



aws INNOVATE

EVERY APPLICATION EDITION

SEPTEMBER 01, 2023

Scaling from 1 to million customers on AWS

Venkatramana Ameth Achar

Solutions Architect

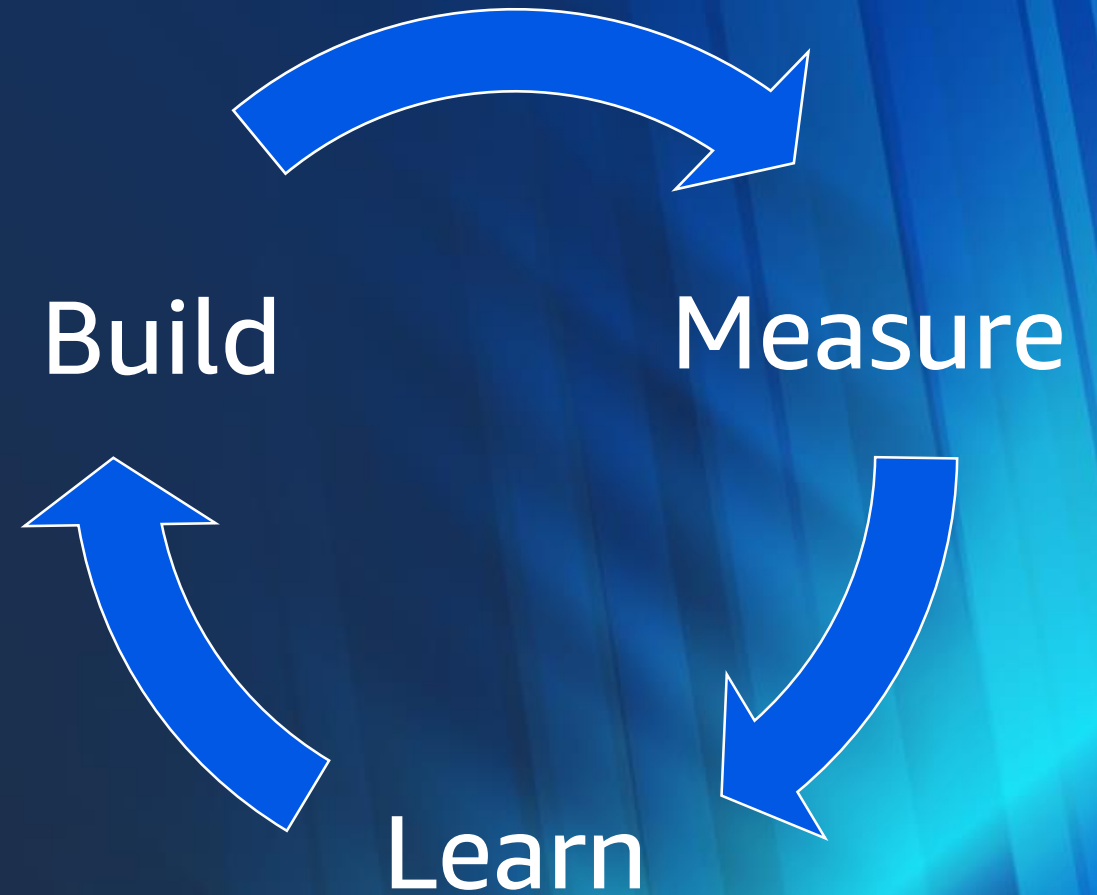
AWS India

“There is no compression algorithm for experience.”

Andy Jassy
CEO, AWS

An approach to scaling...

- 1-way door and 2-way door decisions
- Build Measure Learn
- Avoid undifferentiated heavy lifting



Scaling to 100+ customers

Scaling to 100+ customers : Build

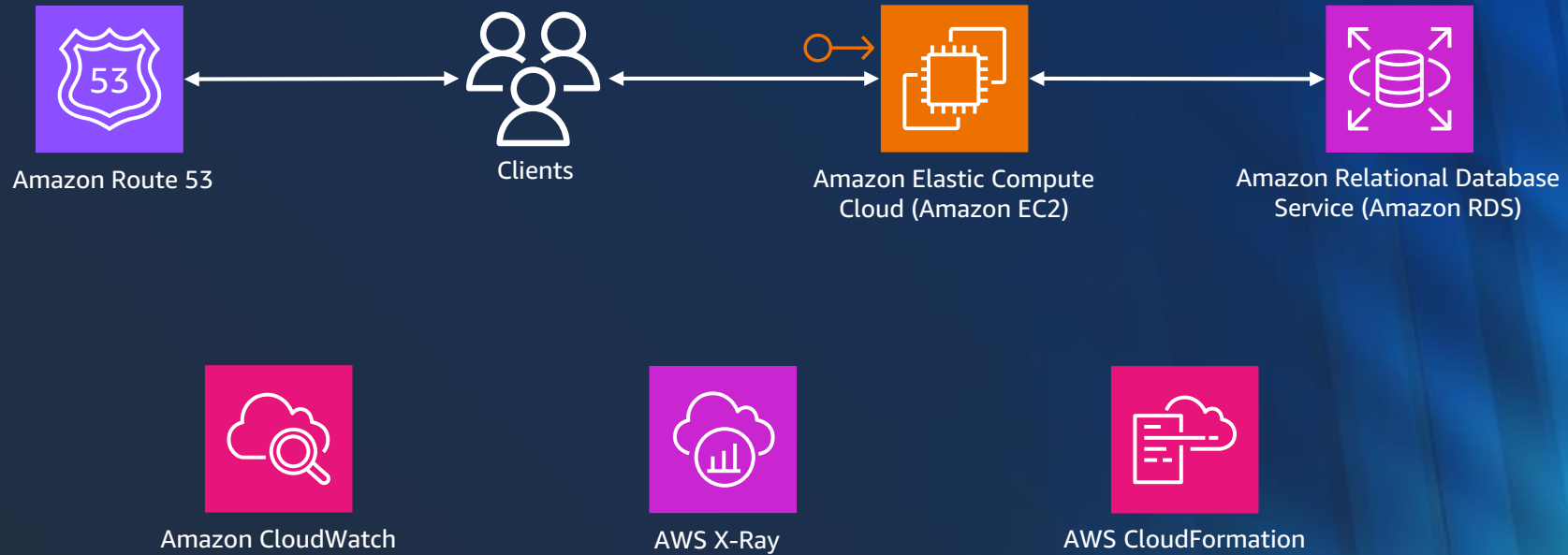
Development

- Scope
- Application stack
- Loosely coupled components
- Application performance metrics

Deployment

- Infrastructure sizing
- Infrastructure monitoring
- Repeatability in deployment

Scaling to 100+ customers : Measure



Reference Architecture

Scaling to 100+ customers : Measure (dive deep)



Scaling to 100+ customers : Learn

Performance

- Latency
- Spiky workloads
- Unoptimized and read only queries

Availability

- Single point of failures
- Only one of three availability zones used

Cost

- Explore cost optimization options

Security

- No identity context within the application
- Application open to Layer 7 attacks

Development

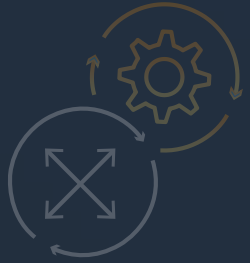
- Application development is taking too long
- Manual release management

Positives

- Quick proof-of-concept
- Customer feedback

Scaling to 100+ customers : Best Practices check

AWS Well-Architected Framework



Operational
excellence



Security



Reliability



Performance
efficiency



Cost
optimization



Sustainability

Scaling to 10000+ customers

Scaling to 10000+ customers : Build

Scaling development

- Accelerate coding
- Automate release management
- Stateless application

Performance improvements

- Object storage and CDN
- Auto scale horizontally
- Use cache to offload database
- Database read replica

Security improvements

- Customer identity management
- Web Application Firewall

Availability improvements

- Use multiple availability zones

Cost improvements

- Explore cost optimization

Well-Architected report

- Remediate the risks

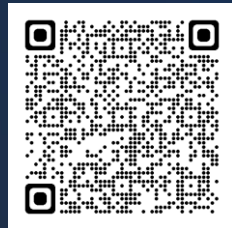
Scaling development

Build applications faster

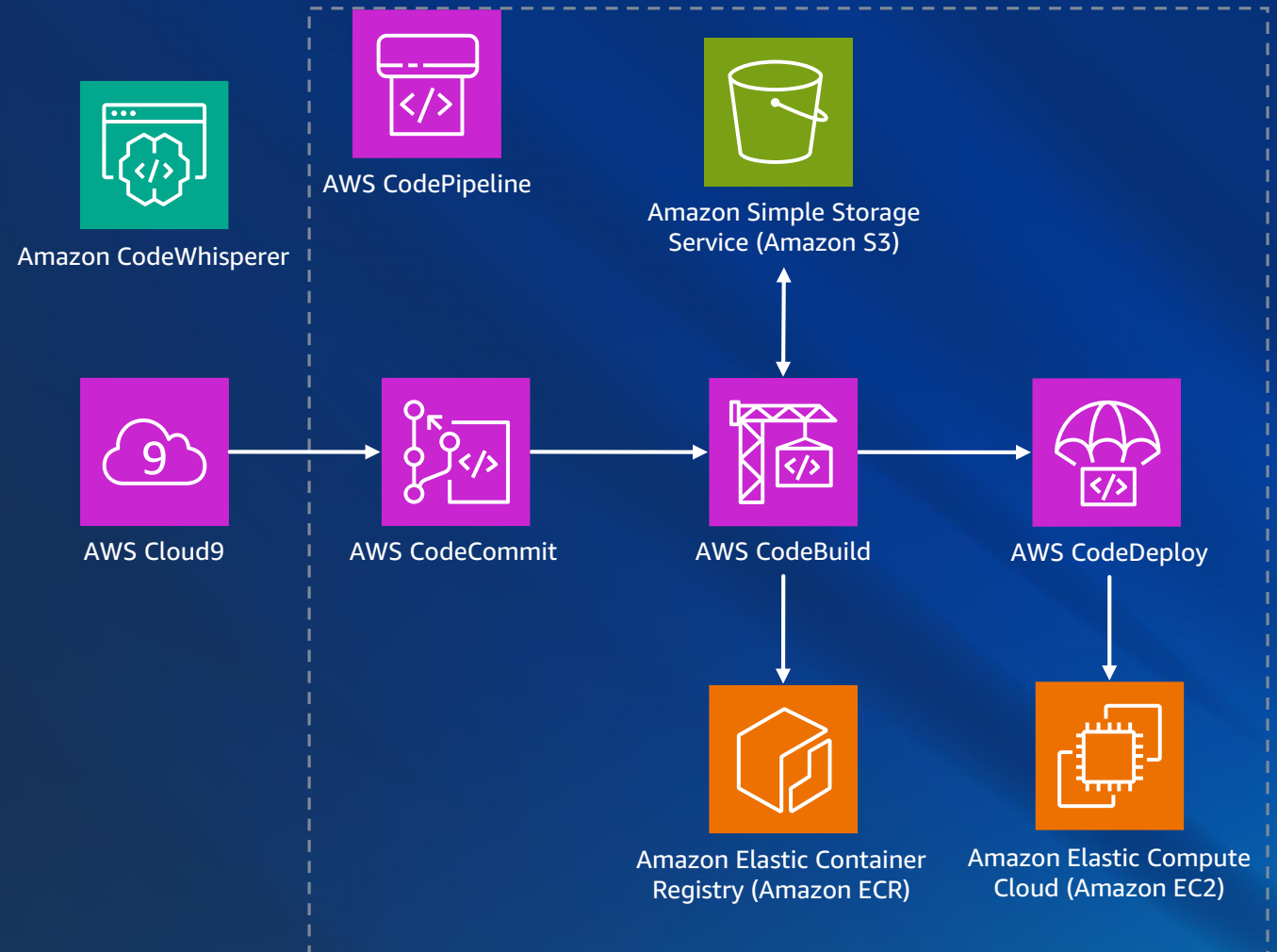
- AI coding companion accelerated development

CI/CD pipeline

- Cloud based IDE to scale development efforts
- Automated build on code commit
- Build image and push to Amazon ECR
- Deploy build to Amazon EC2
- Artifact stored in Amazon S3



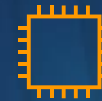
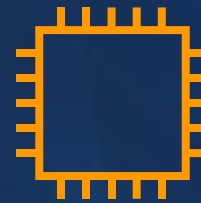
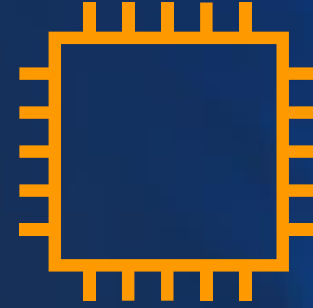
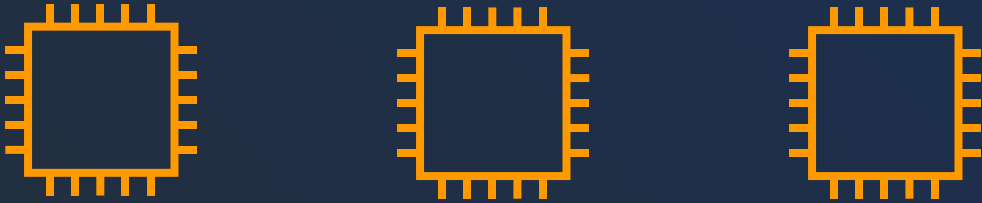
Workshop : Introduction
to AWS Code Family



Reference Architecture

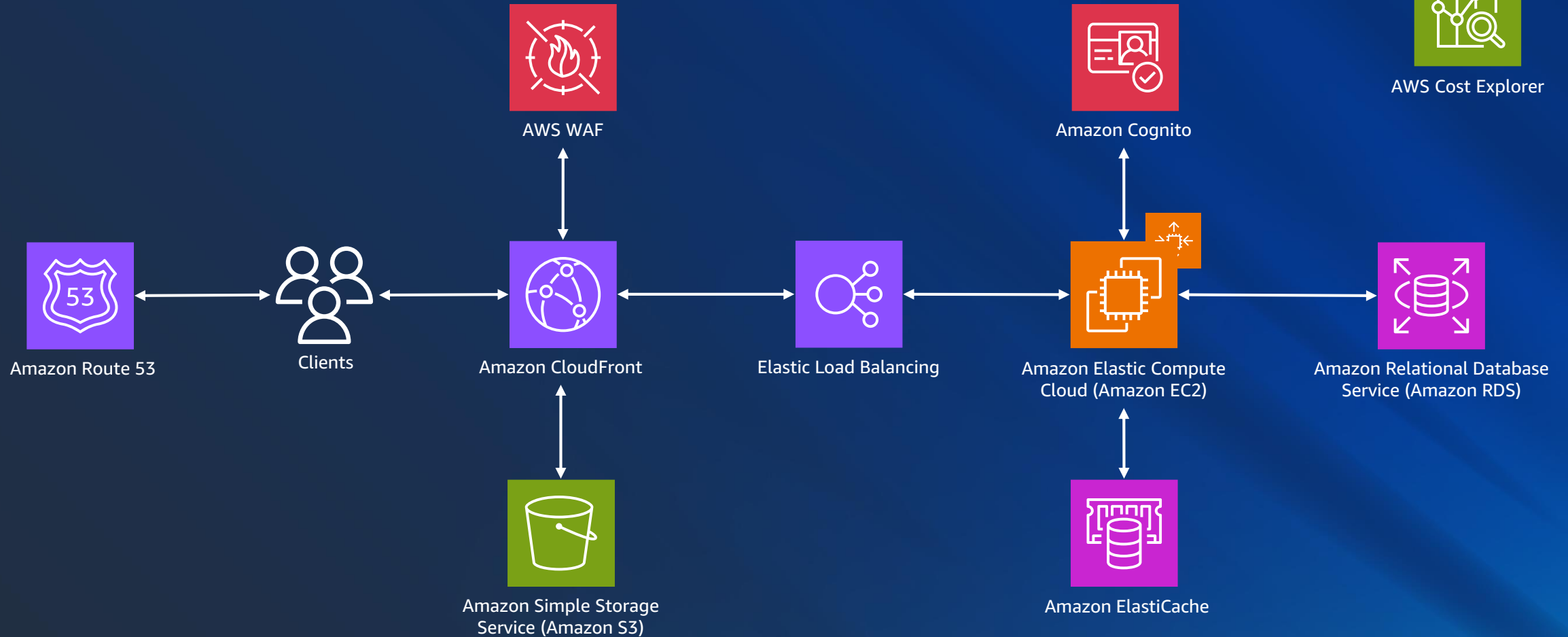
Scaling compute

Horizontally



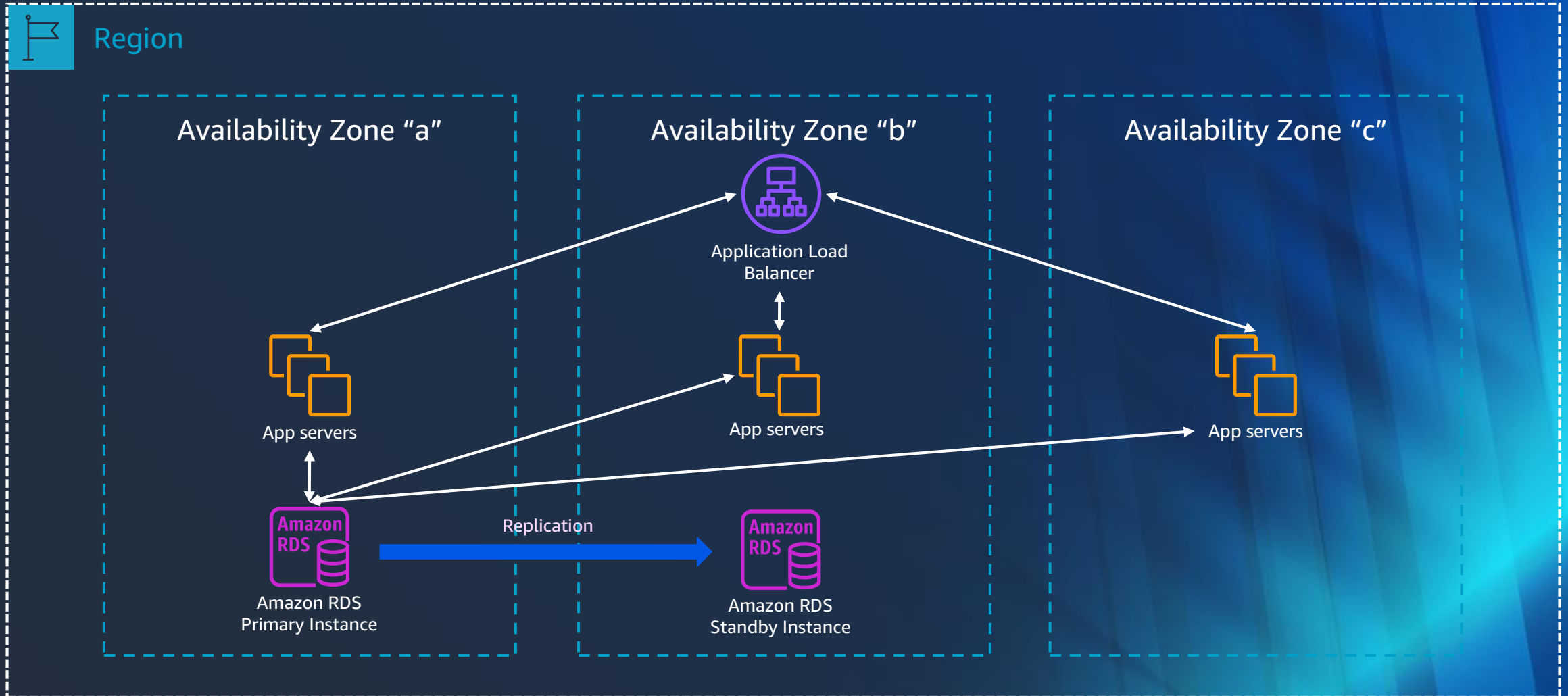
Vertically

Scaling to 10000+ customers : Measure

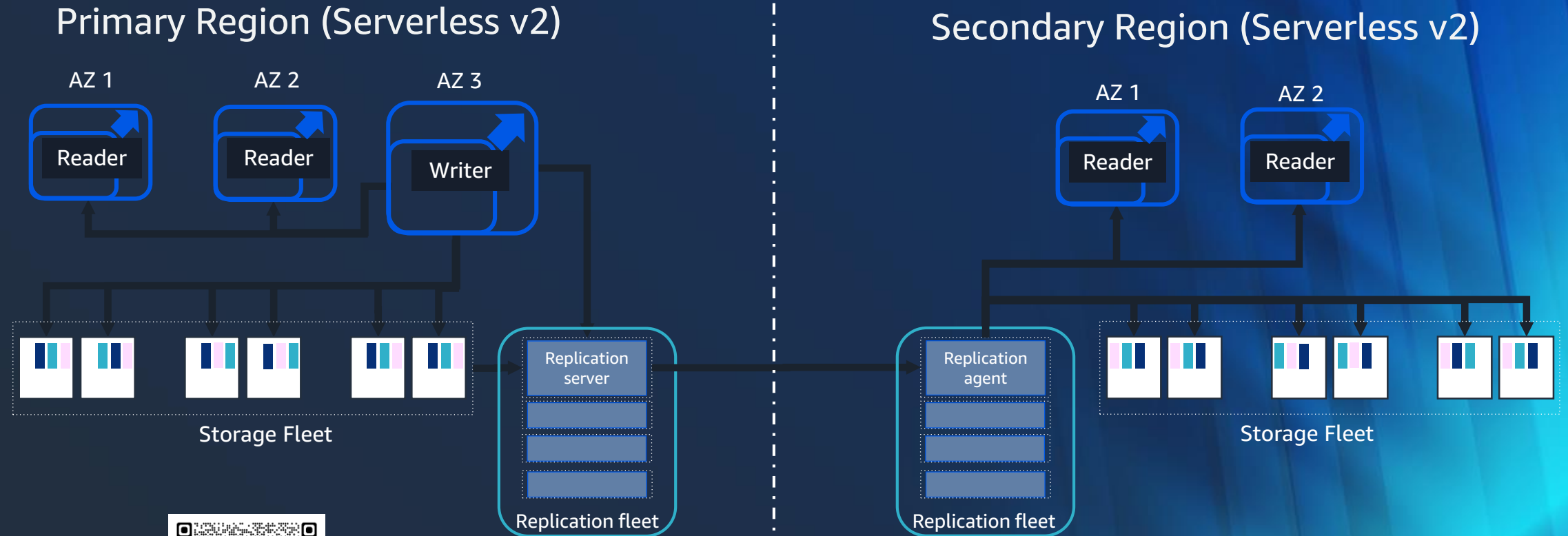


Reference Architecture

Scaling to 10000+ customers : Measure (dive deep)



Aurora Global Database & Aurora Serverless v2



Blog : Evaluate
Amazon Aurora Serverless v2

Scaling to 10000+ customers : Learn

Access patterns

- Interact beyond web interface
- Repeated joins and database access patterns

Operational issues

- Undifferentiated heavy lifting
- Release management related concerns
- Multi-region disaster recovery

Scaling to 10000+ customers : Best practices check

AWS Well-Architected Framework



Scaling to 1M+ customers

Scaling to 1M+ customers : Build

Design

- Managed API service
- Muti-region disaster recovery strategy
- Purpose-built database

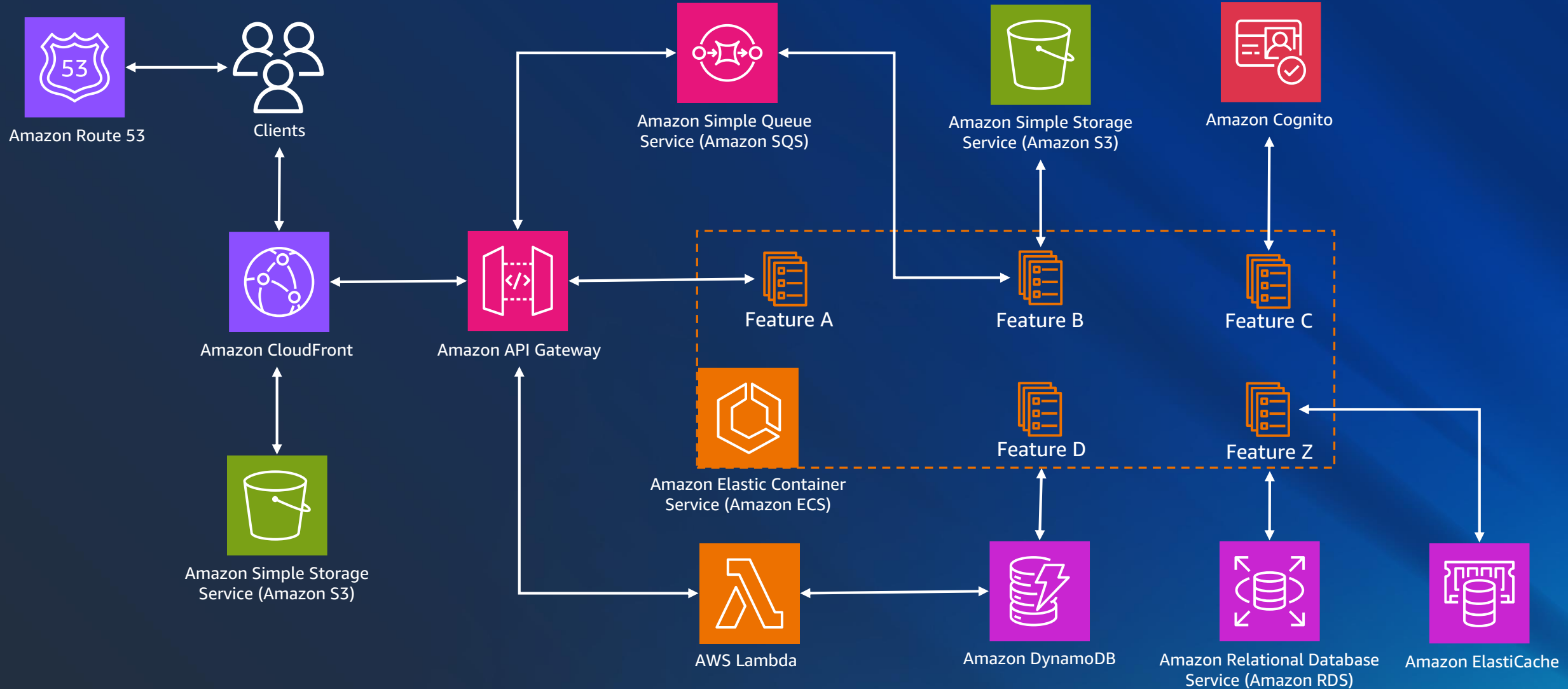
Well-Architected report

- Remediate the risks

Operational

- Container orchestration
- Manage and scale features independently

Scaling to serve 1000000+ customers : Measure



Strategies for disaster recovery

Active/passive

Backup and
restore

Pilot light

Warm
standby

Multi-site
active/active

RPO/RTO:
Hours

RPO/RTO:
Tens of minutes

RPO/RTO:
Minutes

RPO/RTO:
Real time

- Low-priority use cases
- Provision all AWS resources after event
- Restore backups after event
- Cost \$

- Data live
- Services idle
- Provision some AWS resources and scale after event
- Cost: \$\$

- Always running, but smaller
- Business critical
- Scale AWS resources after event
- Cost \$\$\$

- Zero downtime
- Near-zero data loss
- Mission-critical services
- Cost \$\$\$\$

Multi-region active/active

Multi-writer, multi-region



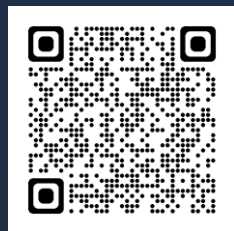
Multi-region architecture

Design principles

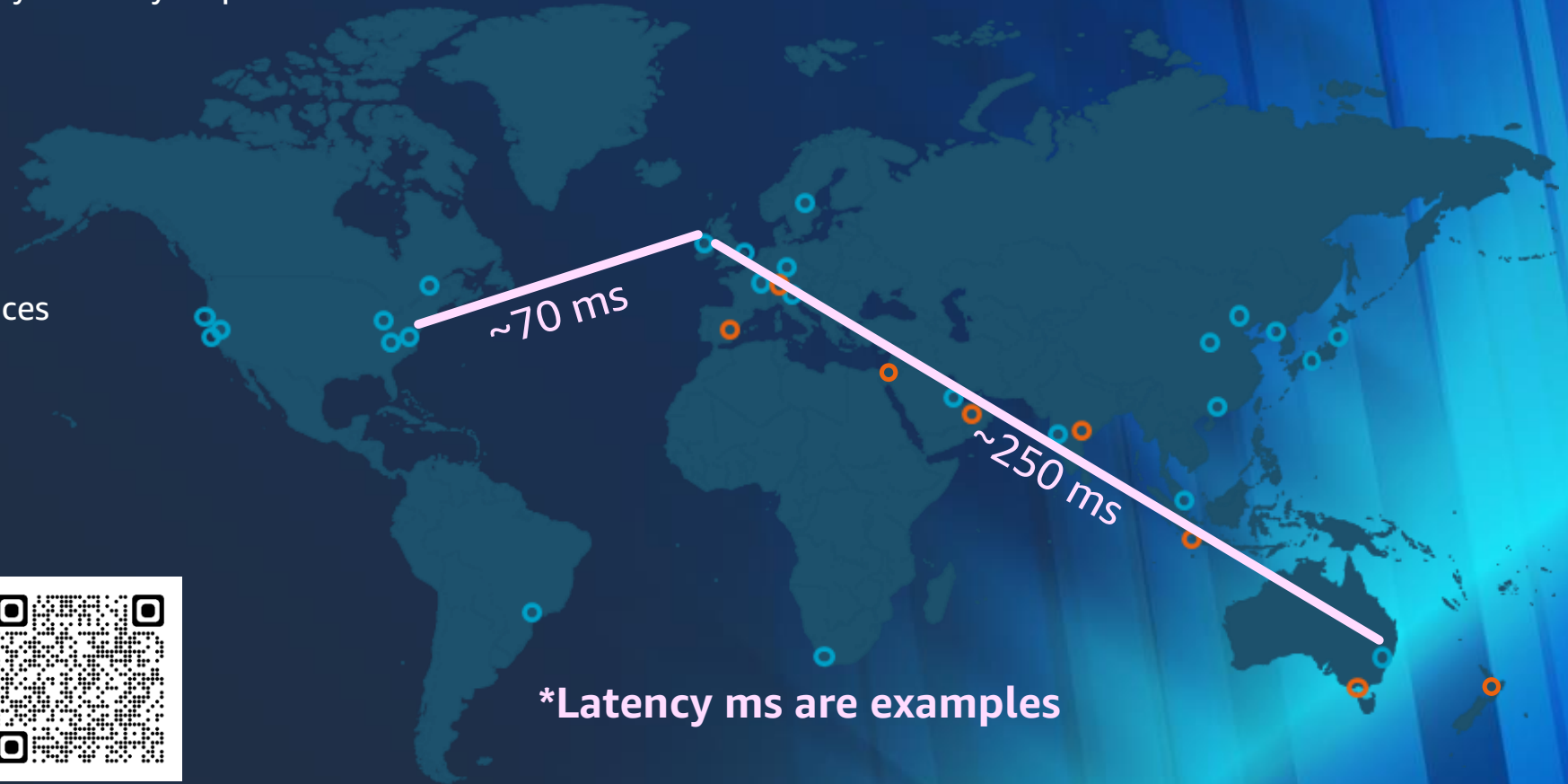
- Regional independence
- Understand dependencies and simplify recovery steps
- Minimize data replication
- Monitoring and observability
- Infrastructure as Code (IaC)
- Distributed system design best practices
- Test regularly



Blog : Implementing Multi-Region
Disaster Recovery

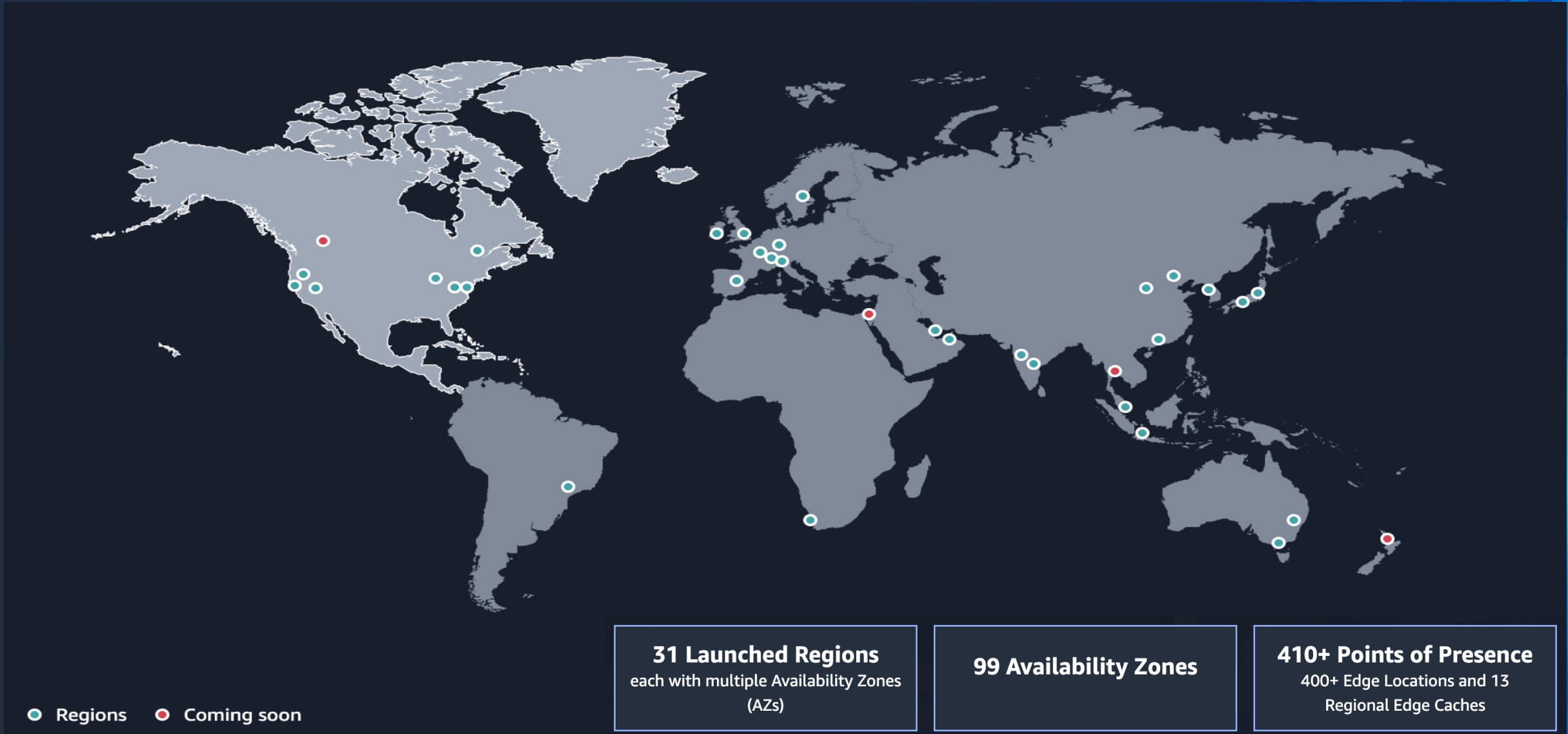


Video : Multi-Region design
patterns and best practices



***Latency ms are examples**

AWS Global Infrastructure



Scaling to 1M+ customers : Learn

Design

- Scaling from 1 to 1M+ customers is a journey
- Your journey is going to be unique to you
- Try not to be rigid about your architecture on Day 1
- Build Measure Learn
- Avoid undifferentiated heavy lifting

Scaling to 1M+ Customers : Best Practices check

AWS Well-Architected Framework



Operational
excellence



Security



Reliability



Performance
efficiency



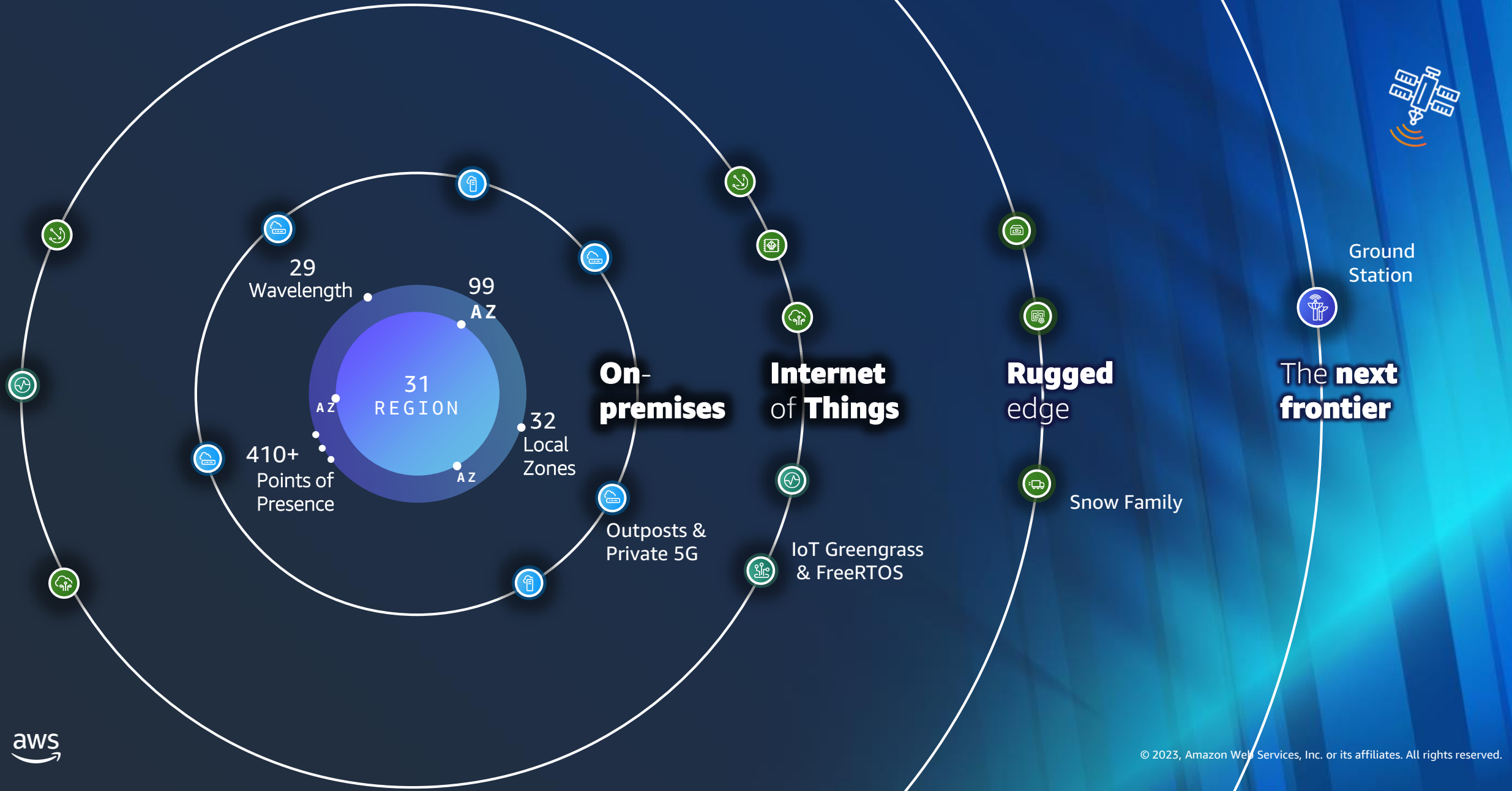
Cost
optimization



Sustainability

Scaling beyond a million customers

AWS Can Help – Going Everywhere



Summary

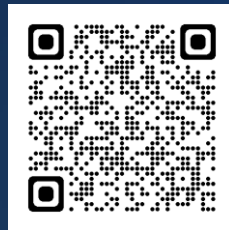
Takeaways

Scaling design

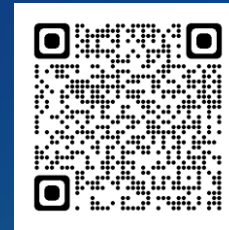
- Look for “best fit” technology based on your need
- Build, Measure, Learn
- Loose coupling, independently scale every layer
- Event-driven architecture, asynchronous processing
- SOA or independent application features
- Disaster recovery and business continuity
- Leverage AWS Partners, AWS Professional Services & AWS Prototyping team



AWS Workshops



AWS Solutions Library



AWS Free Tier

Scaling implementation

- Cache at edge, and within application
- Use purpose-built database
- Accelerate code development with Generative AI
- Automate CI/CD
- Use managed services, let AWS do the heavy lifting
- Use Amazon API Gateway to route traffic to application
- Focus on application features more than infrastructure

Visit the AWS Every App Resource hub

Start building upon a scalable, reliable, and globally available infrastructure so that you can focus on innovation and bringing new applications to market. Dive deeper with these resources today.

- Accelerate innovation with AWS
- Global-scale solutions
- Get more performance for your applications at lower costs with AWS
- Improving application delivery with AWS
- ... and more!



<https://tinyurl.com/for-every-app-hub-aws>

[Visit resource hub](#)



AWS Training and Certification

Access 600+ free digital courses with AWS Skill Builder

Focus on the cloud skills and services that are most relevant to you across 30+ AWS solutions, including digital self-paced learning plans and ramp-up guides

- Build your future in the AWS Cloud at your own pace
- Advance your skills and knowledge with learning plans
- Validate your cloud expertise with AWS Certification



LEARN YOUR WAY [SKILLBUILDER.AWS](https://skillbuilder.aws)



Working with AWS Partner Network

AWS Partners are uniquely positioned to help your organization at any stage of your cloud adoption journey.

- **Innovation** – innovative and cost-effective scalable cloud solutions and capabilities for your organization - helping you keep pace with cutting edge technology changes.
- **Expertise** – strategic experts and experienced builders, providing groundbreaking, relevant and reliable solutions to help your business grow.
- **Global Reach** - Choose from the global community of trusted AWS Partners across software, hardware, and services.



[Find and connect with AWS Partners](#)



[Explore partner offerings on AWS Marketplace](#)

Thank you for attending AWS Innovate – Every App 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!