

# Going Global with Serverless!

Adrian Hornsby, Cloud Architecture Evangelist





https://xkcd.com/1428/

#### JOBS I'VE BEEN FIRED FROM

FEDEX DRIVER
CRANE OPERATOR
SURGEON
AIR TRAFFIC CONTROLLER
PHARMACIST
MUSEUM CURATOR
WAITER
DOG WALKER
OIL TANKER CAPTAIN
VIOLINIST
MARS ROVER DRIVER
MASSAGE THERAPIST



11

Failures are a given and everything will eventually fail over time.

Werner Vogels
CTO – Amazon.com





#### Netflix 2013

Applause from Iramesh and 33 others



#### Netflix Technology Blog

Learn more about how Netflix designs, builds, and operates our systems and engineering organizations

#### Active-Active for Multi-Regional Resiliency

by Ruslan Meshenberg, Naresh Gopalani, and Luke Kosewski

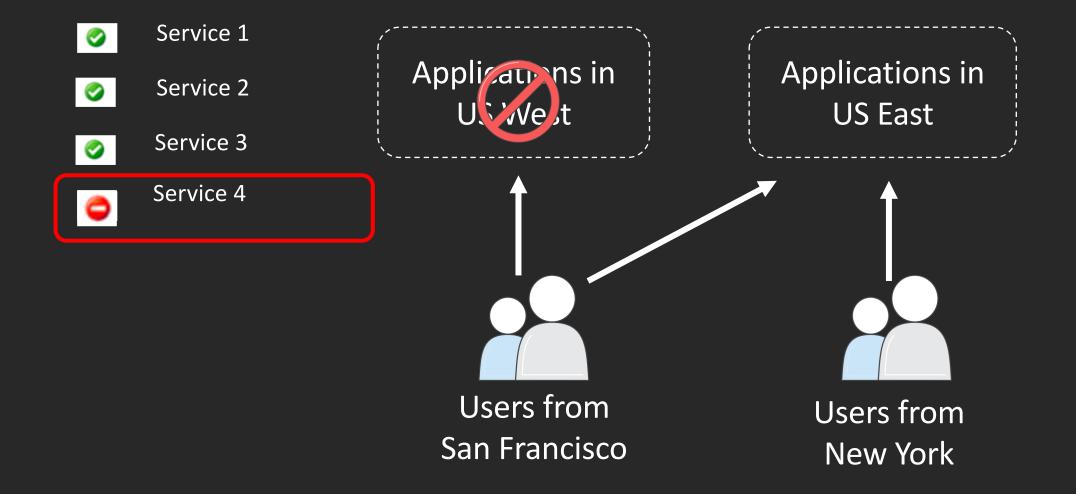
In June, we talked about Isthmus—our approach to achieve resiliency against region-wide ELB outage. After completing the Isthmus project we embarked on the next step in our quest for higher resiliency and availability—a full multi-regional Active-Active solution. This project is now complete, and Netflix is running Active-Active across the USA, so this post aims to highlight some of the interesting challenges and learnings we found along the way.

#### Failure — function of scale and speed.

In general, failure rates an organization is dealing with depend largely on 2 factors: scale of operational deployment and velocity of change. If both scale and speed are small, then most of the time things just work. Once scale starts to grow, even with slow velocity, the chance of hardware failure will increase. Conversely, even at small scale, if velocity is fast enough, chance of software failure will increase. Ultimately, if you're running at scale and still pursuing high velocity—things will break all the time.



#### Disaster recovery





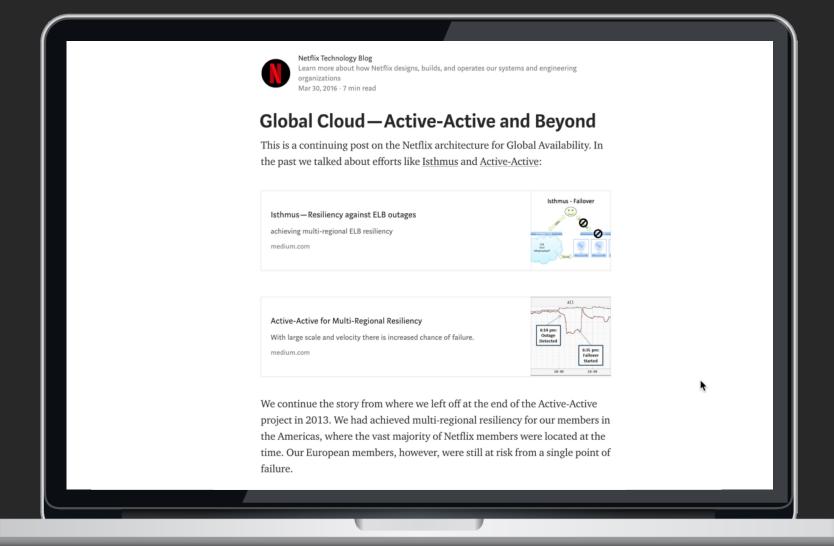




Service 4

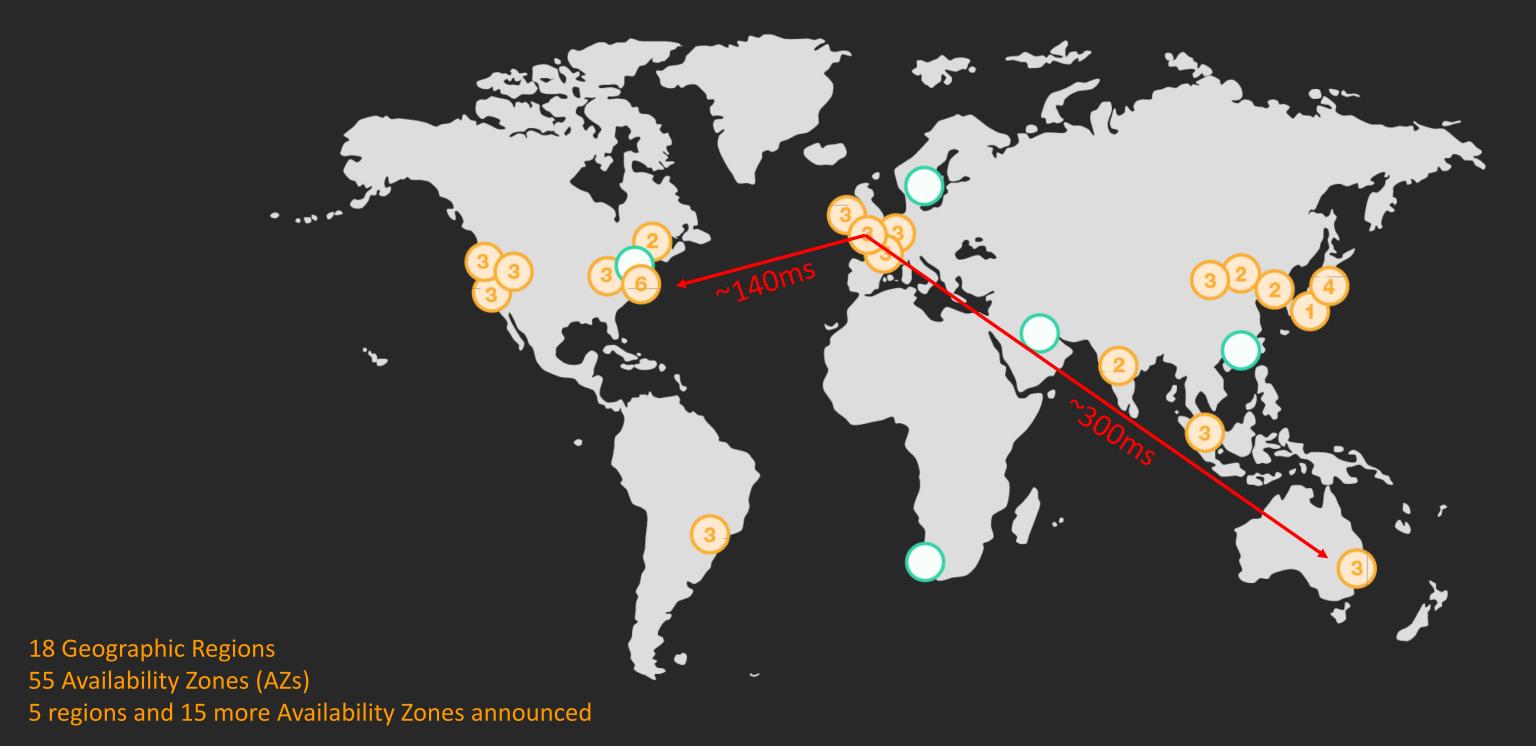


#### Netflix 2016

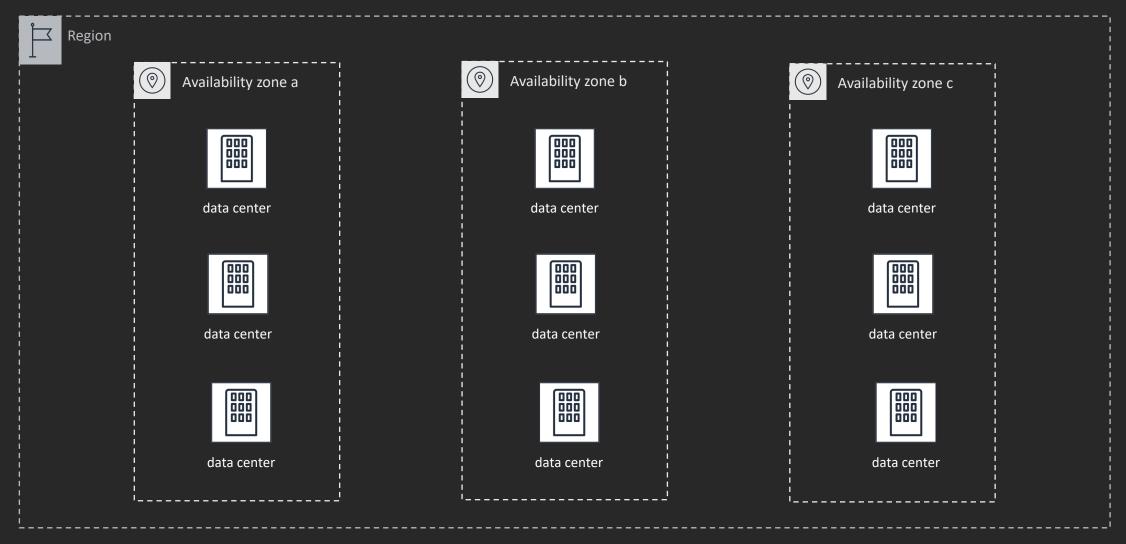




# Improve latency for end-users



# AWS Region and availability zones

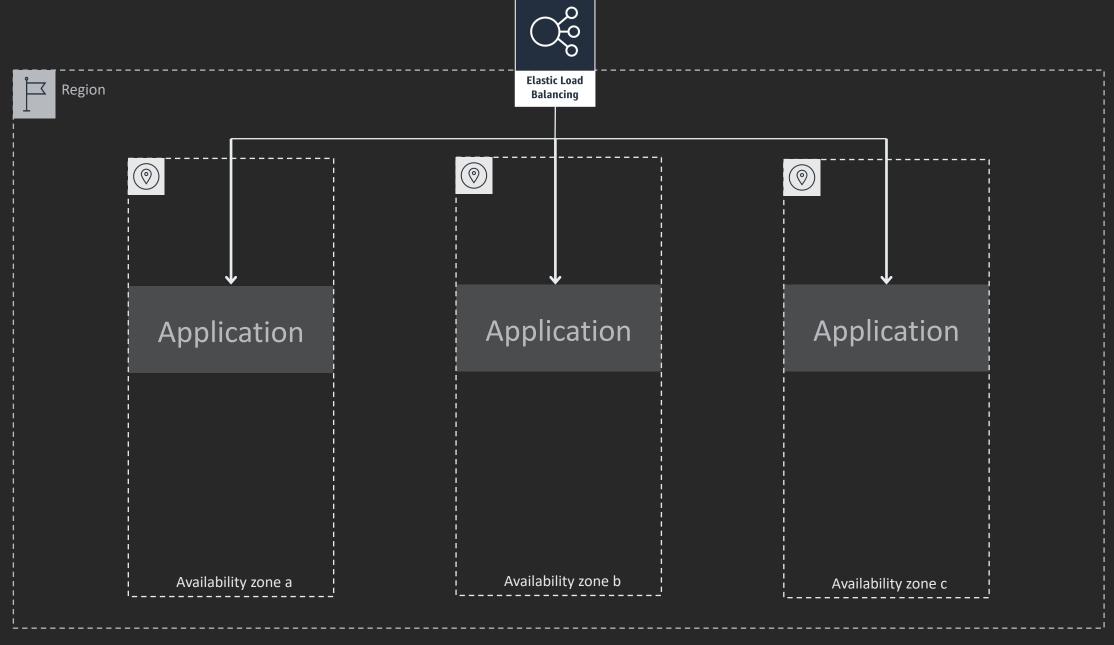


<sup>\*</sup> n data center per AZ



<sup>\*\*</sup> AZ per region: 2 to 6

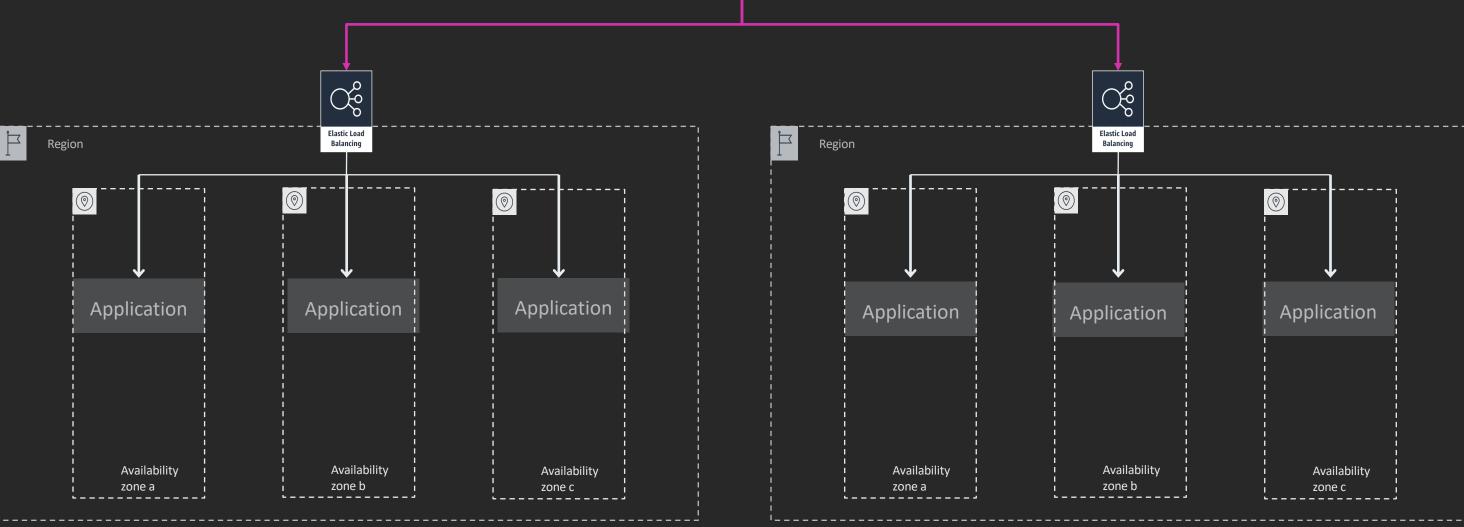
#### Multi-AZ architecture





### Multi-region







#### Why serverless components??



No servers to provision or manage



Scales with usage



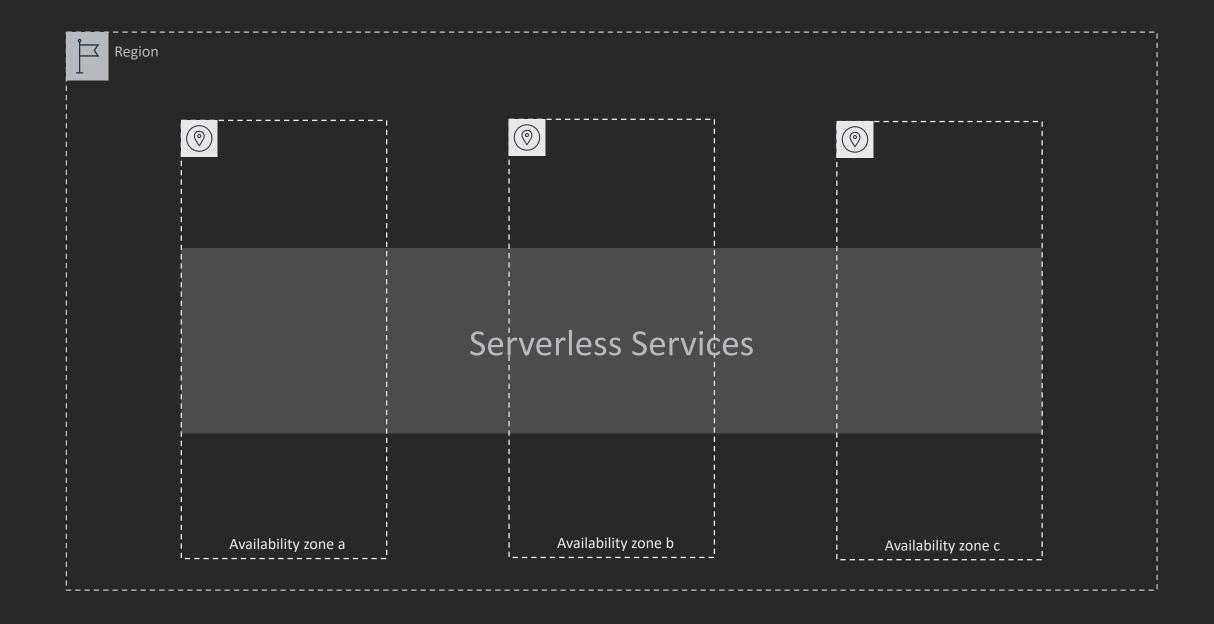
Never pay for idle



Availability and fault tolerance built in



# Serverless components





# Prerequisits to building a multi-region architecture.

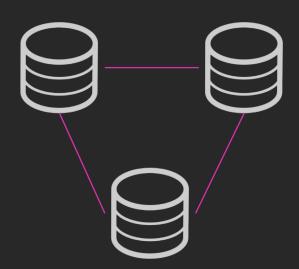


#### **CAP Theorem**

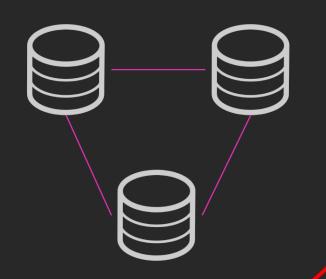
In the presence of a network partition, you must choose between consistency and availability!

#### Consistency

Data is consistent. All nodes see the same state.



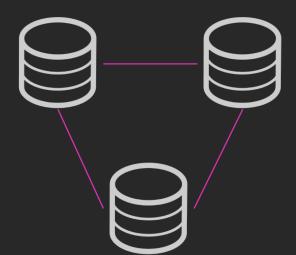
# Availability Every request is non-failing.



#### **Distributed System**

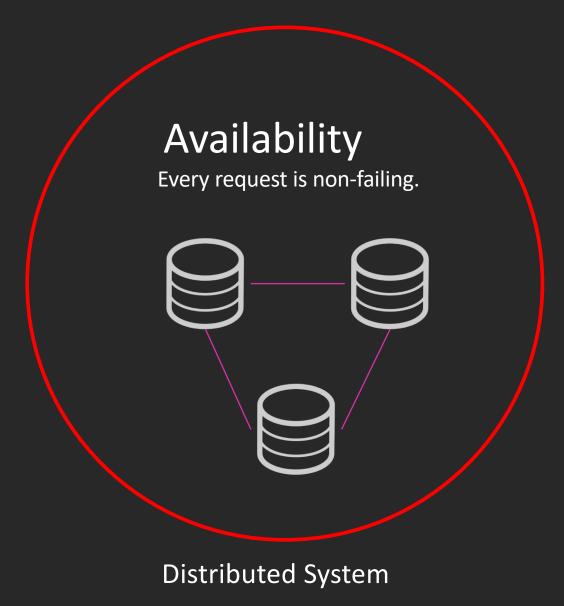
#### Partition Tolerance

Service still responds as expected if some nodes crash.





#### Embrase eventual consistency



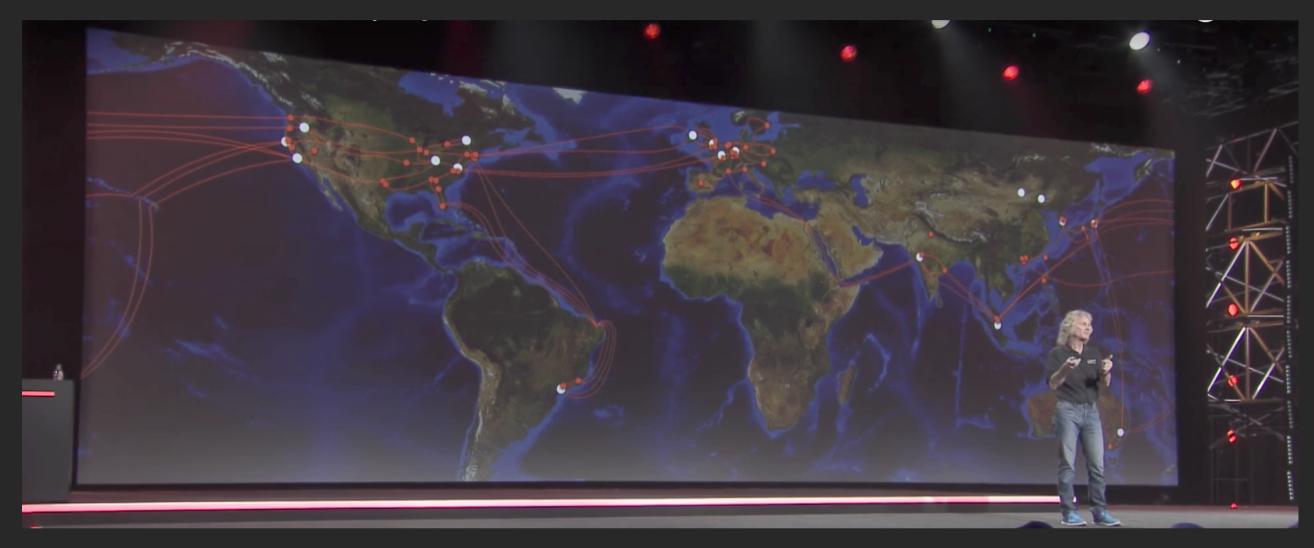
... if no new updates are made to a given data item, eventually all accesses to that item will return the last updated value.

https://en.wikipedia.org/wiki/Eventual\_consistency

An eventually consistent system can return any value before it converges!!

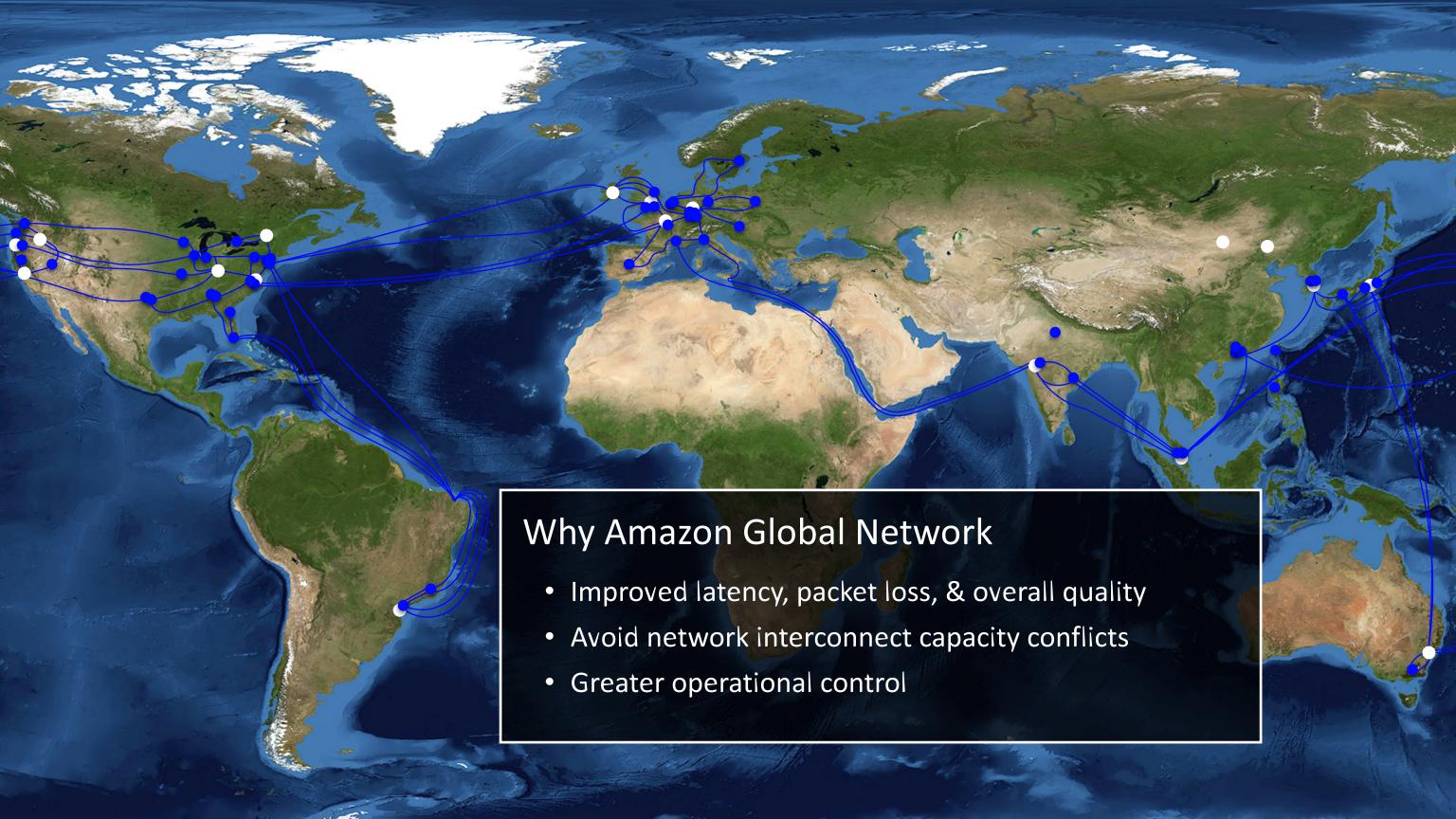


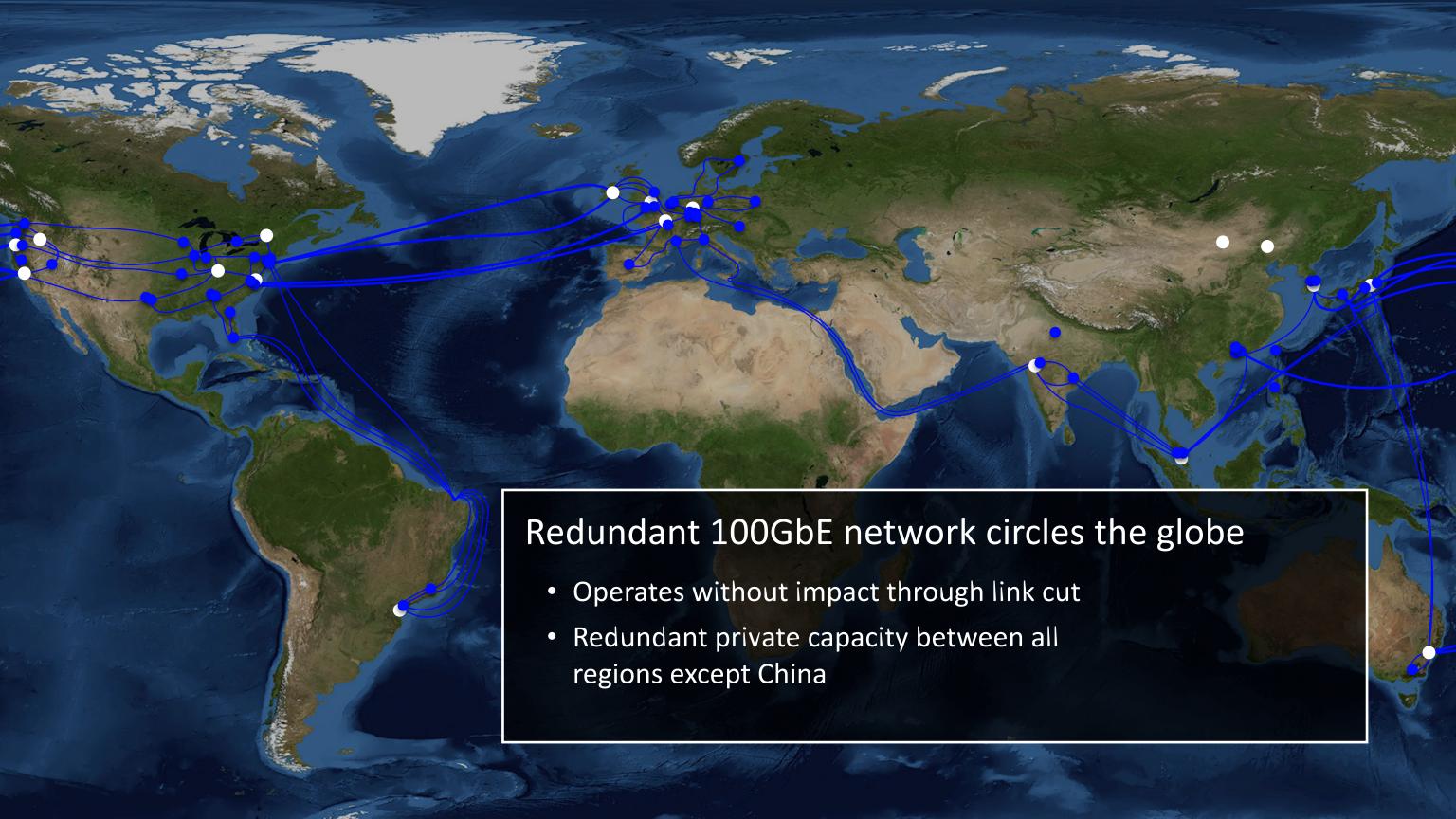
# Secure and reliable global network

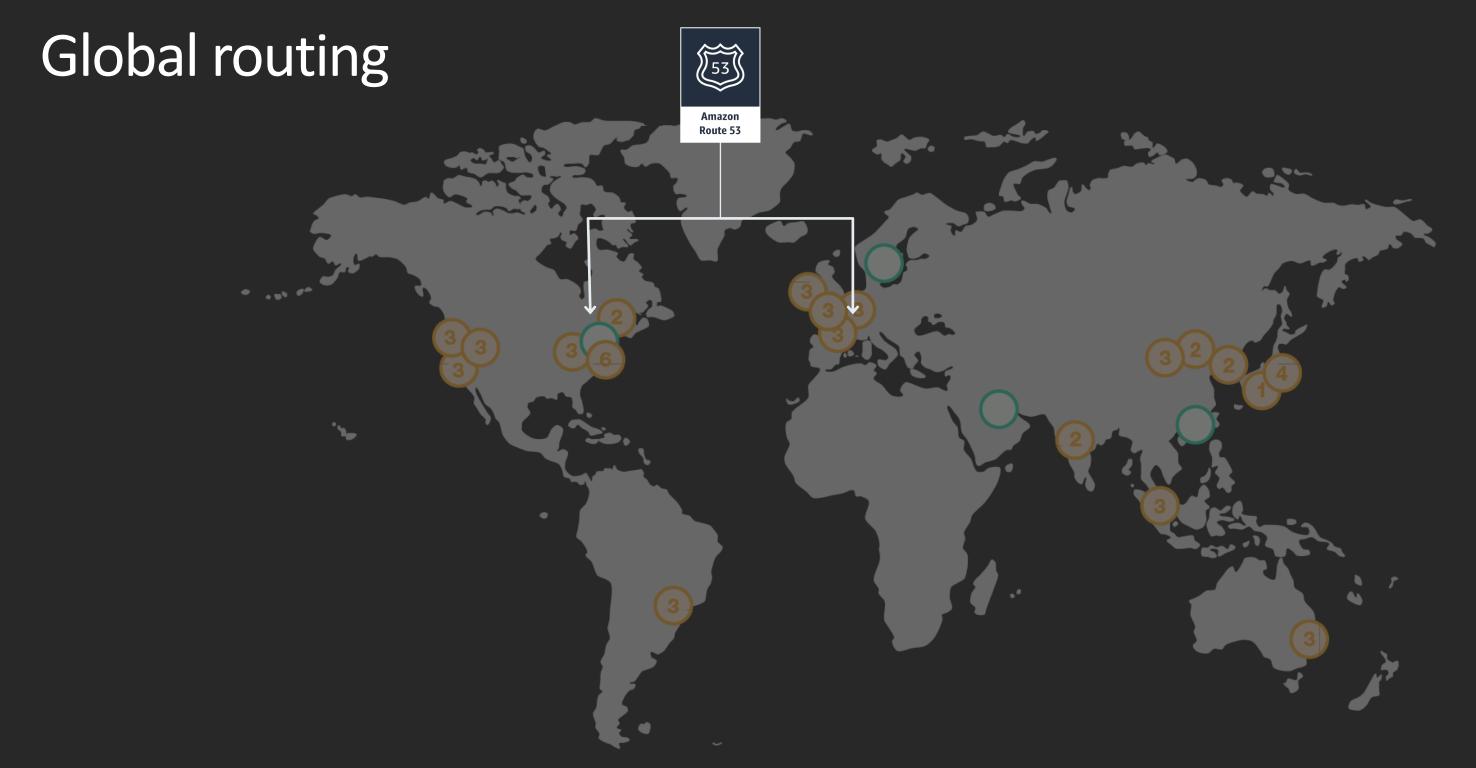


James Hamilton – reinvent 2016



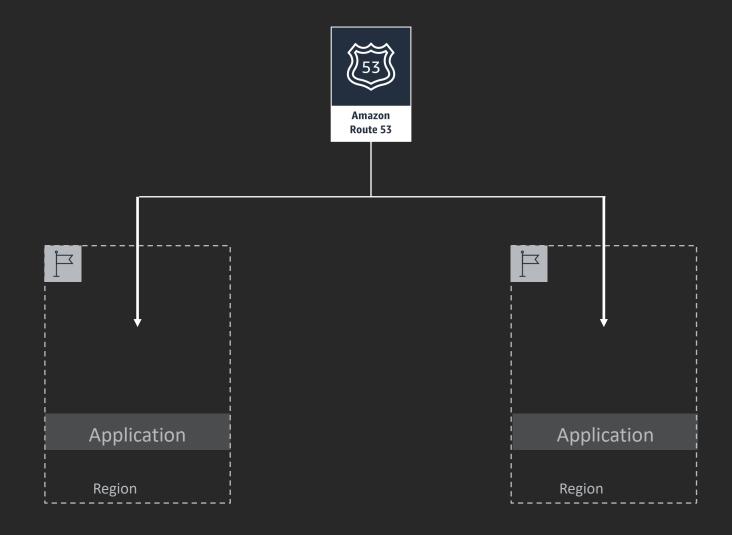






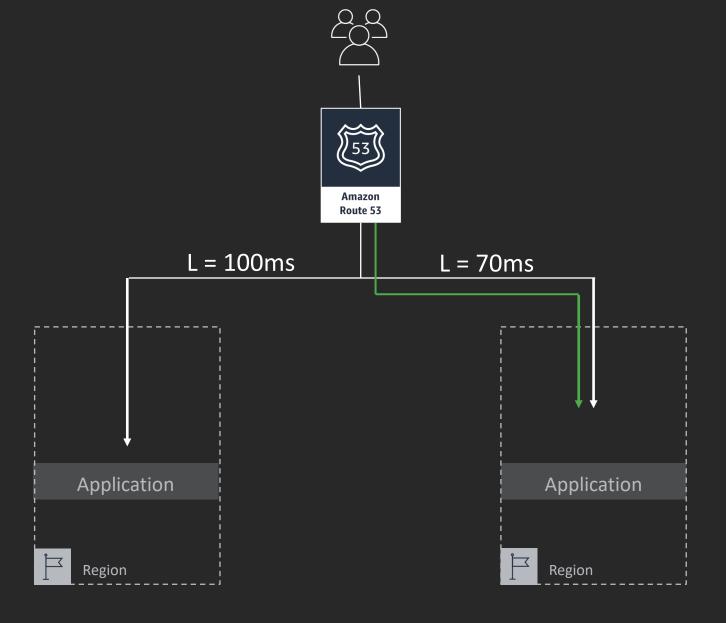


### Routing policies with Route 53



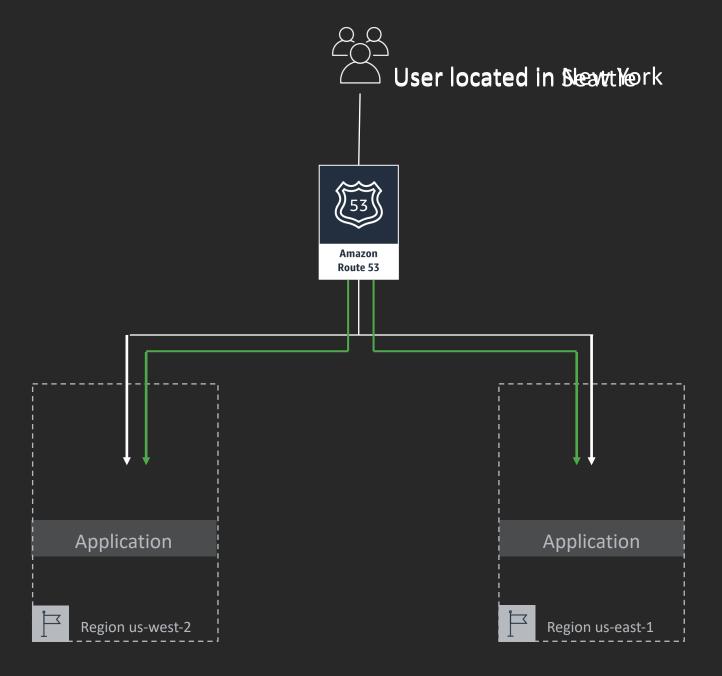


### Latency based routing





### Geo-based routing





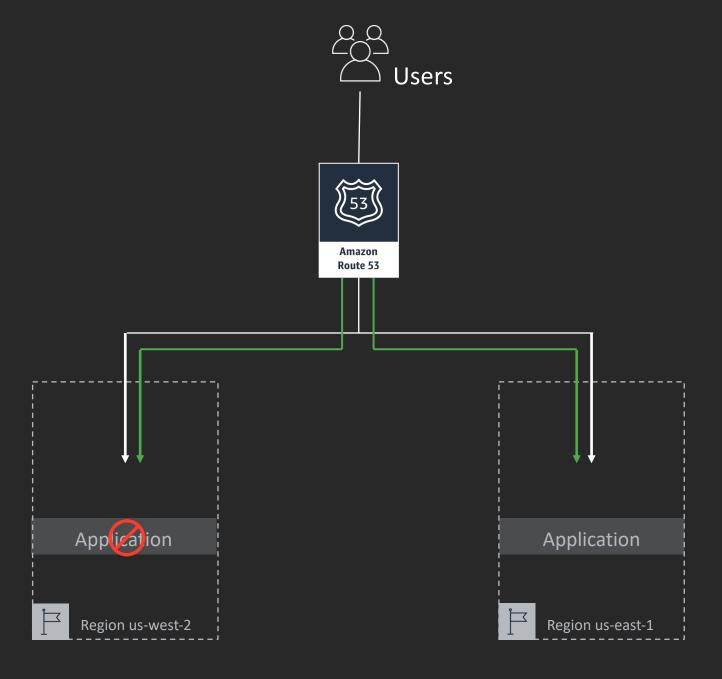
# Weighted round robin routing User located in Seattle Route 53 Application Application



Region us-east-1

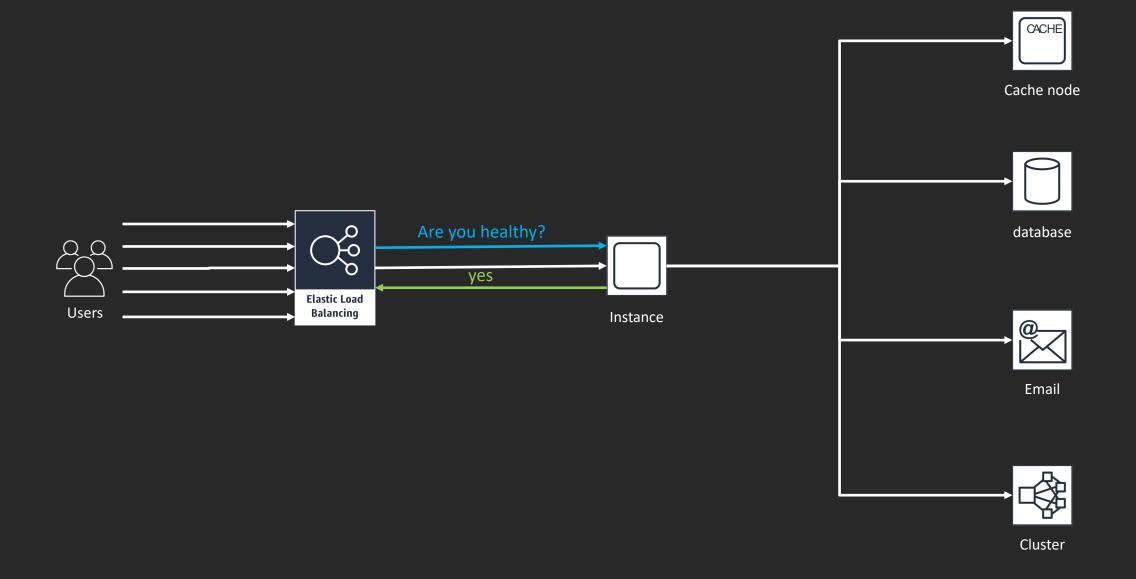
Region us-west-2

### DNS failover



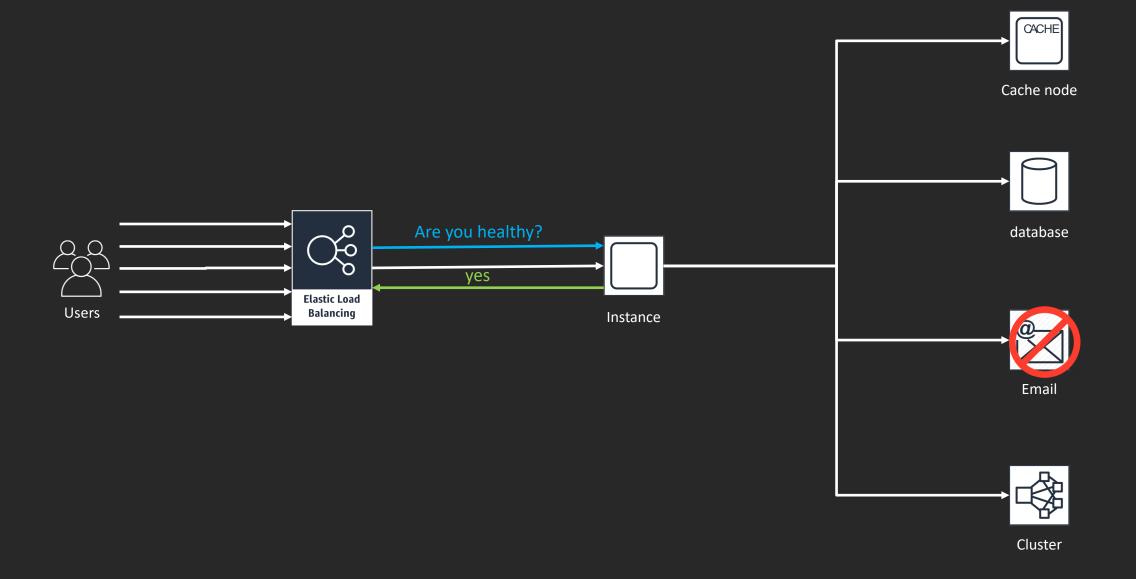


#### Shallow health check



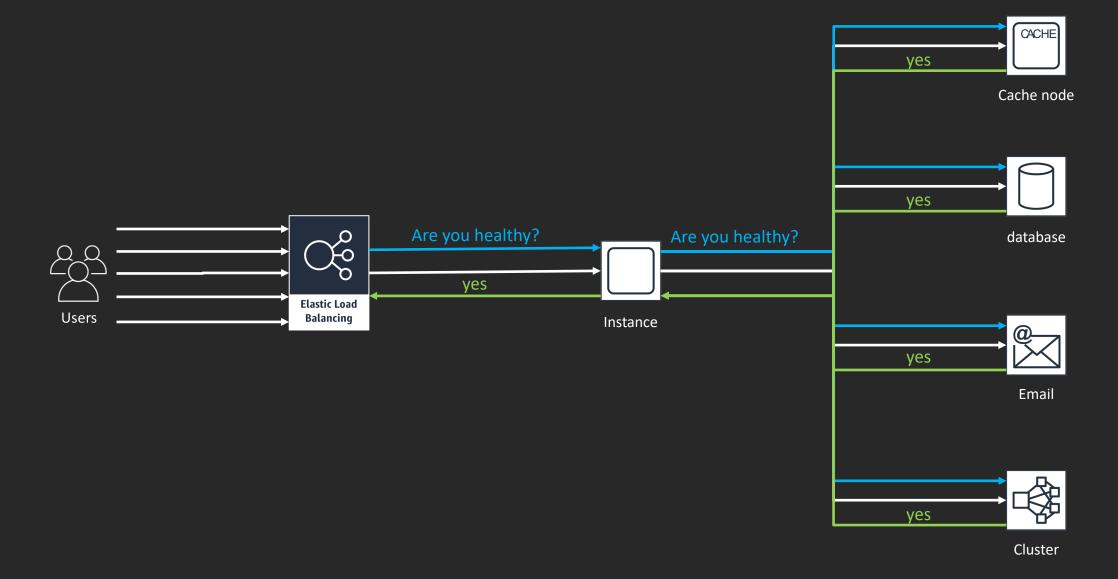


#### Shallow health check



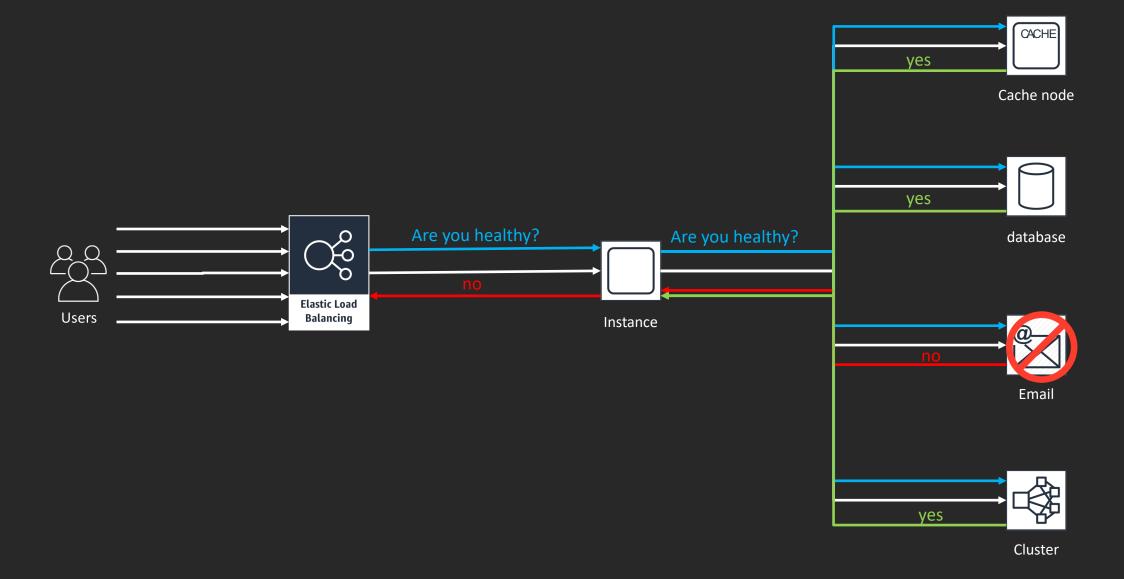


# Deep health check



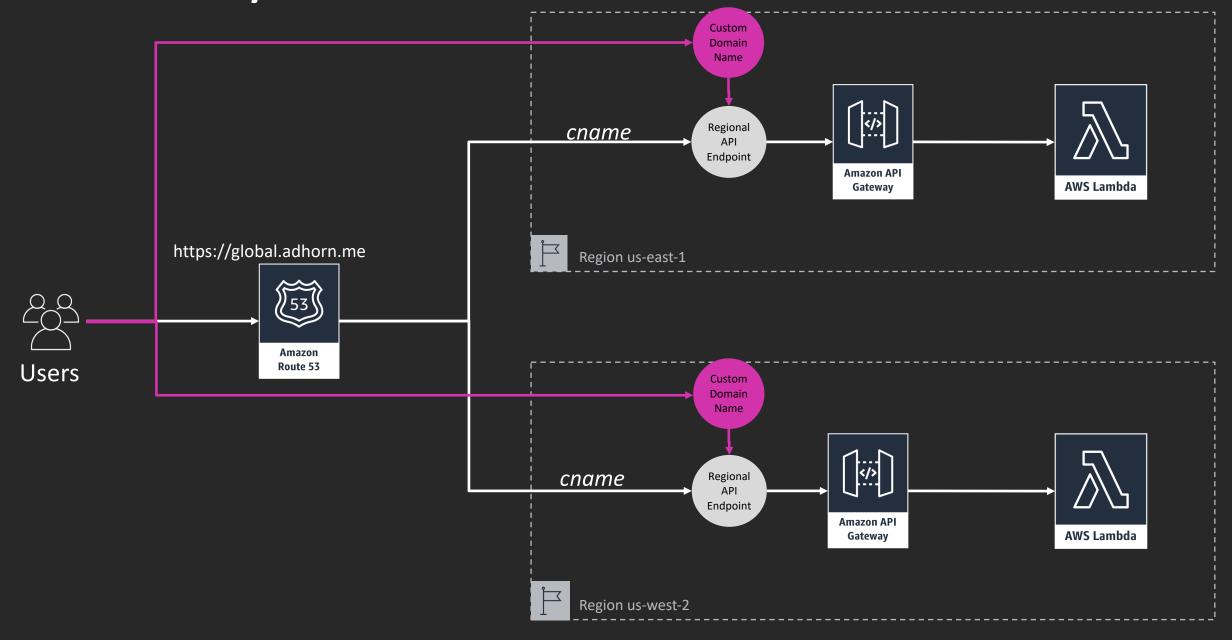


# Deep health check





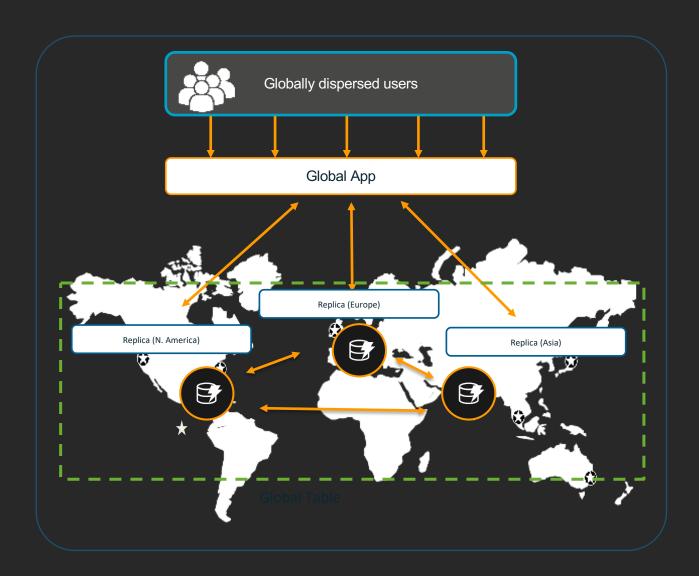
### **API** Gateway





#### Amazon DynamoDB Global Tables

#### Fully managed, multi-master, multi-region database



Build high performance, globally distributed applications

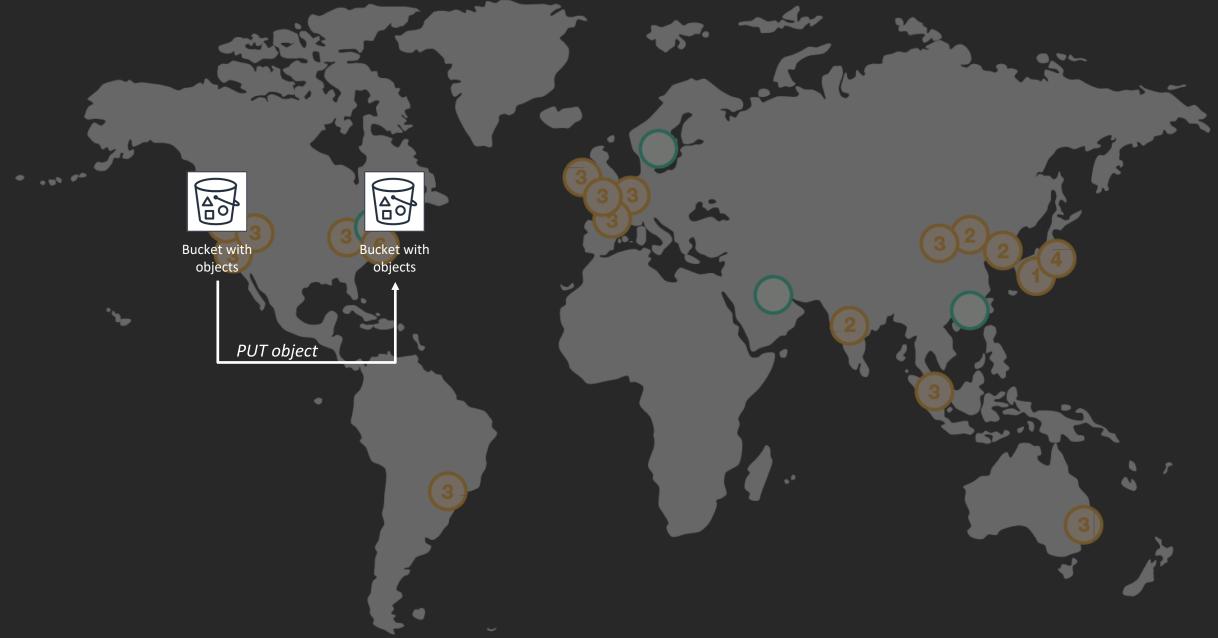
Low latency reads & writes to locally available tables

Disaster proof with multi-region redundancy

Easy to set up and no application rewrites required

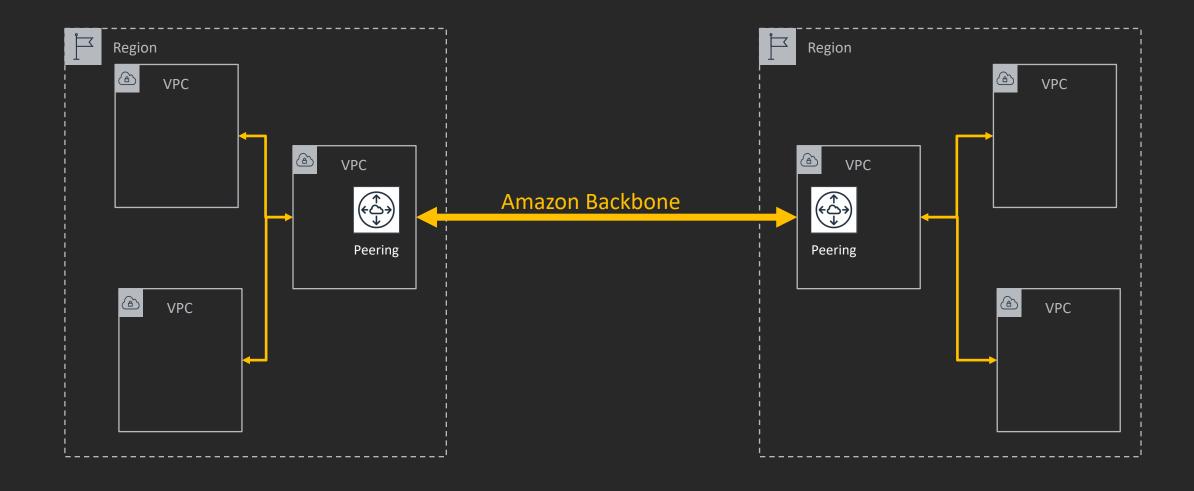


# Amazon S3 cross-region replication





# Multi-region multi-vpc connectivity



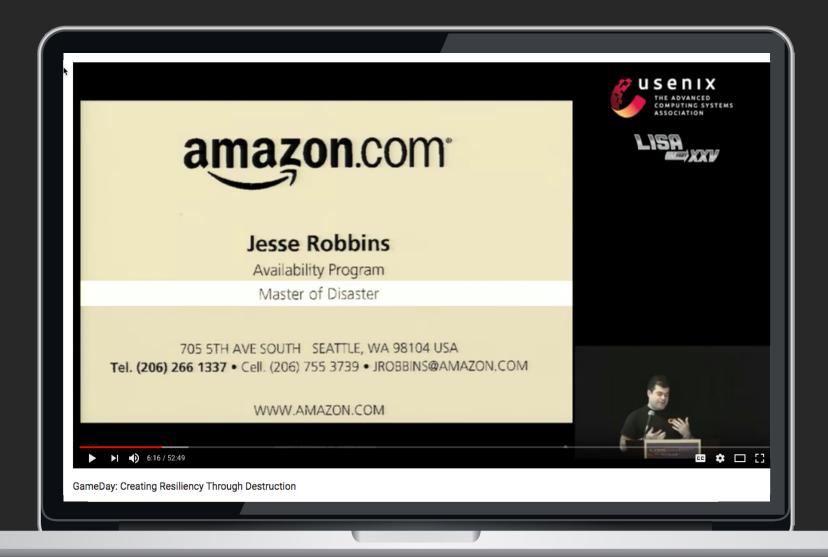
https://aws.amazon.com/answers/networking/aws-multiple-region-multi-vpc-connectivity/



# Testing multi-region architecture.



#### GameDay at Amazon



https://www.youtube.com/watch?v=zoz0ZjfrQ9s



# Chaos engineering

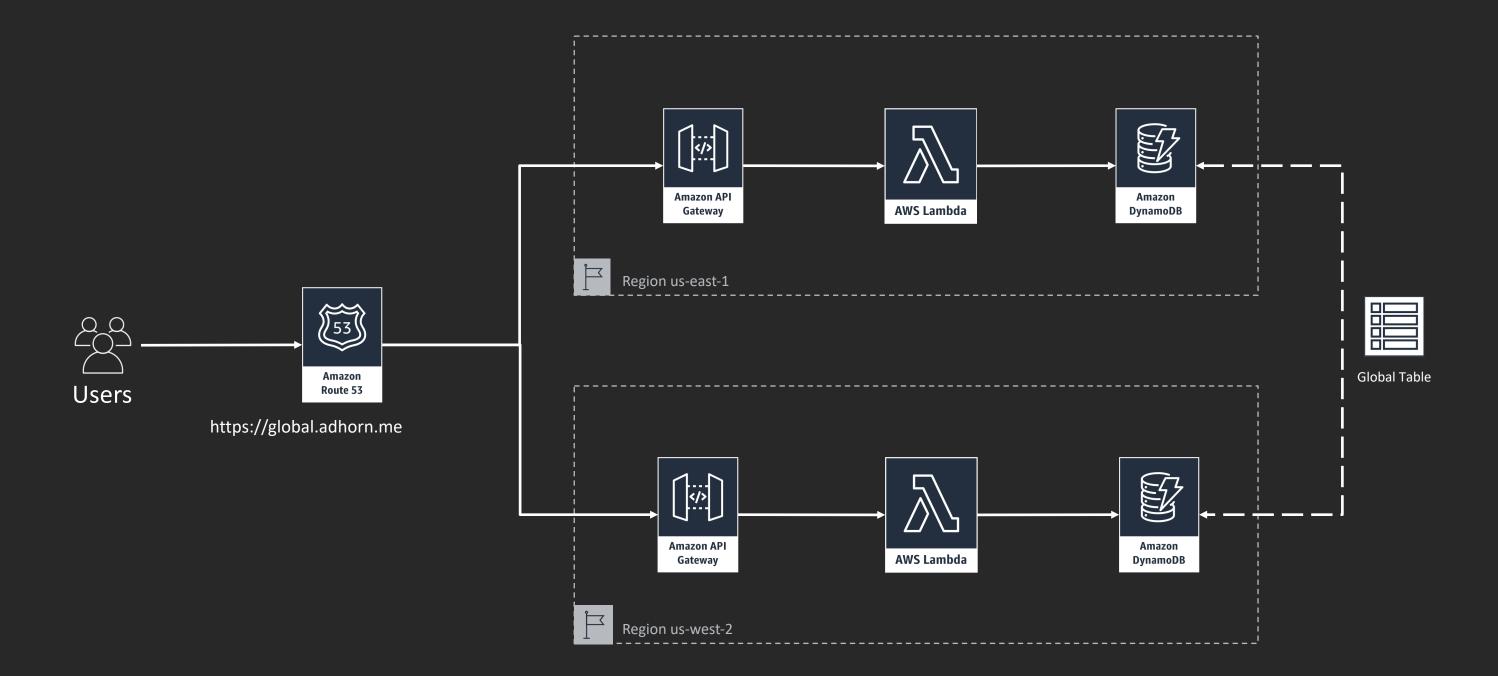


https://github.com/Netflix/SimianArmy

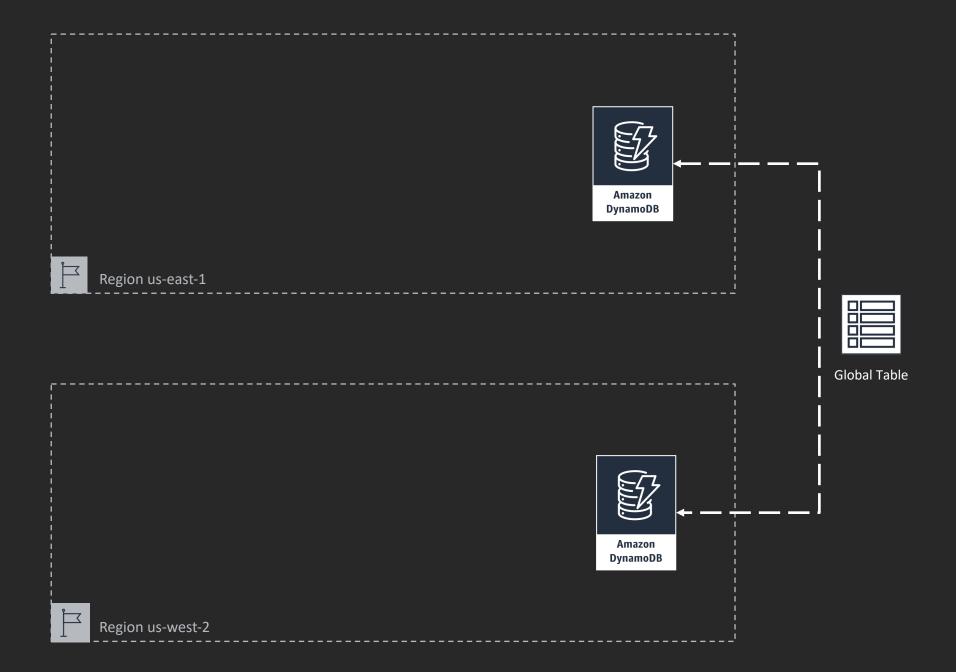


# Hands-on demo

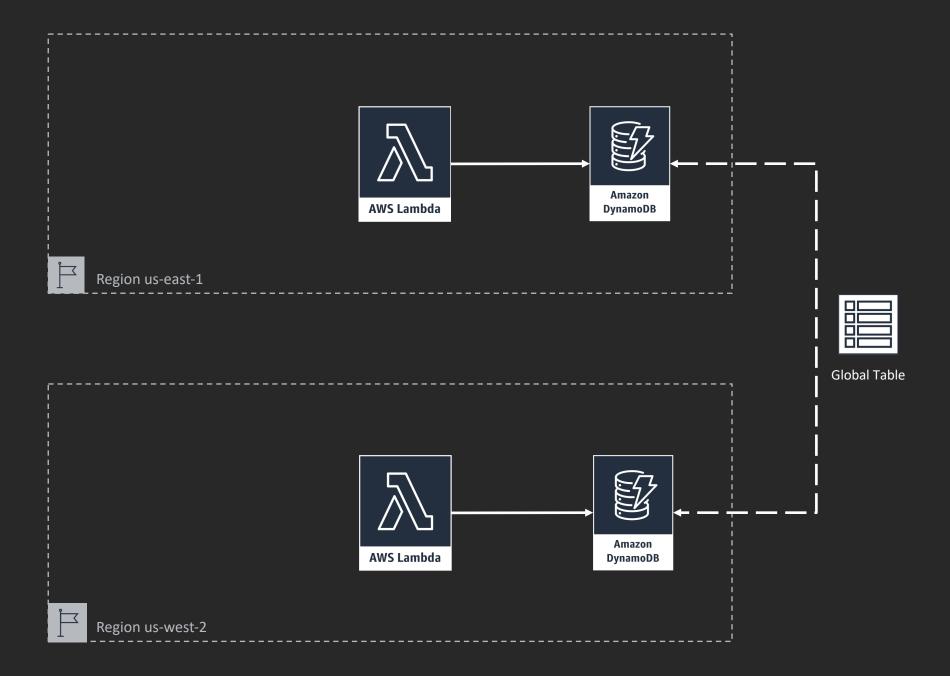




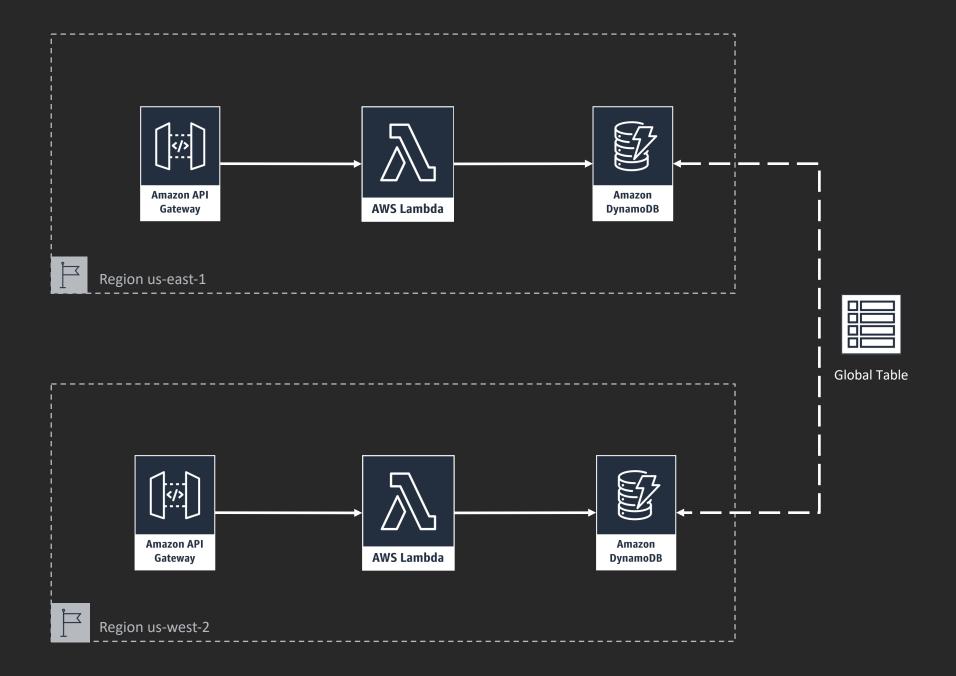






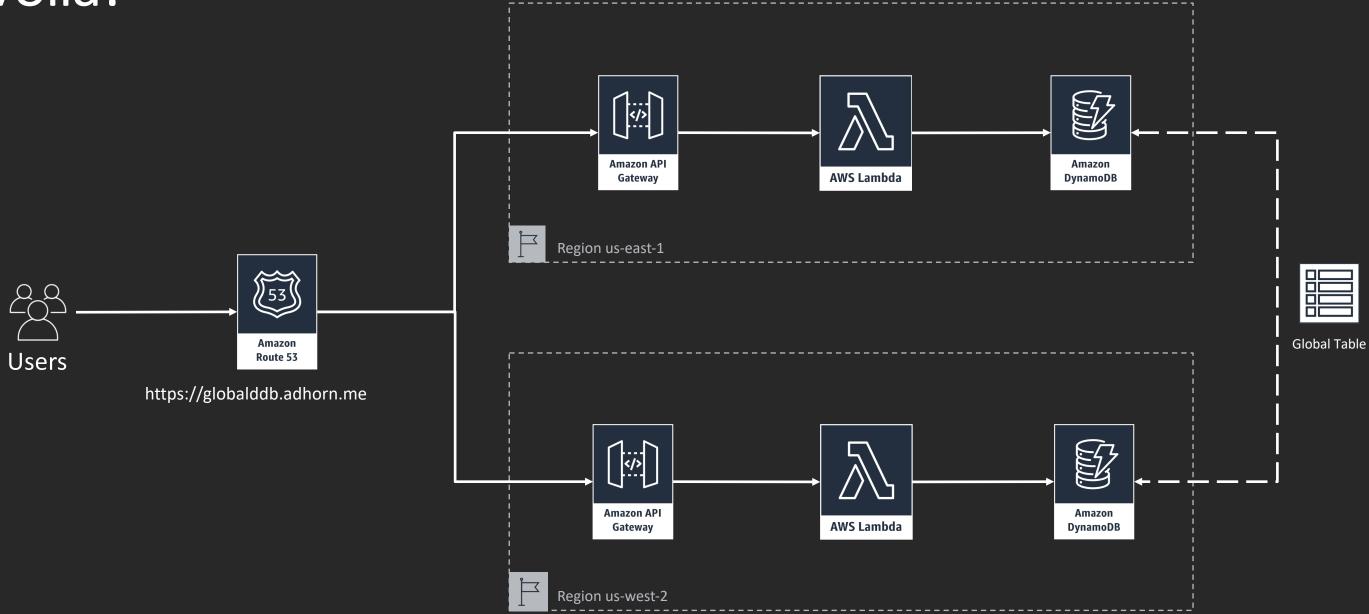








# Voilà!







Build a serverless multi-region, active-active backend solution — within a VPC

A VPC improves privacy...









Build a serverless multi-region, active-active backend solution in an hour

The solution is built using...



5 responses 🔲 🗸





Adrian Hornsby in A Cloud Guru

## RESILIENT ARCHITECTURE Multi-Region and Active-Active

How to build a multi-region active-active architecture on AWS

Everything fails all the time - build...





## https://medium.com/@adhorn



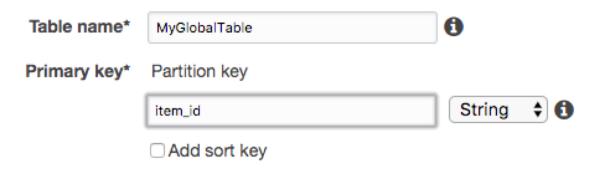
# Thanks you!





### Create DynamoDB table

DynamoDB is a schema-less database that only requires a table name and primary key. The table's primary key is made up of one or two attributes that uniquely identify items, partition the data, and sort data within each partition.



#### Table settings

Default settings provide the fastest way to get started with your table. You can modify these default settings now or after your table has been created.

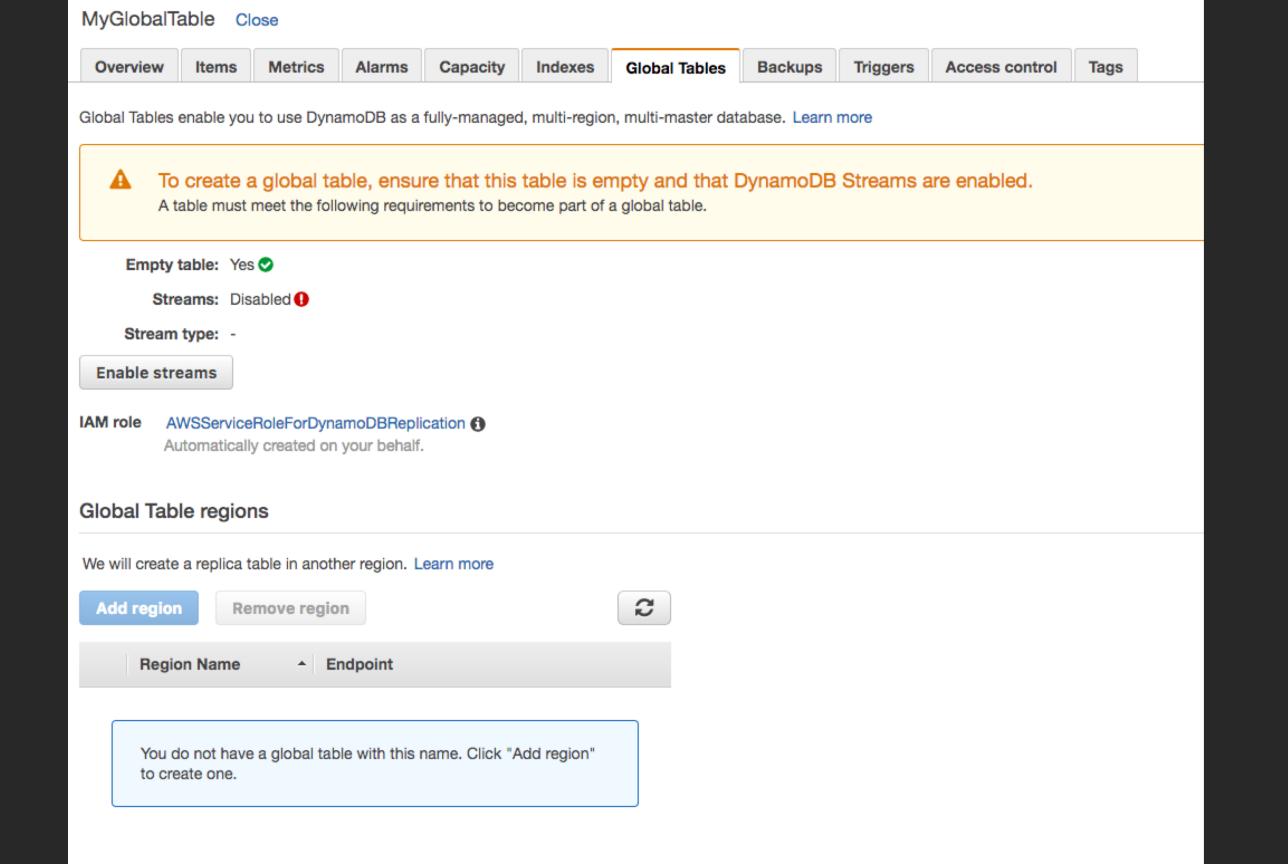
#### Use default settings

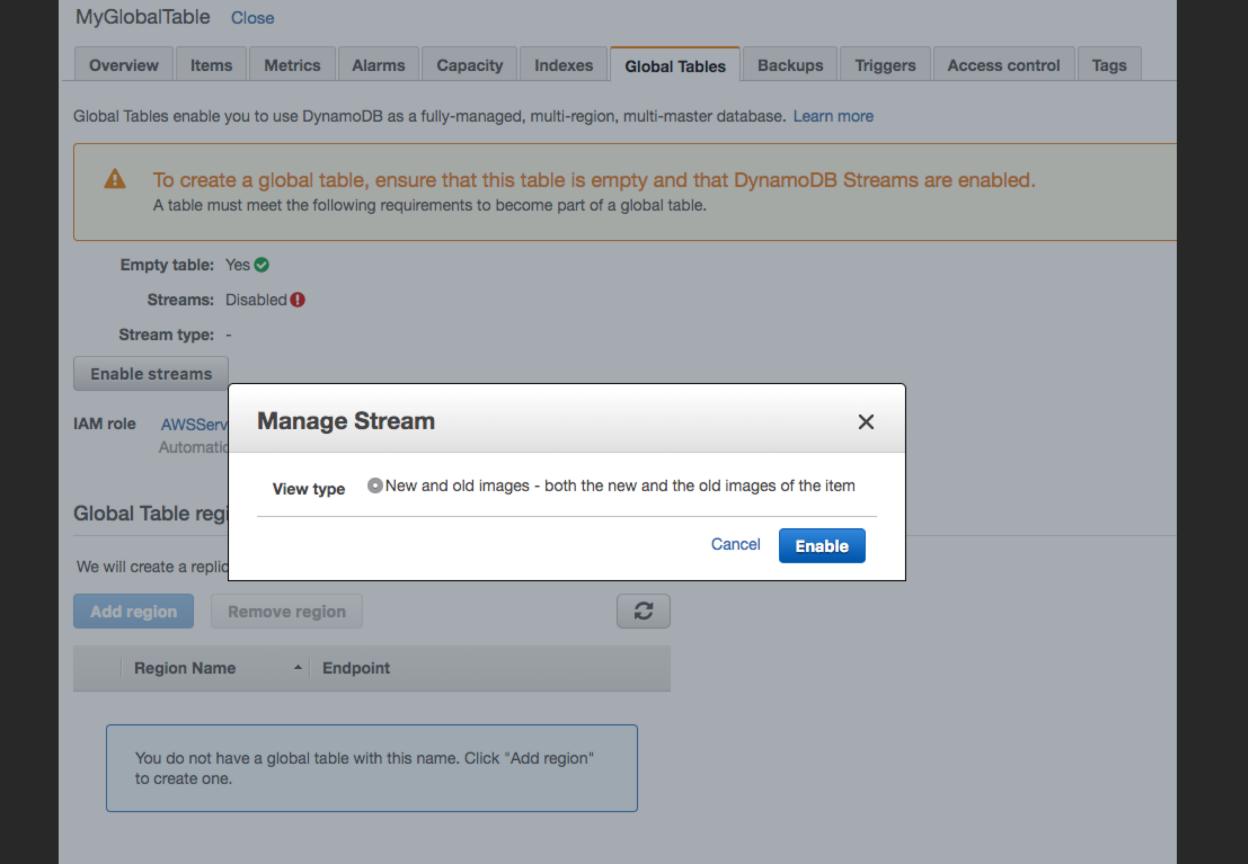
- · No secondary indexes.
- Auto Scaling capacity set to 70% target utilization, at minimum capacity of 5 reads and 5 writes
- On-Demand Backup and Restore Enabled NEW!

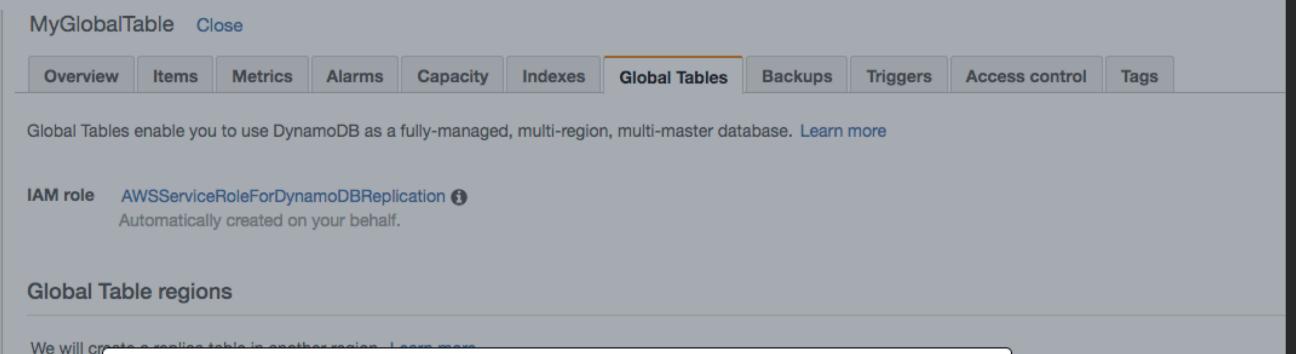
Additional charges may apply if you exceed the AWS Free Tier levels for CloudWatch or Simple Notification Service. Advanced alarm settings are available in the CloudWatch management console.

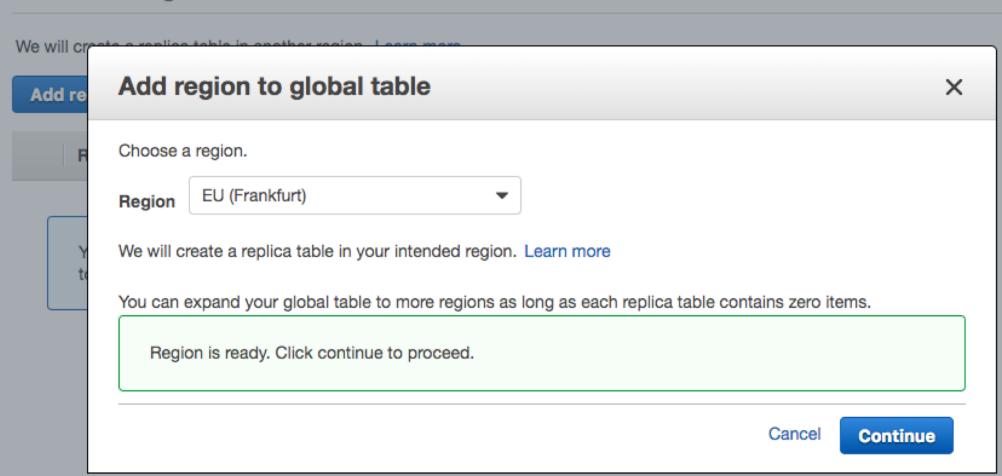
Cancel

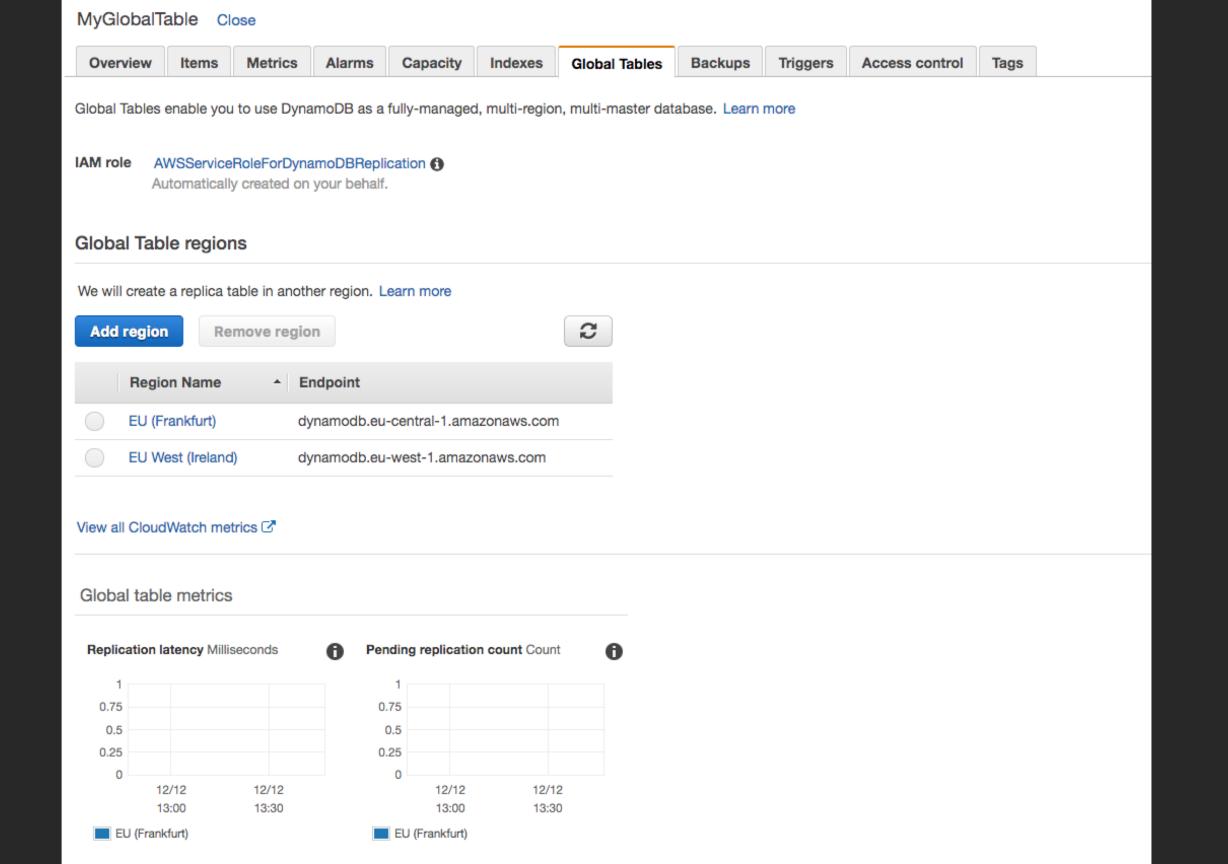




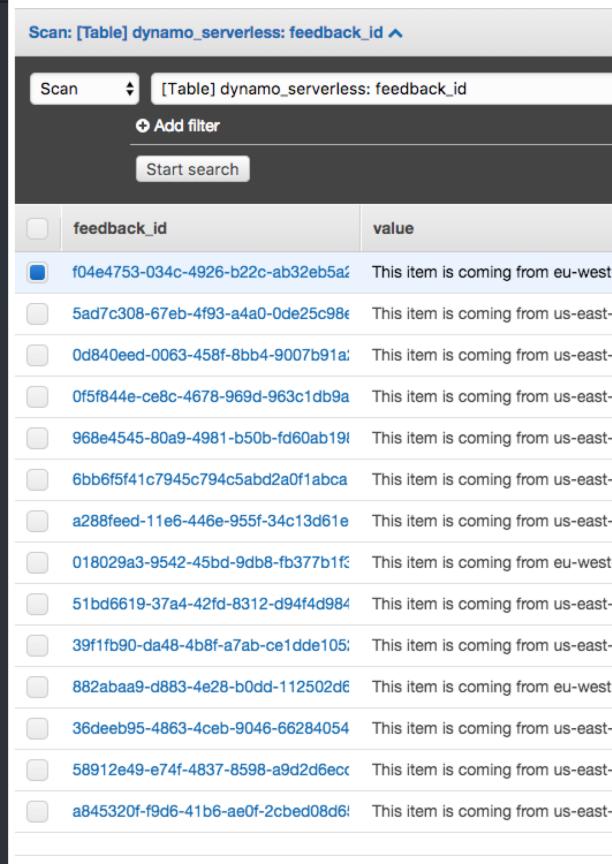




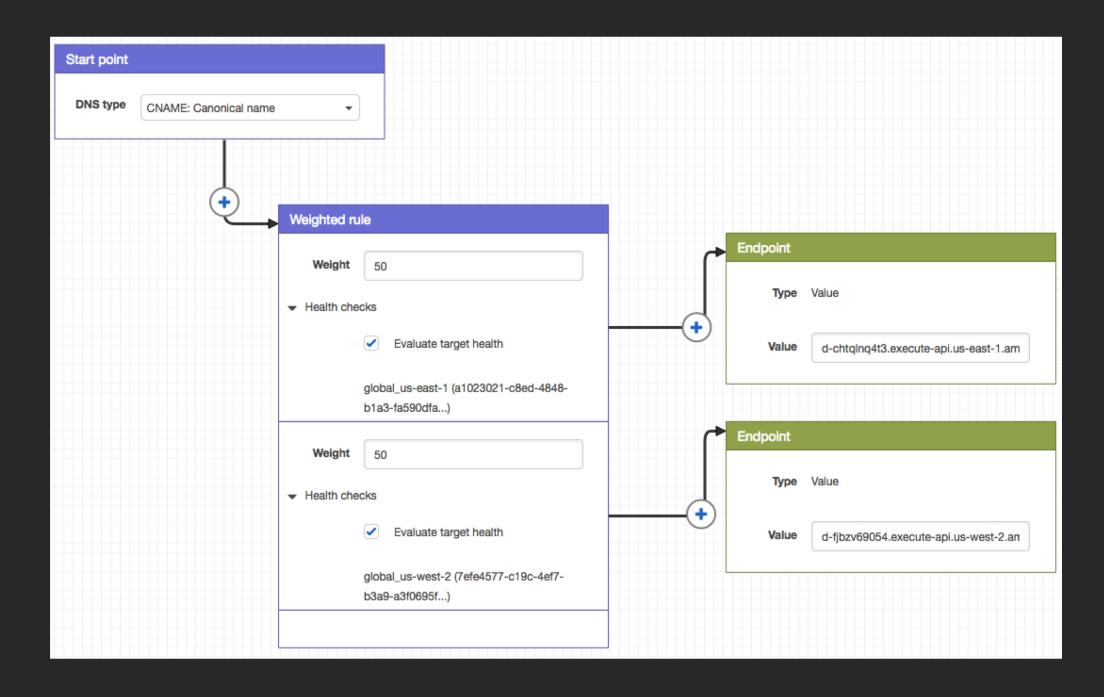




```
import ison
import logging
import boto3
import os
import uuid
log = logging.getLogger()
log.setLevel(logging.DEBUG)
region = os.environ["AWS_REGION"]
dynamodb = boto3.resource('dynamodb', region_name=region)
table = dynamodb.Table('dynamo_serverless')
def put to dynamo(event):
    log.debug("Received in put_to_dynamo: {}".format(json.dumps(event)))
    feedback id = str(uuid.uuid4())
    table.put_item(
        Item={
            'feedback_id': feedback_id,
            'value': "This item is coming from {}".format(region),
    return feedback_id
```



## Route53: Traffic Policy



## Health checks with Route 53

