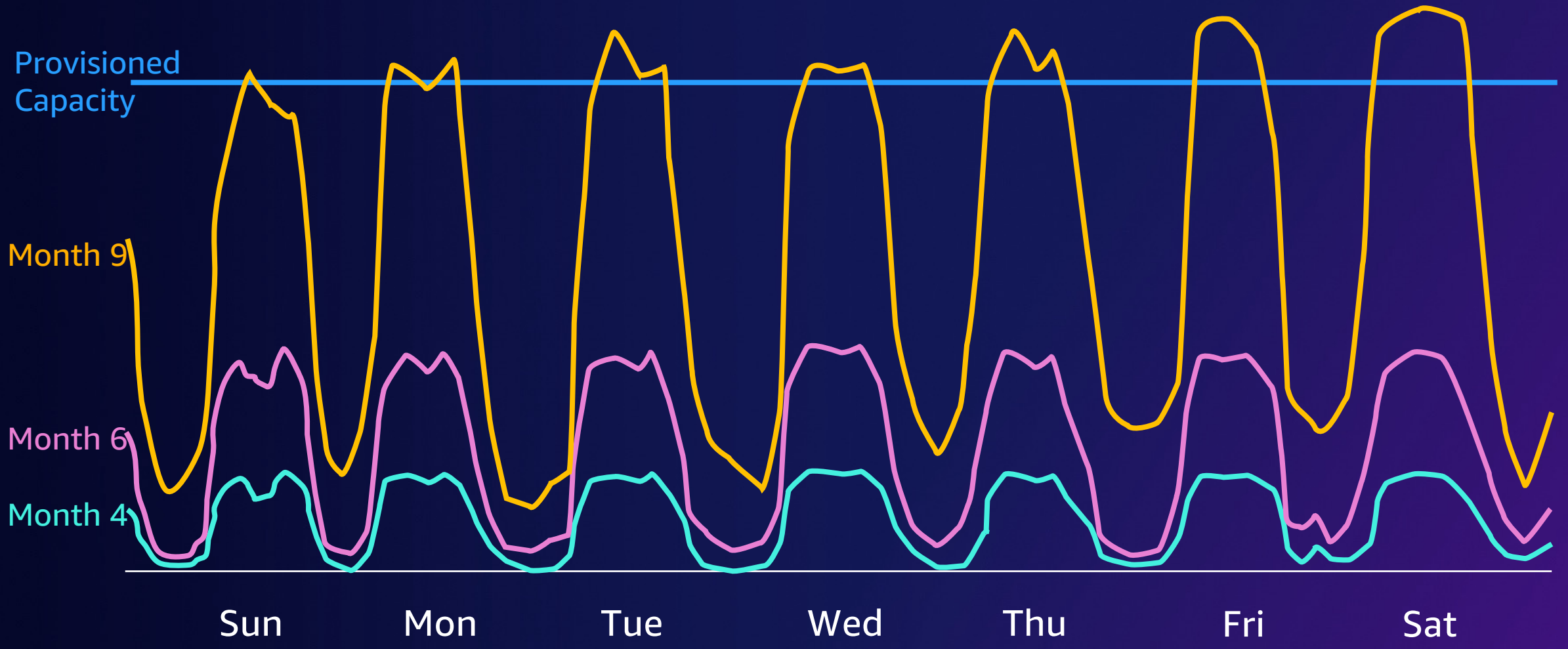
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Weekly traffic to UnicornRentals.com



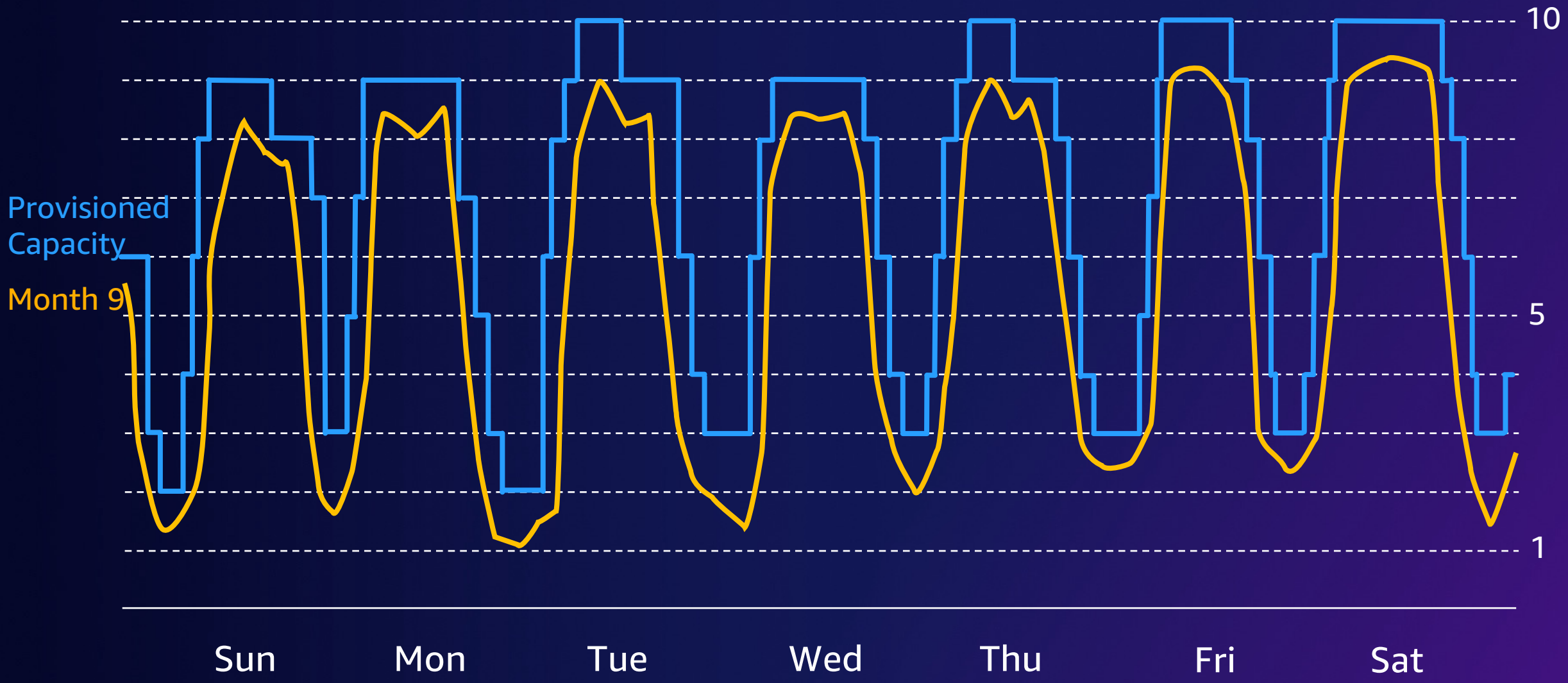
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Weekly traffic to UnicornRentals.com



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How EC2 Auto Scaling can help



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Auto Scaling is not just for EC2 !

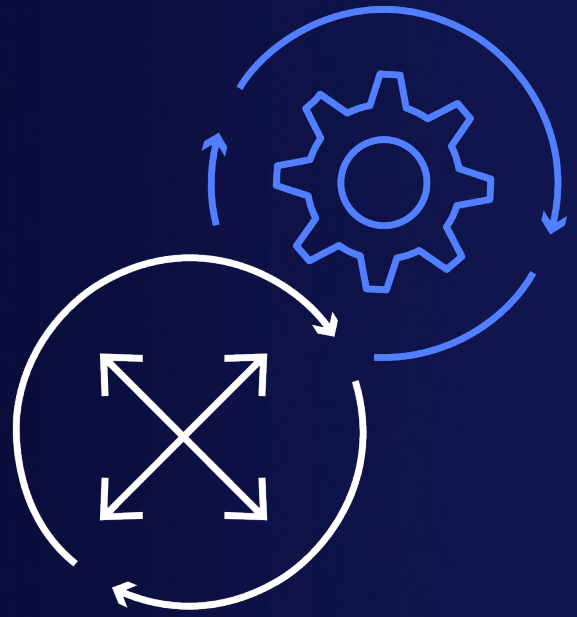
AWS Application Auto Scaling

Use AWS Application Auto Scaling for :

- Amazon Aurora replicas
- Provisioned capacity for Amazon DynamoDB tables
- Other AWS services with scalable resources
- Custom application resources

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Embrace automation!



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automate operations tasks

AWS Systems Manager

- Manage servers in the cloud and on-prem
- Automate common tasks
- Compliance management and reporting

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automate infrastructure build

AWS Cloud Development Kit
(AWS CDK)

AWS Tools
and SDKs

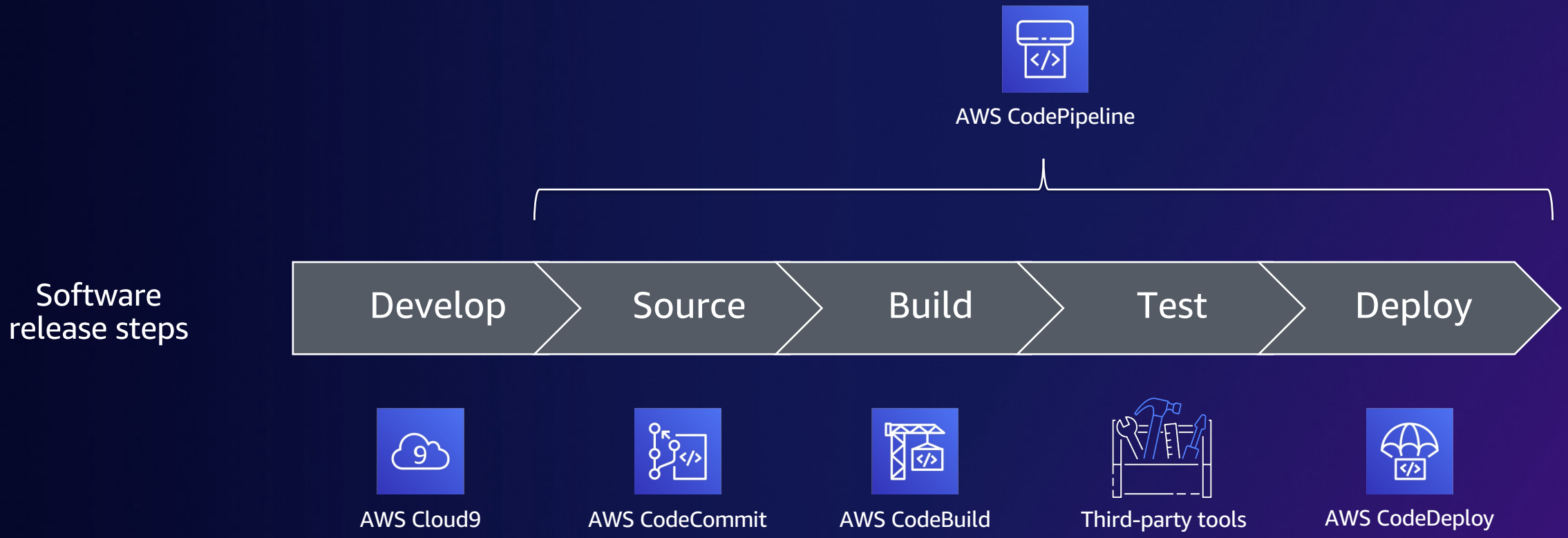
AWS Command Line
Interface (AWS CLI)

AWS CloudFormation

AWS Service APIs

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automate deployment pipeline



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automate monitoring, alarms, and response

Amazon CloudWatch

- Collect: Metrics and logs
- Monitor: Alarms and dashboards
- Act: Auto scaling and events
- Analyze: Trends and metric math

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Approaching 1 million monthly active users…

What further improvements
can be made in the
web / application layer?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Monolithic architecture

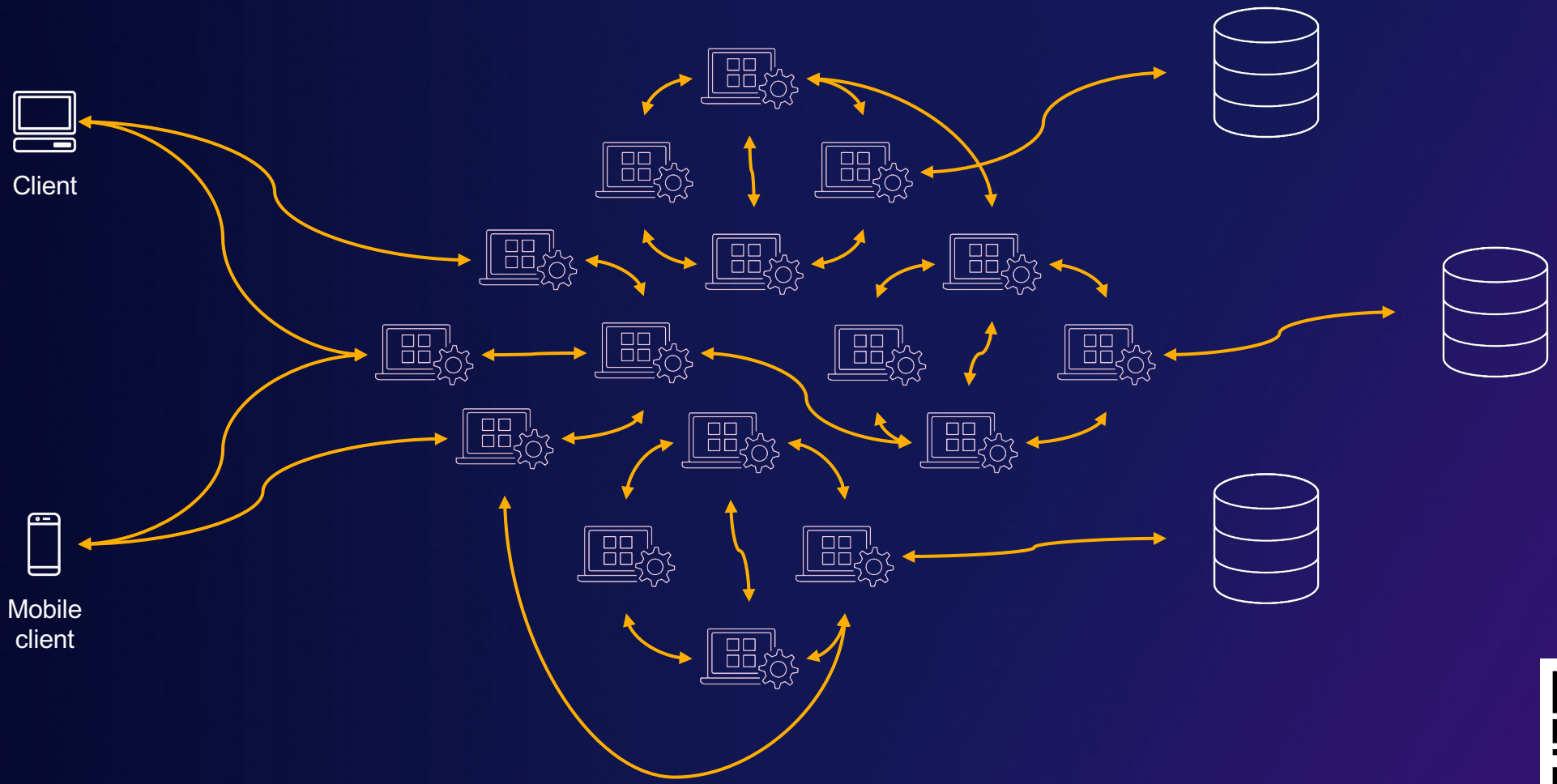
Mix of business functions including:

- Ordering
- Invoicing
- Fulfilment
- Forecasting
- Analytics / Reporting

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Microservices architecture



https://aws.amazon.com/microservices/

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Compute options

Function as a service

AWS Lambda

Scales automatically

Containers

Amazon ECS

Amazon EKS

Scale service

AWS Application Auto Scaling

Kubernetes Horizontal Pod Autoscaler

Scale cluster

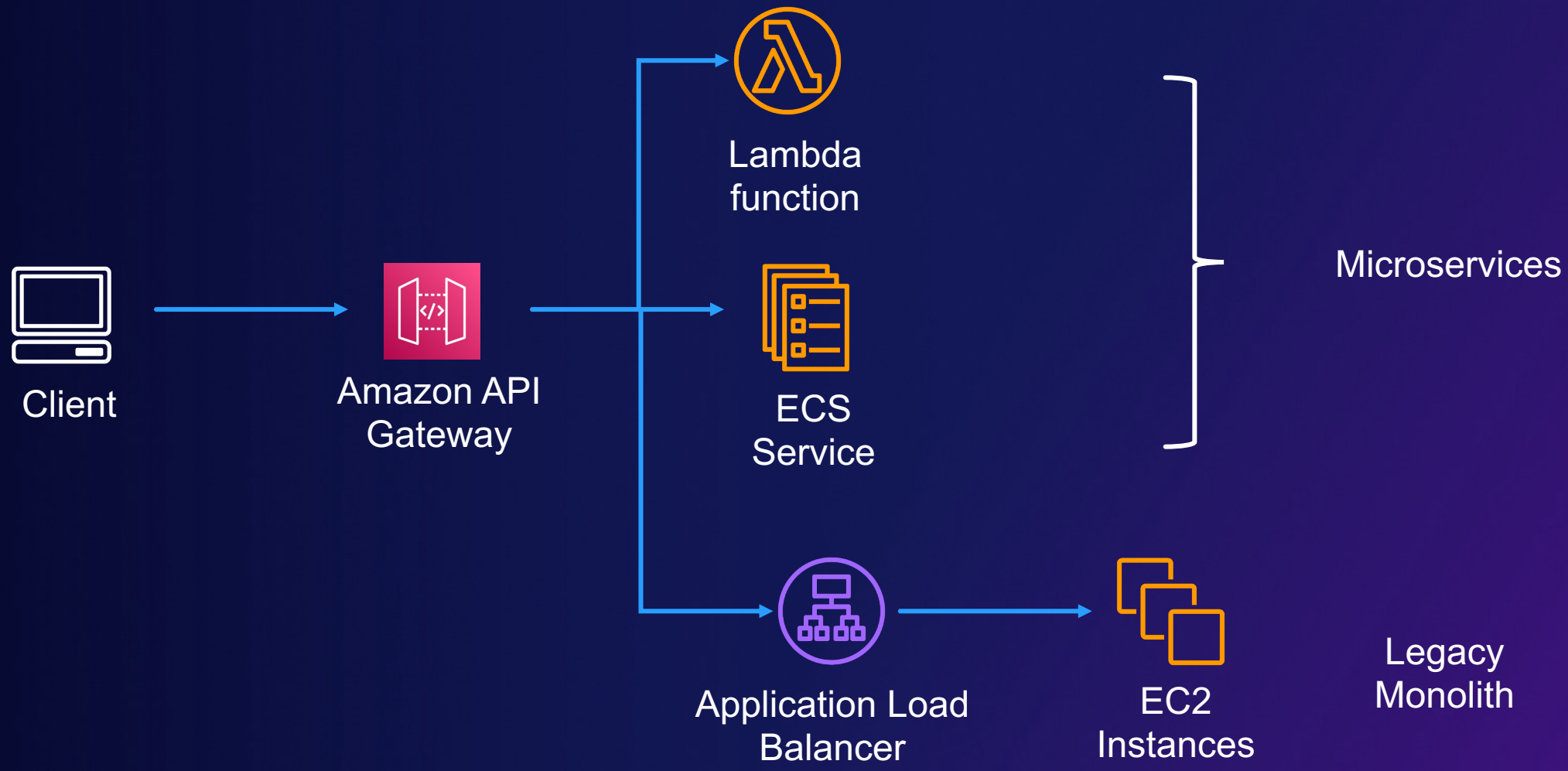
AWS Auto Scaling

AWS Karpenter

aws

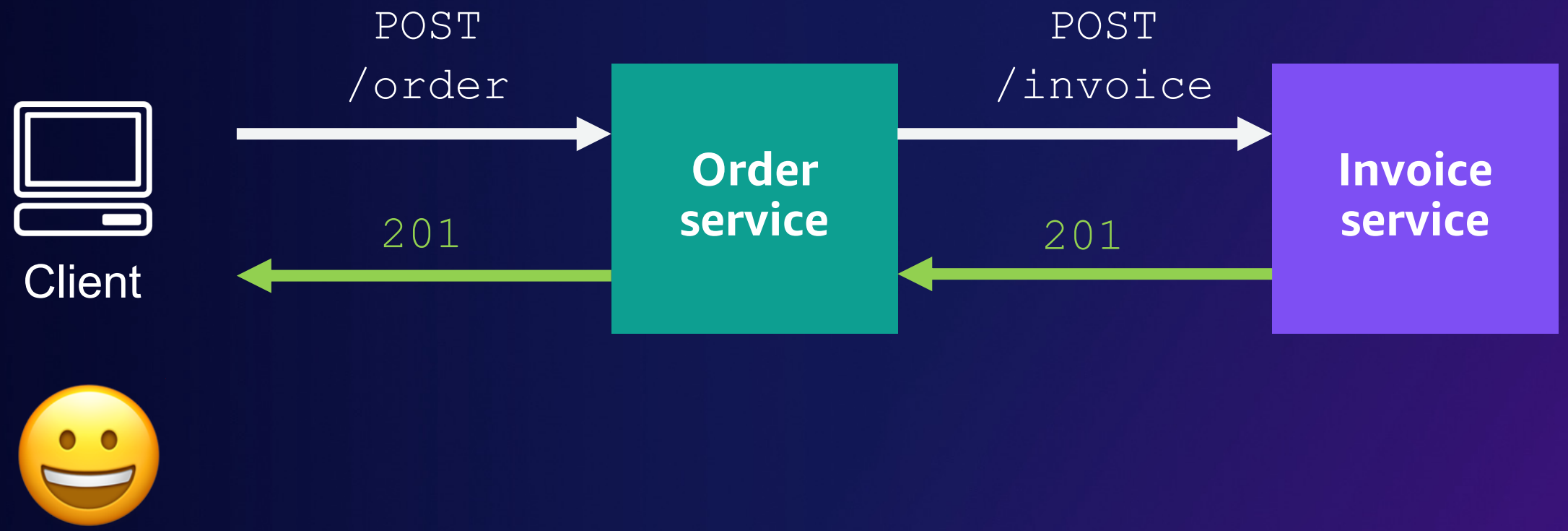
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Evolving to microservices: Strangler pattern



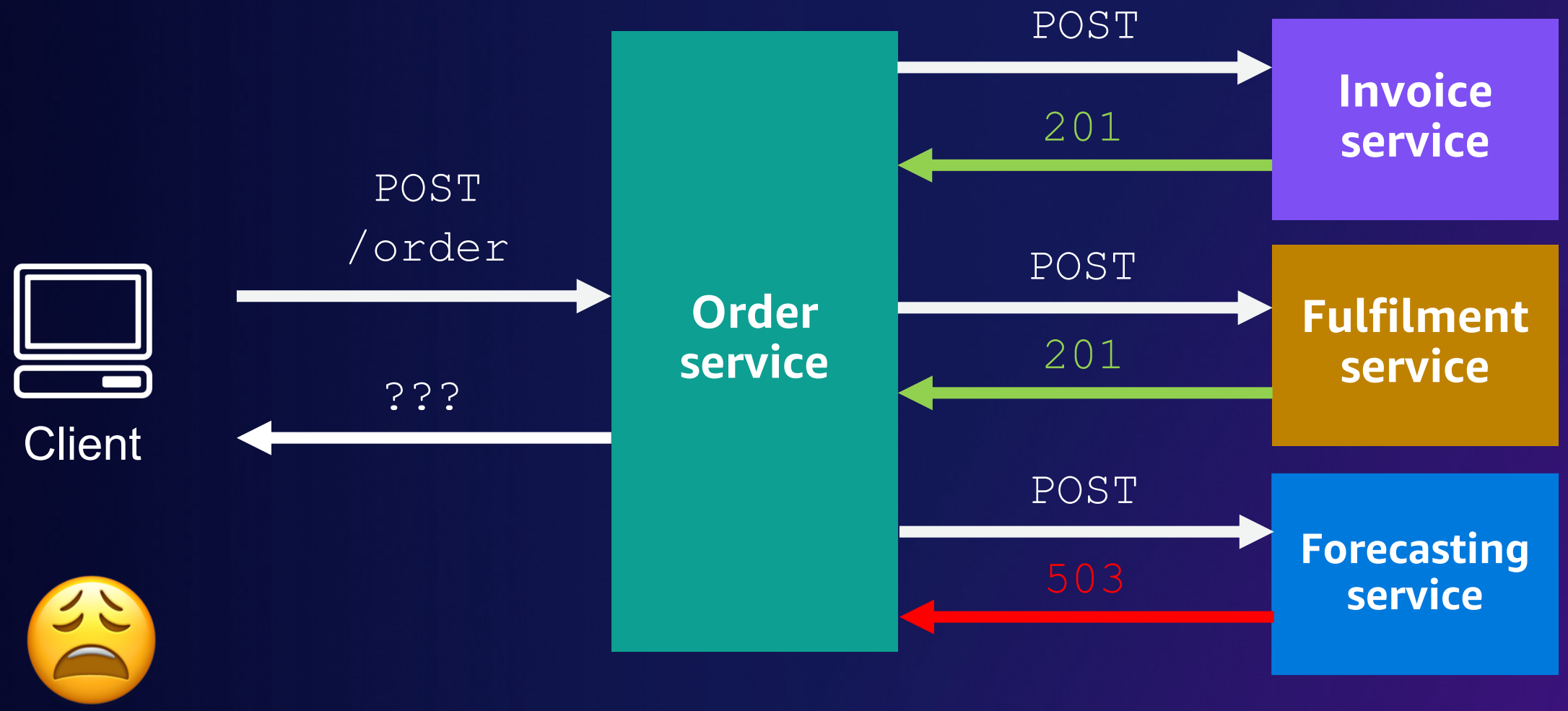
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Inter-service communication



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Synchronous API challenges



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

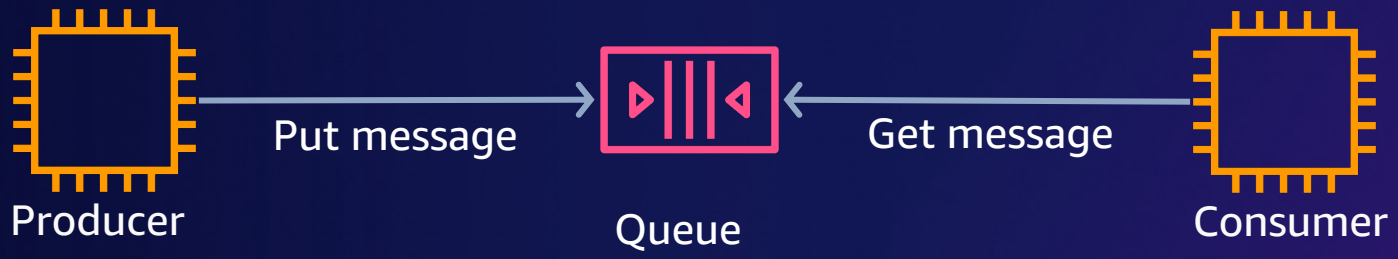
How do we decouple inter-service communication?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Message Queue for Asynchronous Communication

Amazon SQS

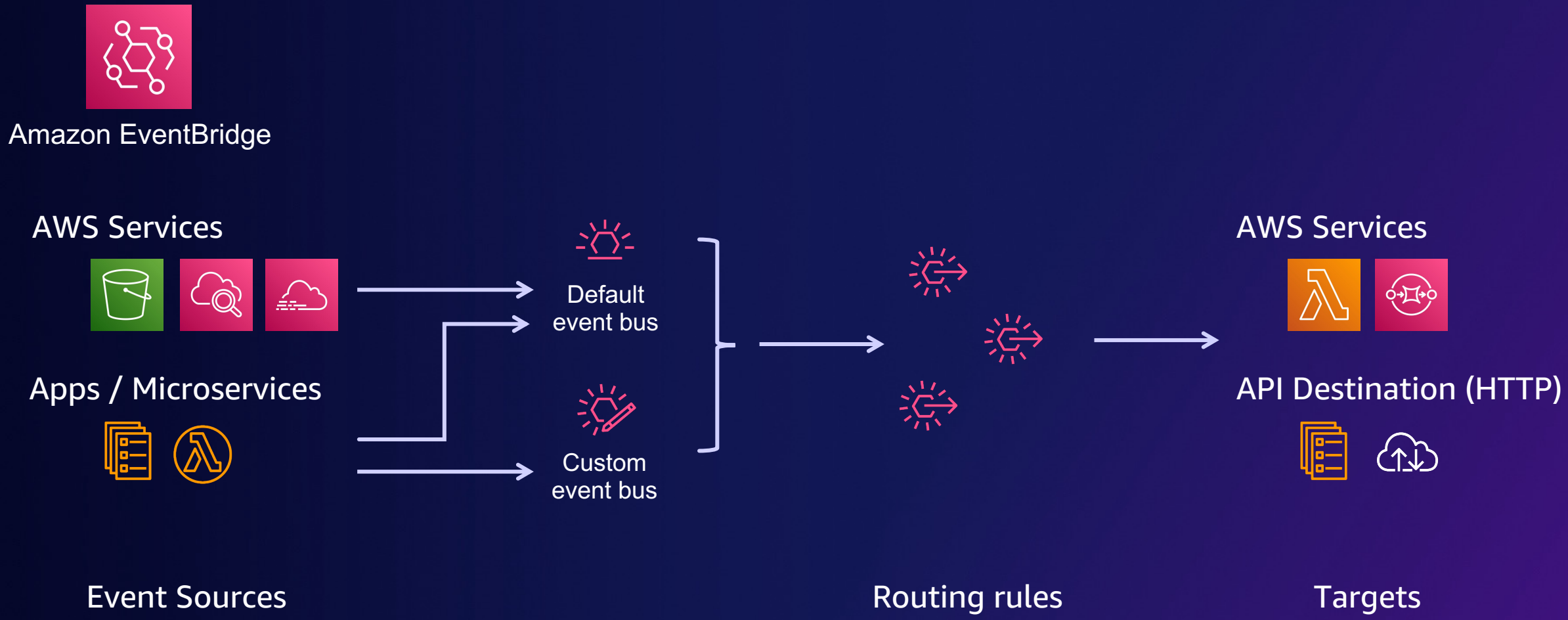
- Reliable (multi-AZ)
- Scalable
- Secure



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

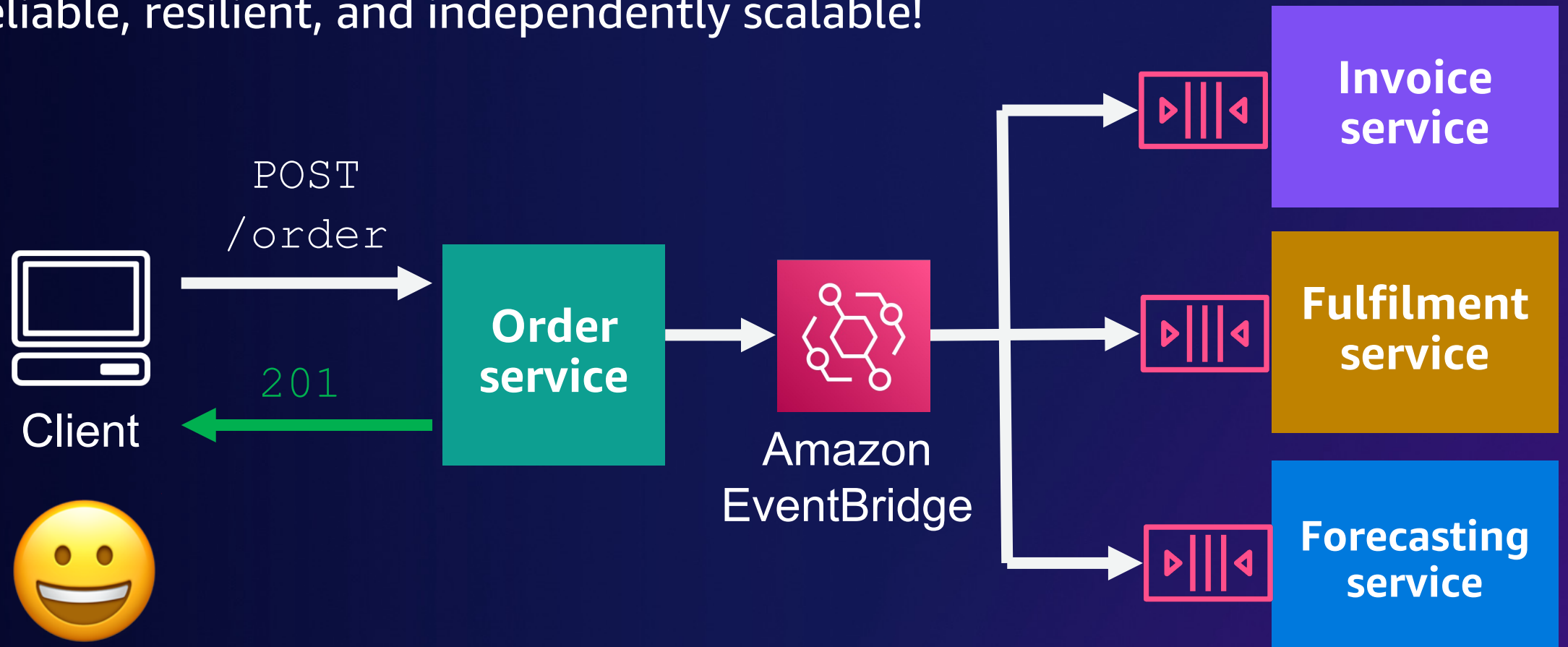
Event routing for 1-*many* Event Notifications



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Event routing with message queue

Reliable, resilient, and independently scalable!

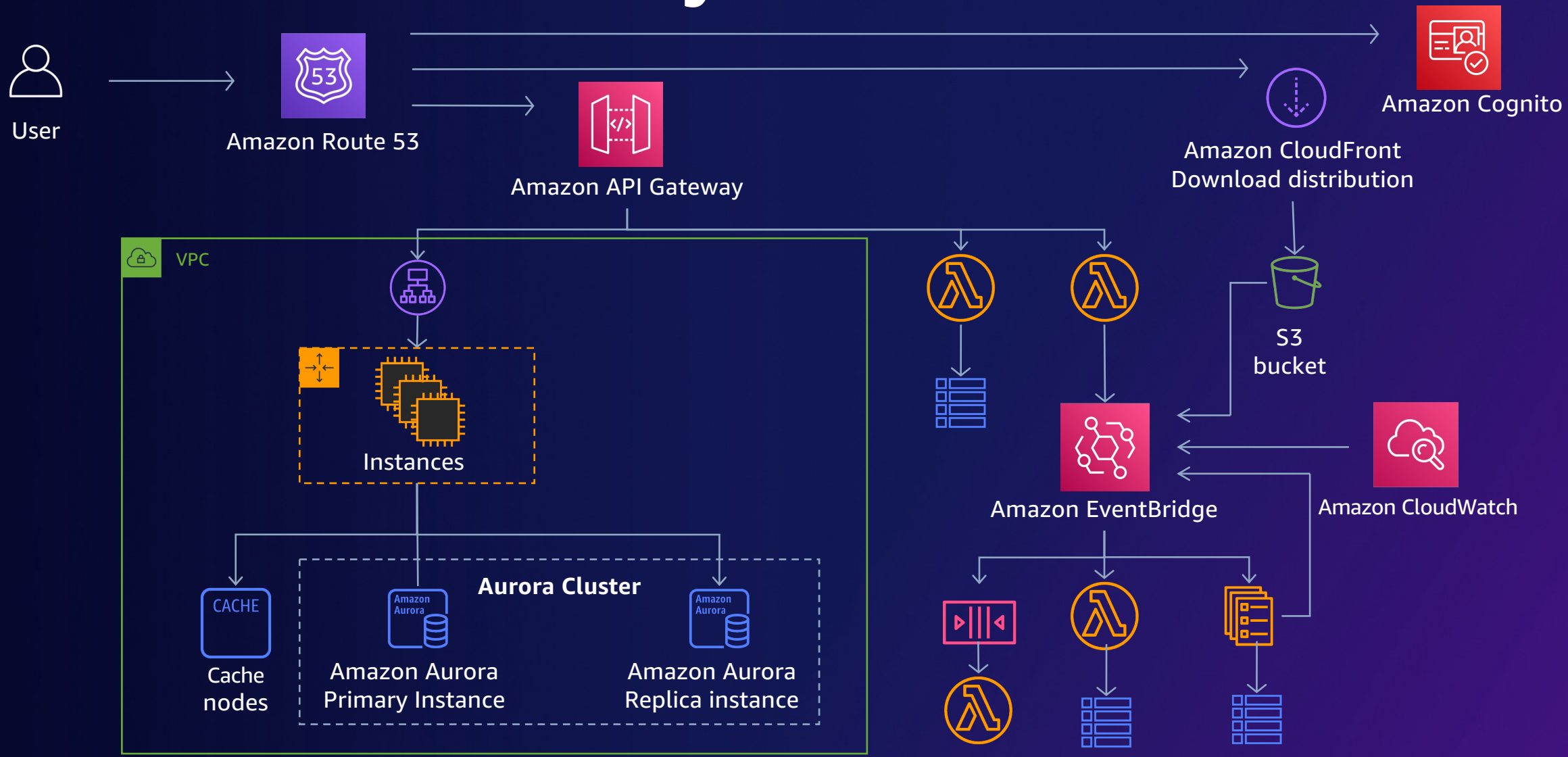


"If your application is cloud-native, or large-scale, or distributed, and doesn't include a messaging or event component, that's probably a bug."

Tim Bray
General-purpose Internet-software geek

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

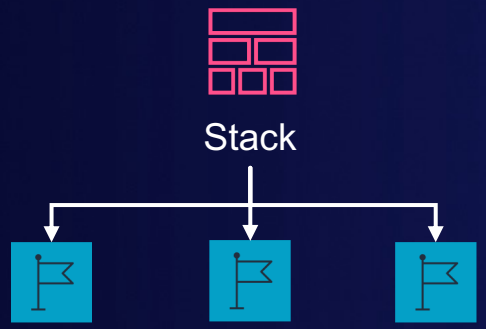
1 million users and beyond...



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

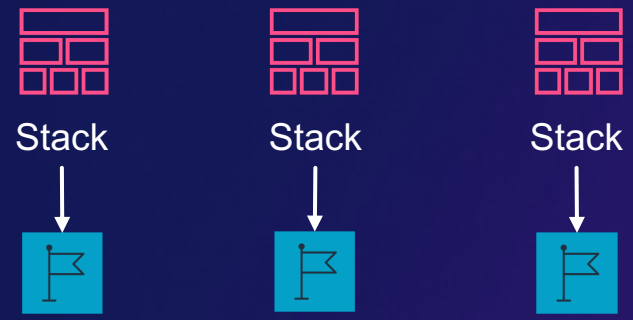
Multi-geo / Multi-region

One multi-region stack



- Data replicated across regions
- Consider multi-master

Stack per region / geo



- Data sharded by region / geo

- Combine approaches for hybrid
- Use Amazon CloudFront
  - Cache objects
  - Shortest path to AWS network

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does scaling impact cost-efficiency and sustainability?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Well-Architected Framework

Describes key concepts, design principles, and architectural best practices for designing and running workloads in the cloud.

Built around six pillars:

- Operational Excellence
- Security
- Reliability
- Performance Efficiency
- Cost Optimization
- Sustainability

https://aws.amazon.com/architecture/well-architected/

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Some relevant best practices and examples

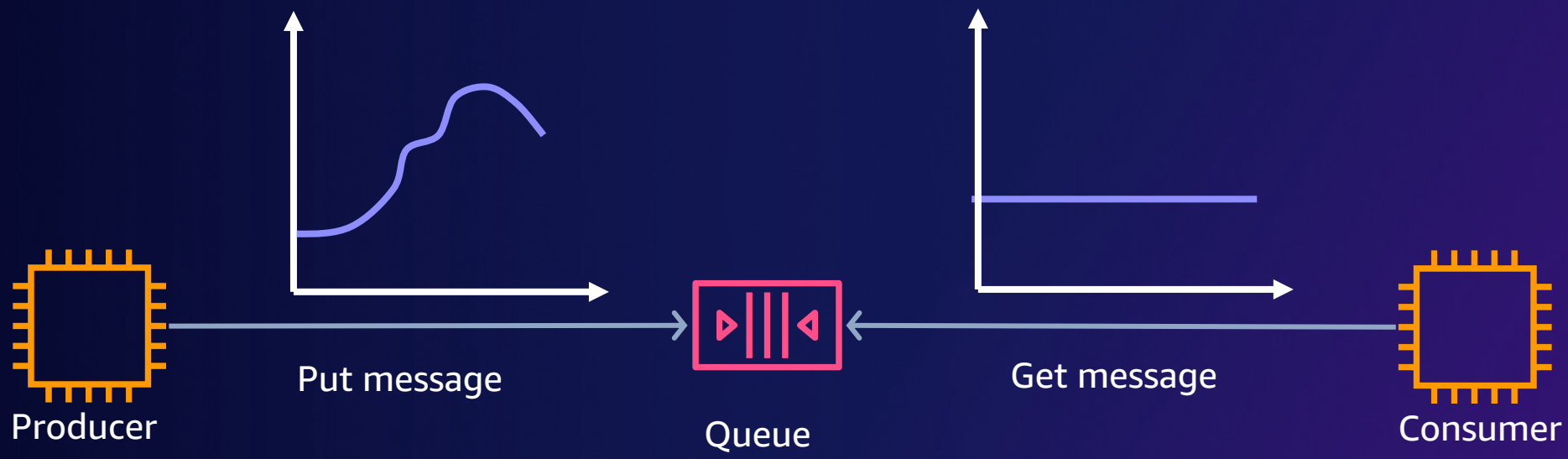
Adopt new, more efficient hardware and software offerings	• Graviton processors, new instance types
Eliminate unused assets	• Use lifecycle policies to archive / delete old data
Use efficient file formats	• Parquet better than CSV for analytics
Optimize code	• Use a code profiler to identify targets for optimization

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

For sustainability: Maximize resource utilization

- Use managed services
- Scale capacity in relation to demand
- Alternatively: Use queues to smooth demand



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

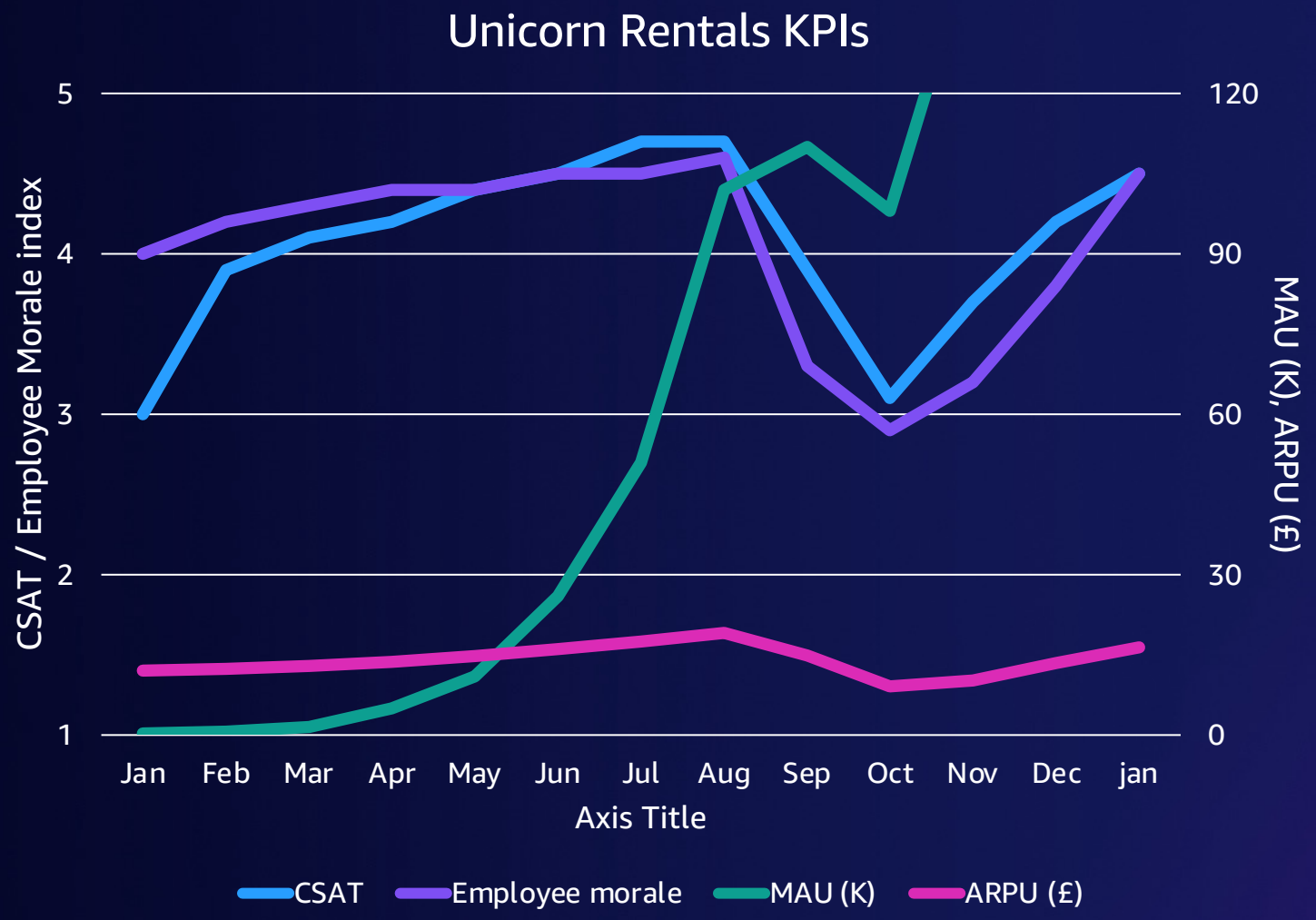
Recap: Key takeaways

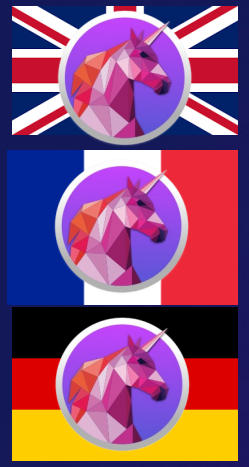
- Break things up into smaller components
- Embrace automation
- Leverage managed services / serverless / edge services
- Monitor resources and re-assess efficiency
- Use an event-driven architecture for loose coupling
- Understand how scalability relates to cost & sustainability targets

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

And Ada is happy again !





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Resources

AWS Well-Architected Framework
Includes a white paper for each of the six Well-Architected Framework pillars

AWS Cloud Databases
Modernize your data infrastructure with fully managed, purpose-built databases

AWS Global Infrastructure
Find out more about AWS Regions, Availability Zones, Edge Locations, Local Zones, and AWS Wavelength.

Microservices on AWS
Microservices architectures make applications easier to scale and faster to develop, enabling innovation and accelerating time-to-market for new features.

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Learn in-demand AWS Cloud skills

AWS Skill Builder

Access **500+ free** digital courses and Learning Plans

Explore resources with a variety of skill levels and **16+** languages to meet your learning needs

Deepen your skills with digital learning on demand

Train now

AWS Certifications

Earn an industry-recognized credential

Receive Foundational, Associate, Professional, and Specialty certifications

Join the **AWS Certified community** and get exclusive benefits

Access **new** exam guides

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Dr Mike Rizzo

@rizzomj

https://www.linkedin.com/in/dr-mike-rizzo/

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Please complete the session survey

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.