Content and Network Delivery Services

David Tucker
TECHNICAL ARCHITECT & CTO CONSULTANT
@_davidtucker_ davidtucker.net

Networking & Content Delivery Services

Amazon Route 53

Amazon VPC

AWS Direct Connect

Amazon API Gateway

Amazon CloudFront

Elastic Load Balancing

Overview

Introducing Virtual Private Clouds on AWS

Understanding the purpose of AWS Direct Connect

Examining DNS with Amazon Route 53

Reviewing Amazon CloudFront

Reviewing API Gateway

Introducing Elastic Load Balancing and scaling approaches

Amazon VPC and Direct Connect

Amazon Virtual Private Cloud (VPC)

A logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define.

Amazon Virtual Private Cloud (VPC)

Enables virtual networks in AWS

Supports IPv4 and IPv6

Allows for configuration of

- IP address range
- Subnets
- Route tables
- Network gateways

Supports public & private subnets

Can utilize NAT for private subnets

Enables a connection to your data center

Can connect to other VPC’s

Supports private connections to many AWS services

AWS Direct Connect

A cloud service solution that makes it easy to establish a dedicated network connection from your data center to AWS.

Amazon Route 53

Amazon Route 53

Domain name service (DNS)

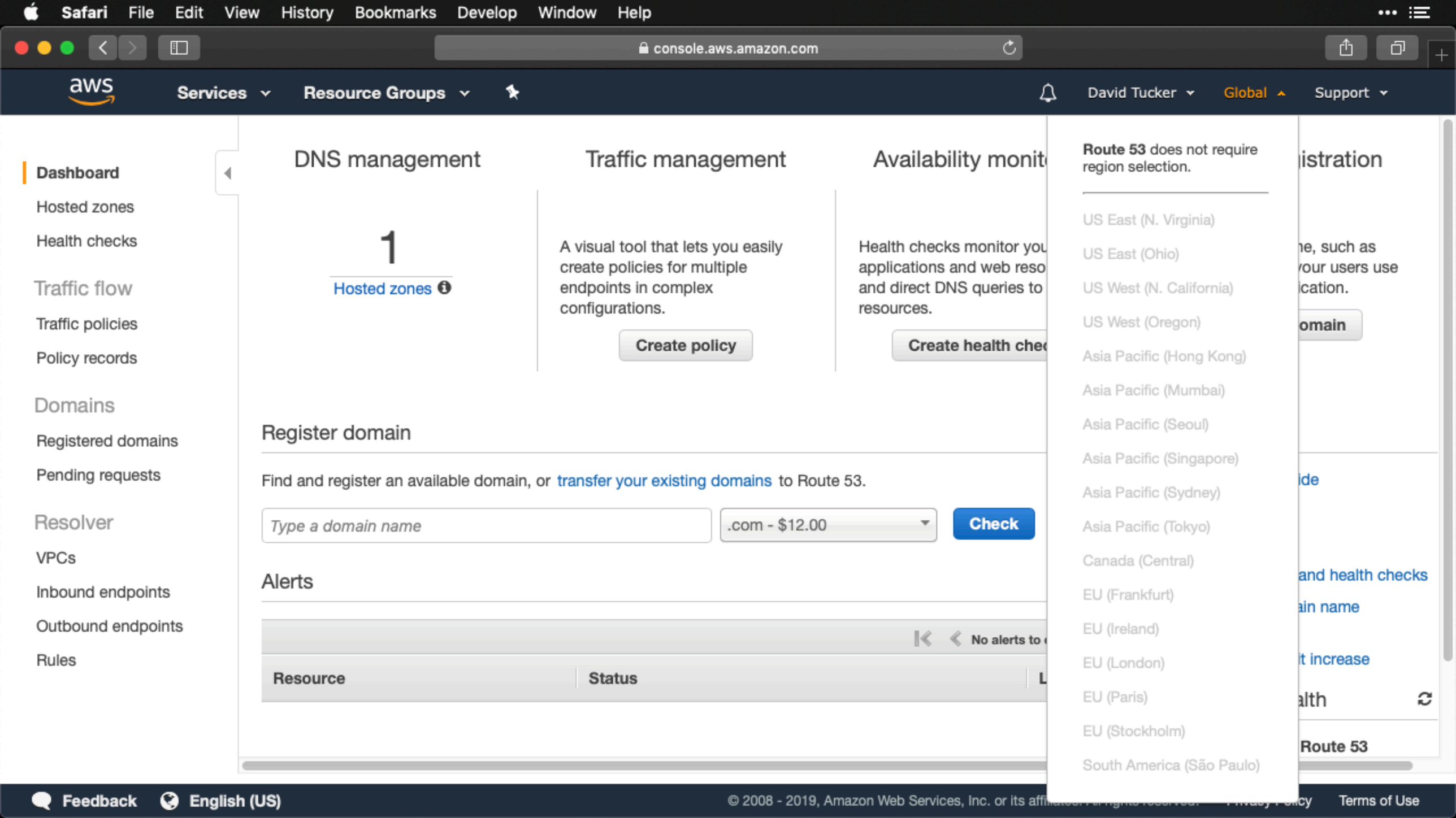
Global AWS service (not regional)

Highly available

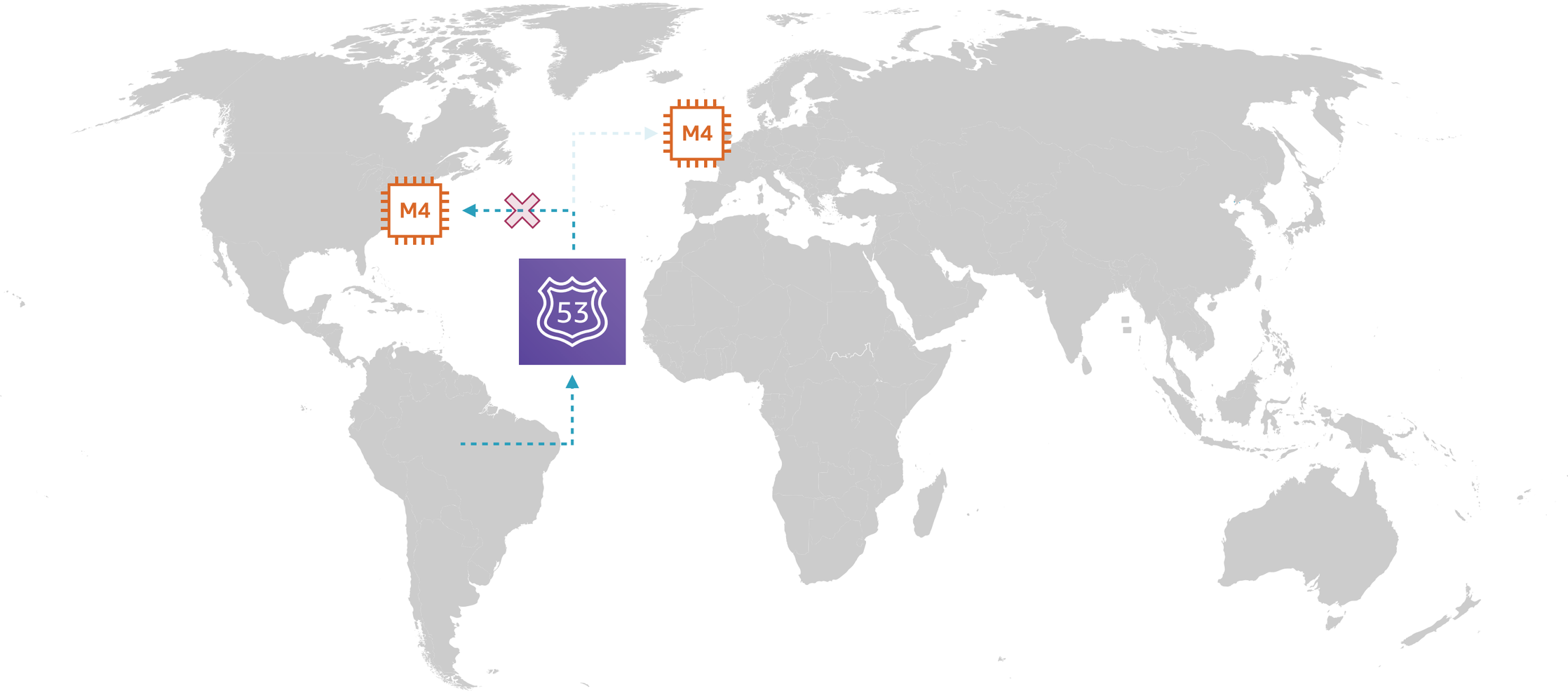
Enables global resource routing

“**DNS** translates more readily memorized domain names to the numerical IP addresses needed for locating and identifying computer services and devices with the underlying network protocols.”

Wikipedia


Safari File Edit View History Bookmarks Develop Window Help
console.aws.amazon.com
aws
Services
Resource Groups
David Tucker
Global
Support
Dashboard
Hosted zones
Health checks
Traffic flow
Traffic policies
Policy records
Domains
Registered domains
Pending requests
Resolver
VPCs
Inbound endpoints
Outbound endpoints
Rules
DNS management
1
Hosted zones
Traffic management
A visual tool that lets you easily create policies for multiple endpoints in complex configurations.
Create policy
Availability monit
Health checks monitor you
applications and web reso
and direct DNS queries to
resources.
Create health chec
istration
ne, such as
your users use
ication.
omain
Register domain
Find and register an available domain, or transfer your existing domains to Route 53.
Type a domain name
.com - $12.00
Check
Alerts
No alerts to
Resource
Status
ide
and health checks
ain name
t increase
alth
Route 53
Route 53 does not require region selection.
US East (N. Virginia)
US East (Ohio)
US West (N. California)
US West (Oregon)
Asia Pacific (Hong Kong)
Asia Pacific (Mumbai)
Asia Pacific (Seoul)
Asia Pacific (Singapore)
Asia Pacific (Sydney)
Asia Pacific (Tokyo)
Canada (Central)
EU (Frankfurt)
EU (Ireland)
EU (London)
EU (Paris)
EU (Stockholm)
South America (São Paulo)
Feedback
English (US)
© 2008 - 2019, Amazon Web Services, Inc. or its affi
Terms of Use

Route 53 High Availability



Elastic Load Balancing

Elasticity

The ability for the infrastructure supporting an application to grow and contract based on how much it is used at a point in time.

Elastic Load Balancing

Distributes traffic across multiple targets

Integrates with EC2, ECS, and Lambda

Supports one or more AZ's in a region

Three types of load balancers

- Application Load Balancer (ALB)
- Network Load Balancer (NLB)
- Classic Load Balancer

Scaling on Amazon EC2

Vertical Scaling

You “scale up” your instance type to a larger instance type with additional resources

Horizontal Scaling

You “scale out” and add additional instances to handle the demand of your application

Amazon CloudFront and API Gateway

Amazon CloudFront

Content delivery network (CDN)

Enables users to get content from server closest to them

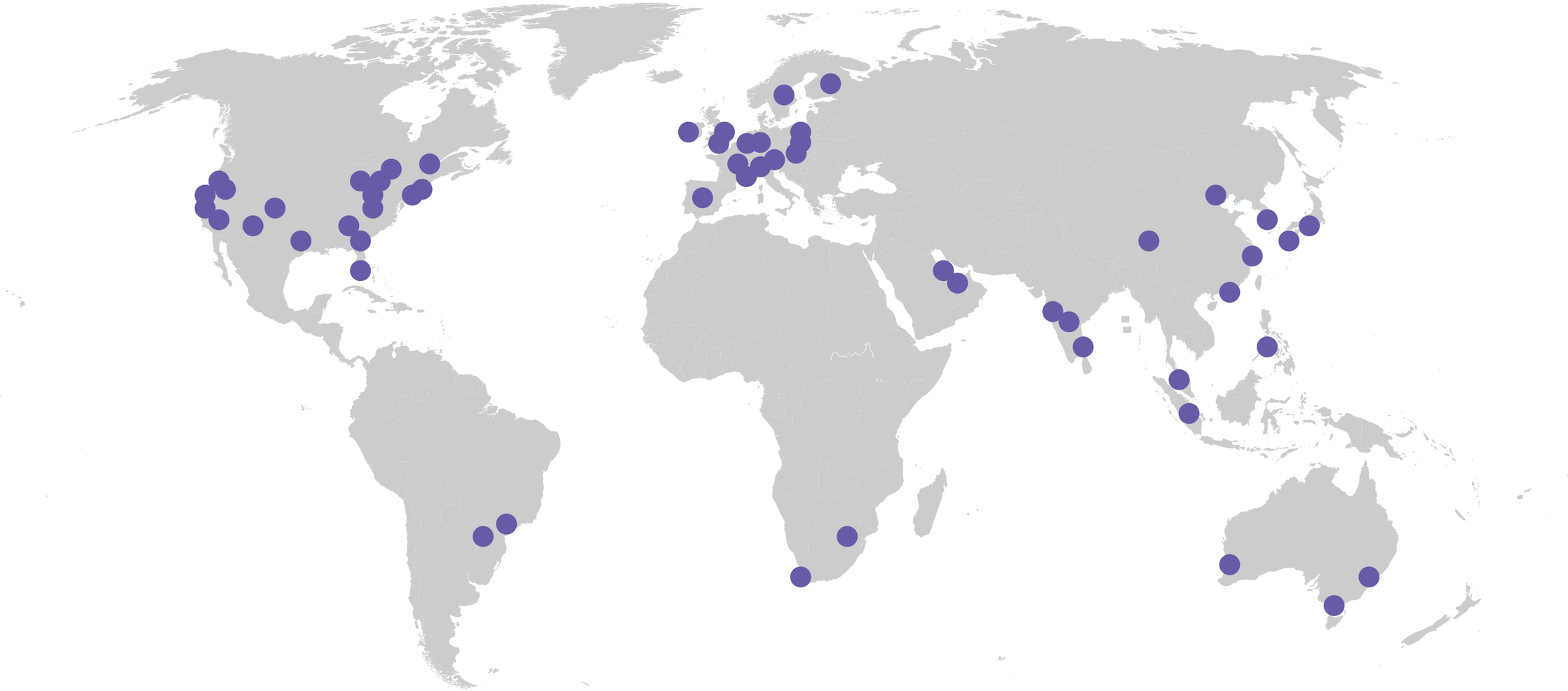
Supports static and dynamic content

Utilizes AWS edge locations

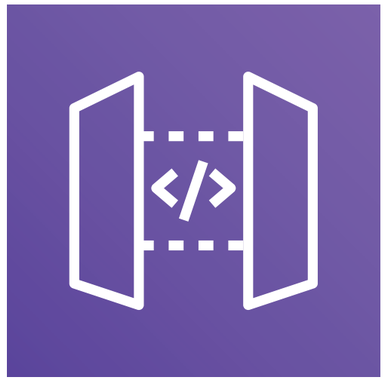
Includes advanced security features

- AWS Shield for DDoS
- AWS WAF

AWS Edge Locations



Amazon API Gateway



- **Fully managed API management service**
- **Directly integrates with multiple AWS services**
- **Provides monitoring & metrics on API calls**
- **Supports VPC and on-premise private applications**

AWS Global Accelerator

“**AWS Global Accelerator** is a networking service that sends your user’s traffic through Amazon Web Service’s global network infrastructure, improving your internet user performance by up to 60%.”

Amazon Web Services

AWS Global Accelerator

Utilizes IP addresses that route to edge locations

Once request reaches edge locations, traffic is routed through AWS network

Can route requests to many AWS resources:

- Network Load Balancer (NLB)
- Application Load Balancer (ALB)
- EC2 Instances
- Elastic IP address

Performance Improvements

Distance between user and initial endpoint is minimized by using edge locations

Traffic is optimized by using AWS network instead of public Internet

Results in improvement of first byte latency, jitter, and throughput

Provides superior fault tolerance by not relying on DNS resolution

AWS Global Accelerator Use Cases

Non-HTTP Protocol

If you are using UDP, MQTT, or VOIP for your solution

Requires Static IP

If you solution needs IP (and not DNS) resolution

Instant Failover

If you need the best approach for failover and fault-tolerance

Scenario Based Review

Scenario 1

Jane's company maintains two corporate data centers

They want their data centers to work alongside AWS for specific workloads

She is wondering if there is a way to have a persistent connection to AWS

What service from AWS would you recommend her company implement?

Scenario 2

Tim's company serves content through their site to users around the globe

They are looking to optimize performance to users around the world

They want to leverage a Content Delivery Network (CDN)

Which service would enable optimized performance globally for their content?

Scenario 3

Ellen's company has an internal application that runs on an EC2 server

Currently there is downtime as demand is greater than capacity for the server

Ellen is trying to decide if she should use bigger servers or more servers

Which scaling approach would you recommend and what services should they use?

Summary

Overview

Introduced Virtual Private Clouds on AWS

Understood the purpose of AWS Direct Connect

Examined DNS with Amazon Route 53

Reviewed Amazon CloudFront

Reviewed API Gateway

Introduced Elastic Load Balancing and scaling approaches

Scenario 1

Jane's company maintains two corporate data centers

They want their data centers to work alongside AWS for specific workloads

She is wondering if there is a way to have a persistent connection to AWS

What service from AWS would you recommend her company implement?

Solution: AWS Direct Connect

Scenario 2

Tim's company serves content through their site to users around the globe

They are looking to optimize performance to users around the world

They want to leverage a Content Delivery Network (CDN)

Which service would enable optimized performance globally for their content?

Solution: Amazon CloudFront

Scenario 3

Ellen's company has an internal application that runs on an EC2 server

Currently there is downtime as demand is greater than capacity for the server

Ellen is trying to decide if she should use bigger servers or more servers

Which scaling approach would you recommend and what services should they use?

Solution: Horizontal Scaling using Elastic Load Balancing