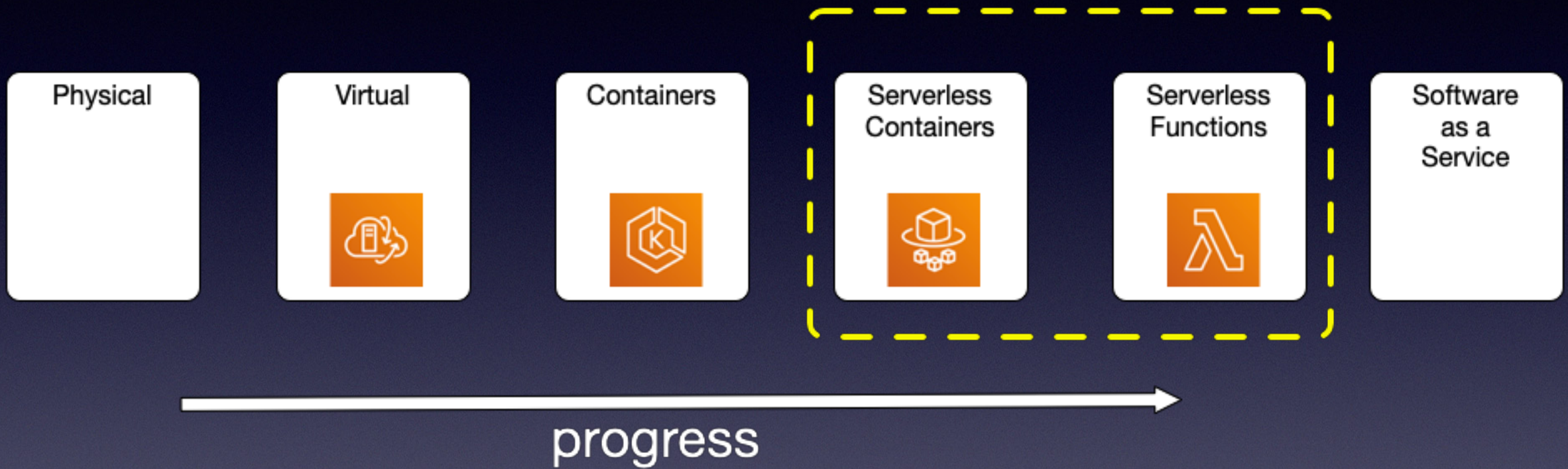
Java & AWS
The Journey to Serverless

Jason Poley / Distinguished Engineer

History of Serverless

- **AWS Lambda - Reinvent 2014**
- Google Cloud Functions 2016
- Azure Functions 2016
- IBM OpenWhisk 2016
- **AWS Fargate - Reinvent 2017**

Journey to Serverless



A few customers already…

airbnb
mlbam
COMCAST
Experian
Coca-Cola
LexisNexis
New York Public Library
Benchling
Pentair
DBRS
instacart
Zillow
ALT/S
VOGUE
23andMe
SCHOLASTIC
CITRIX
McGraw Hill
The Seattle Times
Aditya Birla Group
VidRoll
Localytics
asurion
Expedia
Atlassian
FireEye
vevo
EDIZIONI CONDÉ NAST S.P.A.
BUSTLE
THOMSON REUTERS
ViewLift
SQUARE ENIX
Nextdoor
AirAsia
ZAPPROVED
FannieMae
The Honest Co.
Vertafore
FUNNY OR DIE
Genome Institute of Singapore
MORNINGSTAR
Haier
iRobot
ISCS
la Mutuelle Generale
LA NACION
rachio
Discovery Communications
TTI
FICO
twilio
FINRA
VERACODE
MADSACK Mediengruppe
BEACHBODY ON DEMAND
AUTODESK
WARBY PARKER
doppler labs
Cornelius
WHALEROCK INDUSTRIES
cimpress
Nishitetsu
Optimizely
ASTRO DIGITAL
SAISON INSURANCE
Serenova shine bright
nextel TÁ NAS SUAS MÃOS.
INTERCOM
County of Alameda California
NATIONAL GEOGRAPHIC
AdRoll
DBS
edmunds
teralytic
HomeAway
Fender
Ringier

Defining Serverless

- Scale to zero cost
- Pay for usage (~100 ms inc)
- Scales Horizontally
- Event Driven / schedule placement for you.
- Typically pass in code or artifact (container)

Types of Serverless

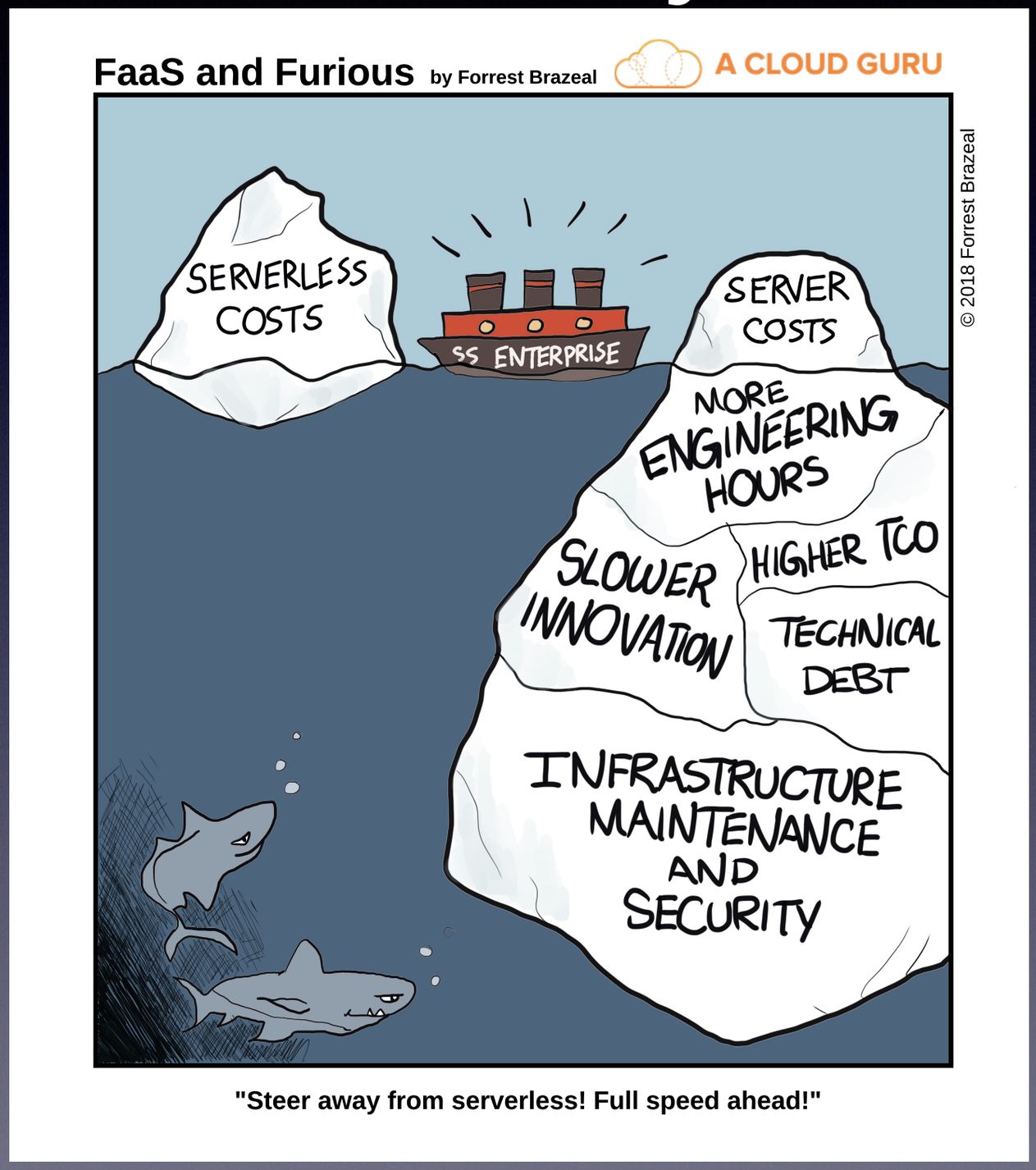
- App tier <= we will focus here
- Serverless databases
- others?

Todays Discussion

AWS Lambda

ECS Fargate

Why Serverless?



"Steer away from serverless! Full speed ahead!"

What about Lock In?

Hidden Figures, Fortran, Diana,
And Multi-Cloud Serverless

https://www.youtube.com/watch?v=oNa3xK2GFKY

You mean you can't rewrite a function wrapper?

https://github.com/cncf/wg-serverless

https://github.com/virtual-kubelet

cloudevents

Its all about the events

https://cloudevents.io/

CNCF Serverless

https://landscape.cncf.io/format=serverless

https://landscape.cncf.io/

Choices to make?

- Secrets Management
- Language to use
- Sizing
- Eventing Model
- Network Endpoints
- Frameworks

Secrets Management

- Hardcode the password
- Use a property
- Use AWS Parameter Store 
- Use AWS Secrets Manager 

Language to Use

AWS OPEN SOURCE

PARTNER SUPPORTED

Sizing

Memory: 128 mb to 3 GB (CPU matches)

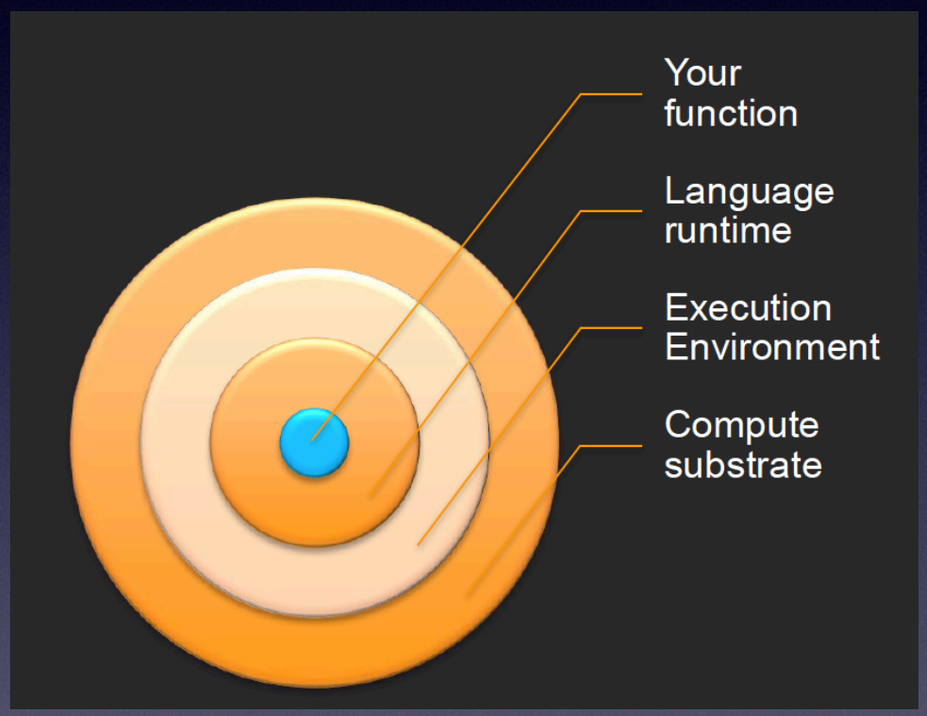
Timeout: 1 sec to 15 min

Payload size: up to 6 MB

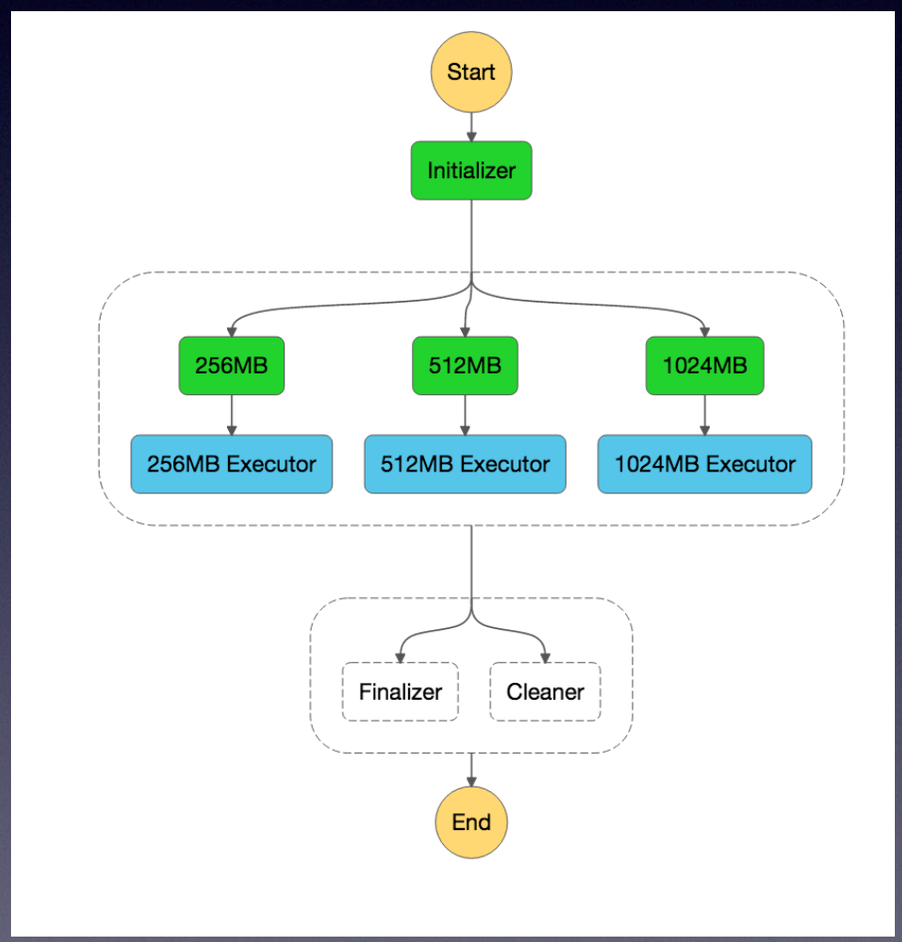
Disk size: up to 512 MB

Event Request: up to 128 KB

Lambda Tuning



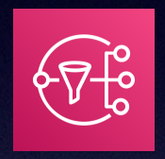
Costing in 100 ms increments



https://github.com/alexcasalboni/aws-lambda-power-tuning

Eventing Model



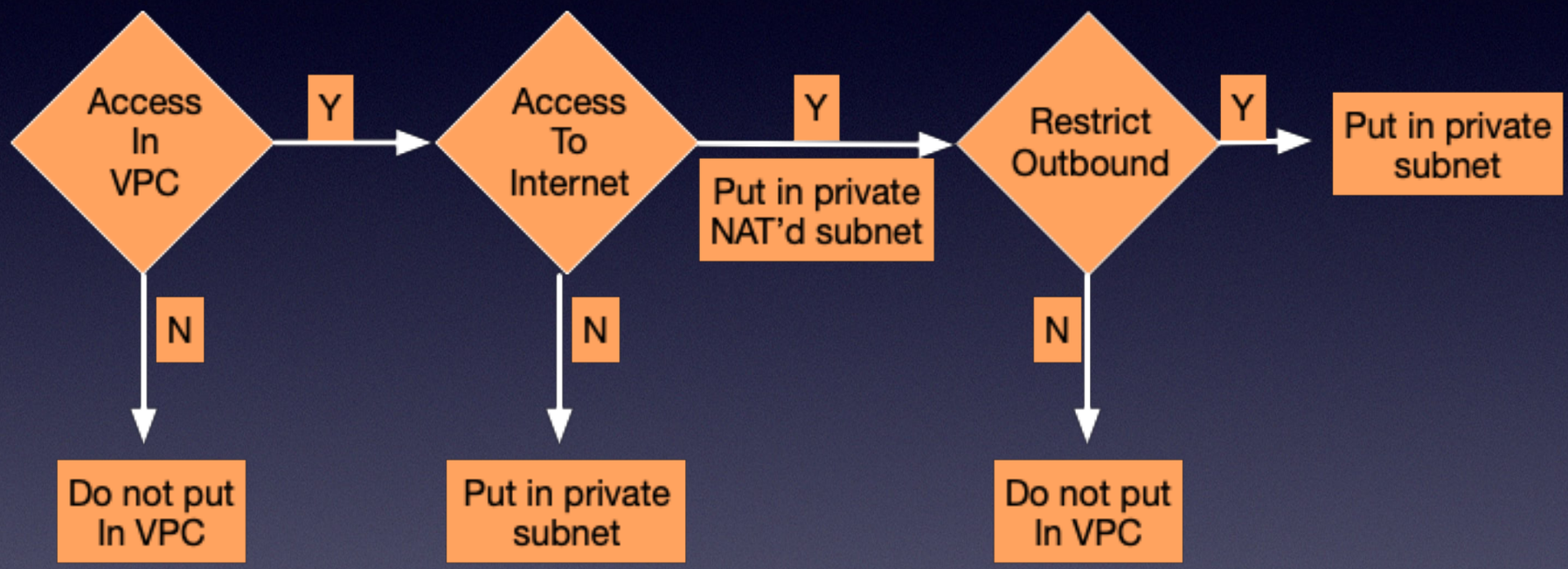


Sync
API Gateway
Application LB

Async
SNS
S3
SQS

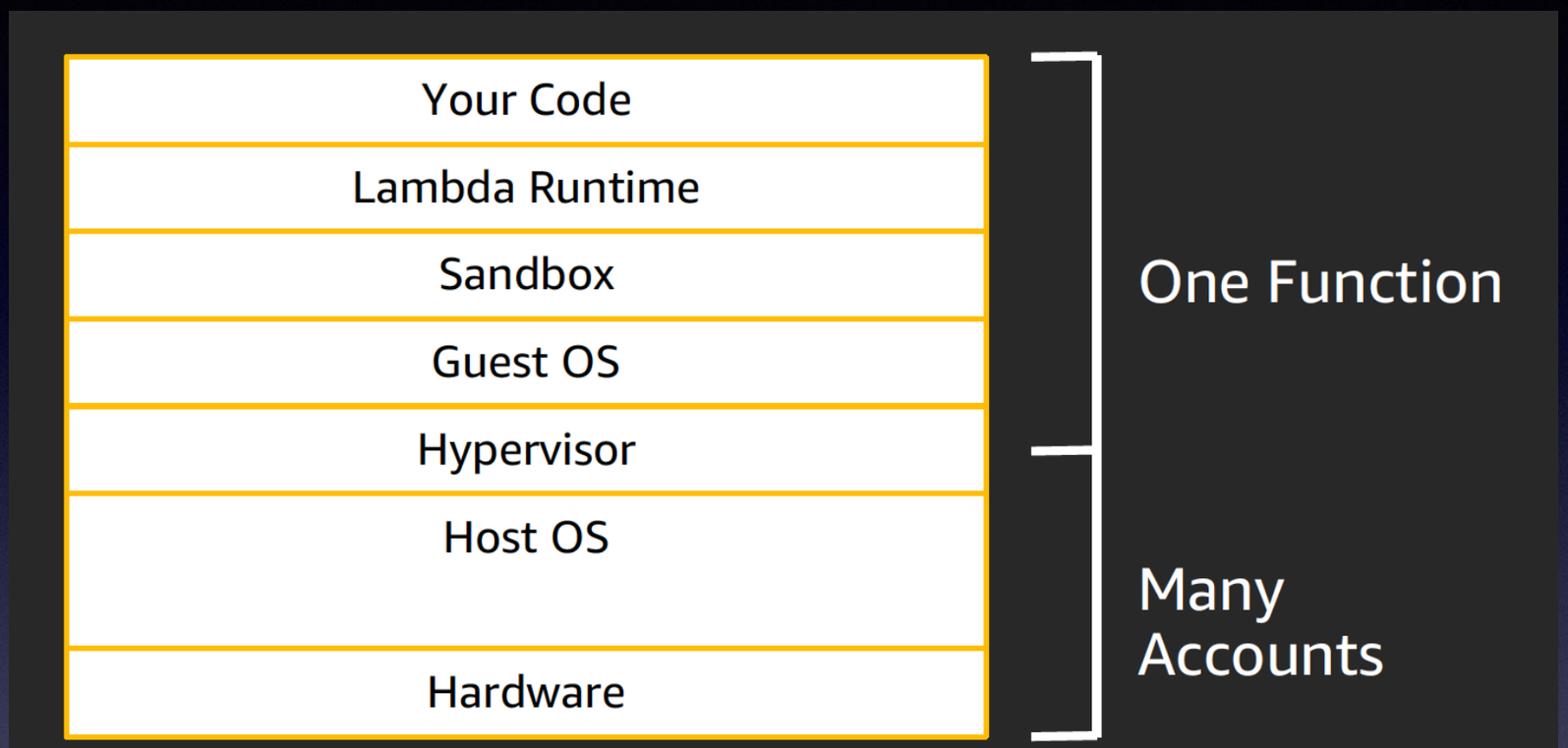

Poll
Dynamo
Kinesis

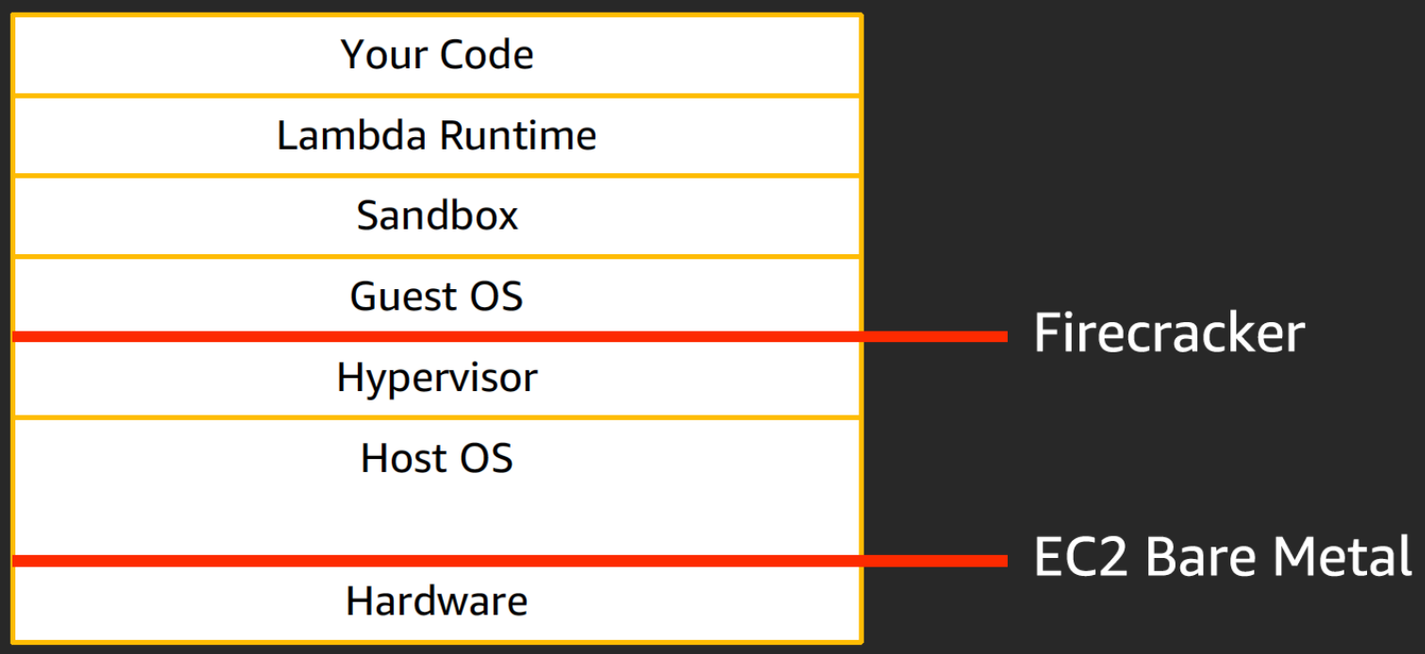
Network Endpoints



Isolation

VPC endpoint
Or not?





Shared VPC
Tenancy only

Frameworks?

- AWS Serverless Application Model - Natively supported with CloudFormation
- Apex - Go/Java/JavaScript/Python/(Others) - Hookable build process, supports deploying infrastructure using Terraform - Go
- Chalice - Python2.7 - Written by AWS, still in preview - Python
- Gordon - Go/Java/JavaScript/Python/Scala - Automatically packages dependencies, integration with pip/npm/gradle, Deploys resources using CloudFormation - Python
- Serverless - Java/JavaScript/Python/Scala - Lots of Plugins, Deploys resources using CloudFormation - JavaScript
- Zappa - Python
- Backand - Node.JS/JavaScript - A serverless development platform built to make AWS Lambda easier. Create your own Lambda functions, or connect your AWS account to use your Lambda functions in the Lambda Launcher tool, providing easy access to running your Lambda functions
- FuseLess - CFML/ColdFusion - a toolkit for running Lambda functions written in CFML.

- Google Dagger2 - A fast dependency injector for Android and Java
- Jackson-Jr - Stand-alone data-binding module designed as a light-weight (and -featured) alternative to `jackson-databind`: will only deal with "Maps, Lists, Strings, wrappers and Java Beans" (jr-objects), or simple read-only trees (jr-stree)

Serverless Application Model

AWS CloudFormation extension optimized for Serverless

Serverless resource types: functions, apis, tables

Local testing with SAM CLI

https://github.com/awslabs/aws-sam-cli

API Gateway authorizers, CORS, DynamoDb Server Side Encryption

https://github.com/awslabs/serverless-application-model

Message Comparison

	Amazon MQ	Kinesis Data Streams	SQS (Standard)	SQS (FIFO)	SNS	MSK
Serverless	No	Yes	Yes	Yes	Yes	No
Guaranteed Ordering	Yes	Yes with shard	No	Yes with Message Group	No	Yes on a topic
Message Access	Pull or Push	Pull or Push (http/2)	Pull	Pull	Push	Pull
Delivery	exactly-once	At least once	At least once	exactly-once	retries	exactly-once
Data retention period	Unlimited	1w	2w	2w	Until retry completes	?
Number of Messages	Unlimited	Unlimited	120,000 limit	20,000 limit	Unlimited	Unlimited
Parallel Consumers	Yes	Multiple consumers per shard with independent iterators	Multiple readers (but only one message per reader at a time)	One consumer at a time	Multiple subscribers per topic	Yes
Delivery destination types	ActiveMQ Clients	Kinesis Consumers (KCL, Lambda SDK)	SQS Readers, Lambda	SQS Readers	http/s, mobile push, sms, email ,sqs, lambda	Kafka clients
Row/Object Size	Unlimited	1 MB	256 KB (TB / S3)	256 KB	256 KB	Unlimited

A few Tips for Lambda

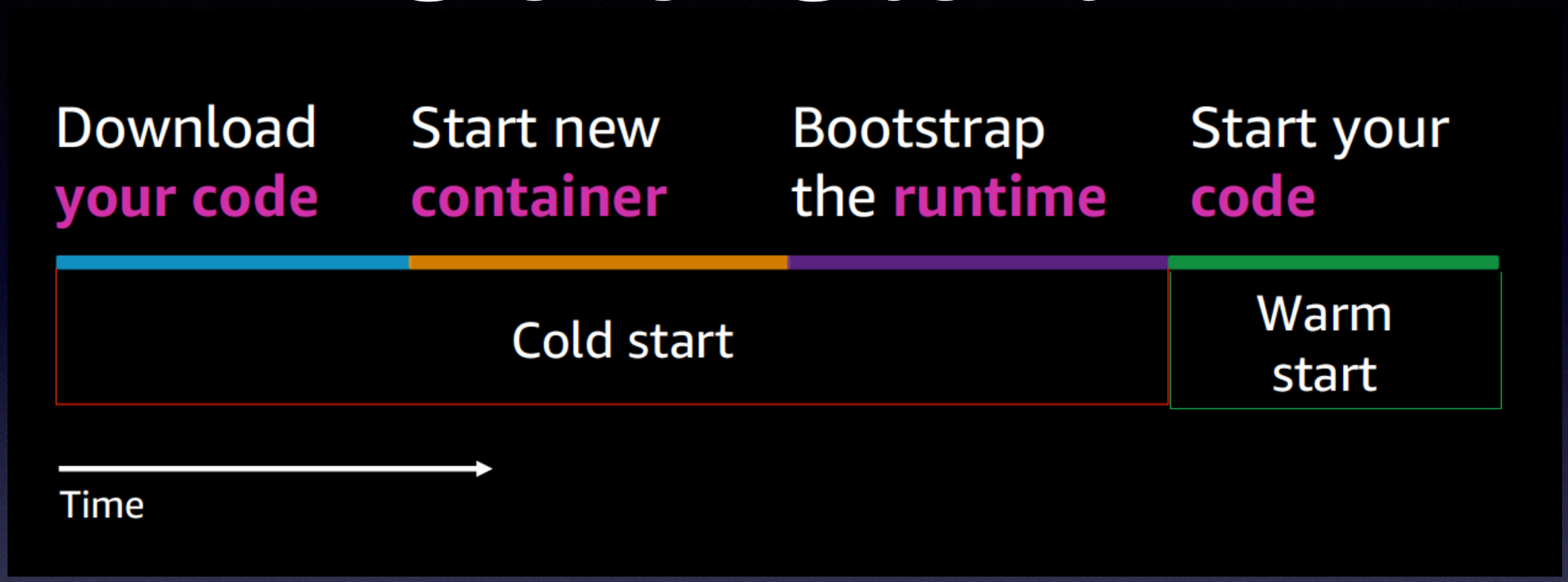
- Minimize your package size, use only needed SDK modules (spring boot probably not, see fargate)
- Put your dependency .jar in separate /lib dir
- Use smaller IoC framework (like dagger2) to load quickly
- Smaller faster data binding choices jackson-jr
- Use environment variables to modify behavior
- Make sure functions invoked via SQS don’t exceed visibility timeout
- Test your code (not unique to lambda)

Top 10 Security Concerns for Serverless

- 1: Function Event Data Injection
- 2: Broken Authentication
- 3: Insecure Serverless Deployment Configuration
- 4: Over-Privileged Function Permissions & Roles
- 5: Inadequate Function Monitoring and Logging
- 6: Insecure 3rd Party Dependencies
- 7: Insecure Application Secrets Storage
- 8: Denial of Service & Financial Resource Exhaustion
- 9: Serverless Function Execution Flow Manipulation
- 10: Improper Exception Handling and Verbose Error Messages

https://github.com/puresec/sas-top-10 https://www.owasp.org/index.php/OWASP_Serverless_Goat

Cold Start

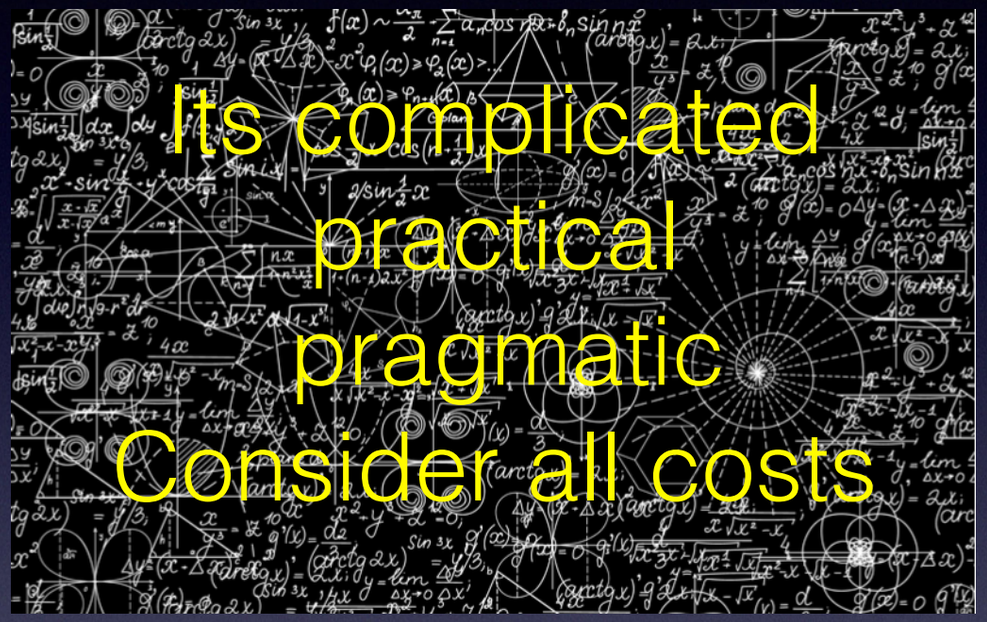


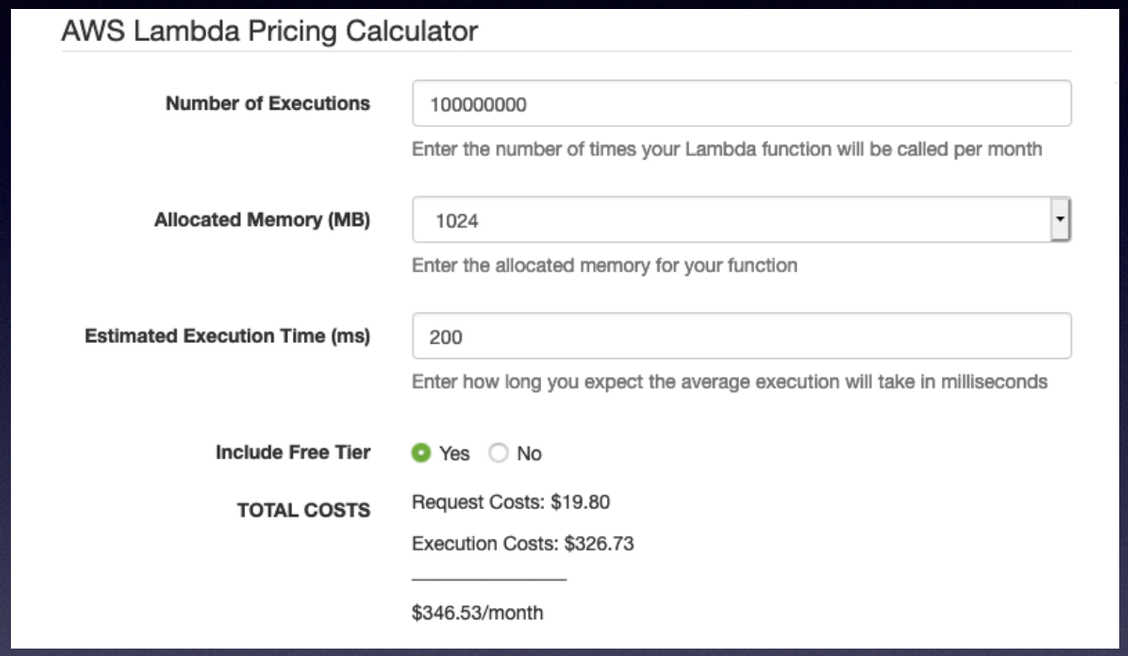
Pre warming instances
Code structure of load vs execution

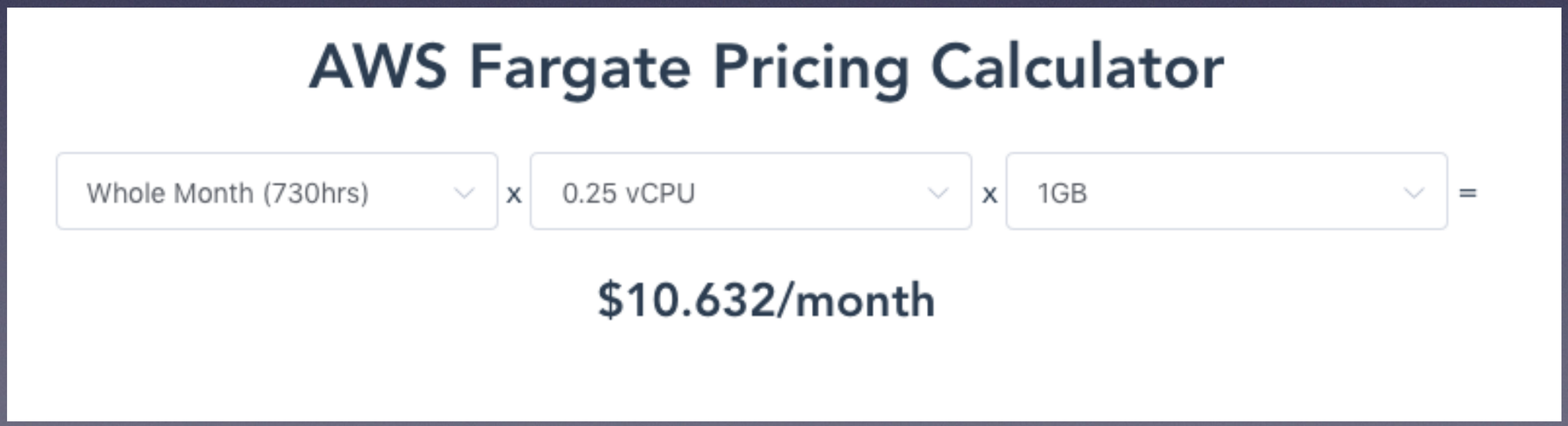
Language
Memory size
Code size
VPC ENI
HTTPS
Classpath scans

Cost Model of Serverless

https://s3.amazonaws.com/lambda-tools/pricing-calculator.html







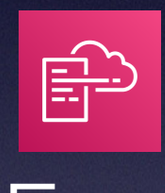
http://fargate-pricing-calculator.site.s3-website-us-east-1.amazonaws.com/

Invoking Serverless

https://docs.aws.amazon.com/lambda/latest/dg/invoking-lambda-function.html

API GW

ALB



CloudFormation

Cognito

Lex

Alexa

Config

CloudFront

CloudWatch

S3

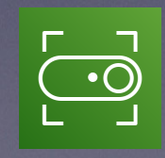
Dynamo

Kinesis

SNS

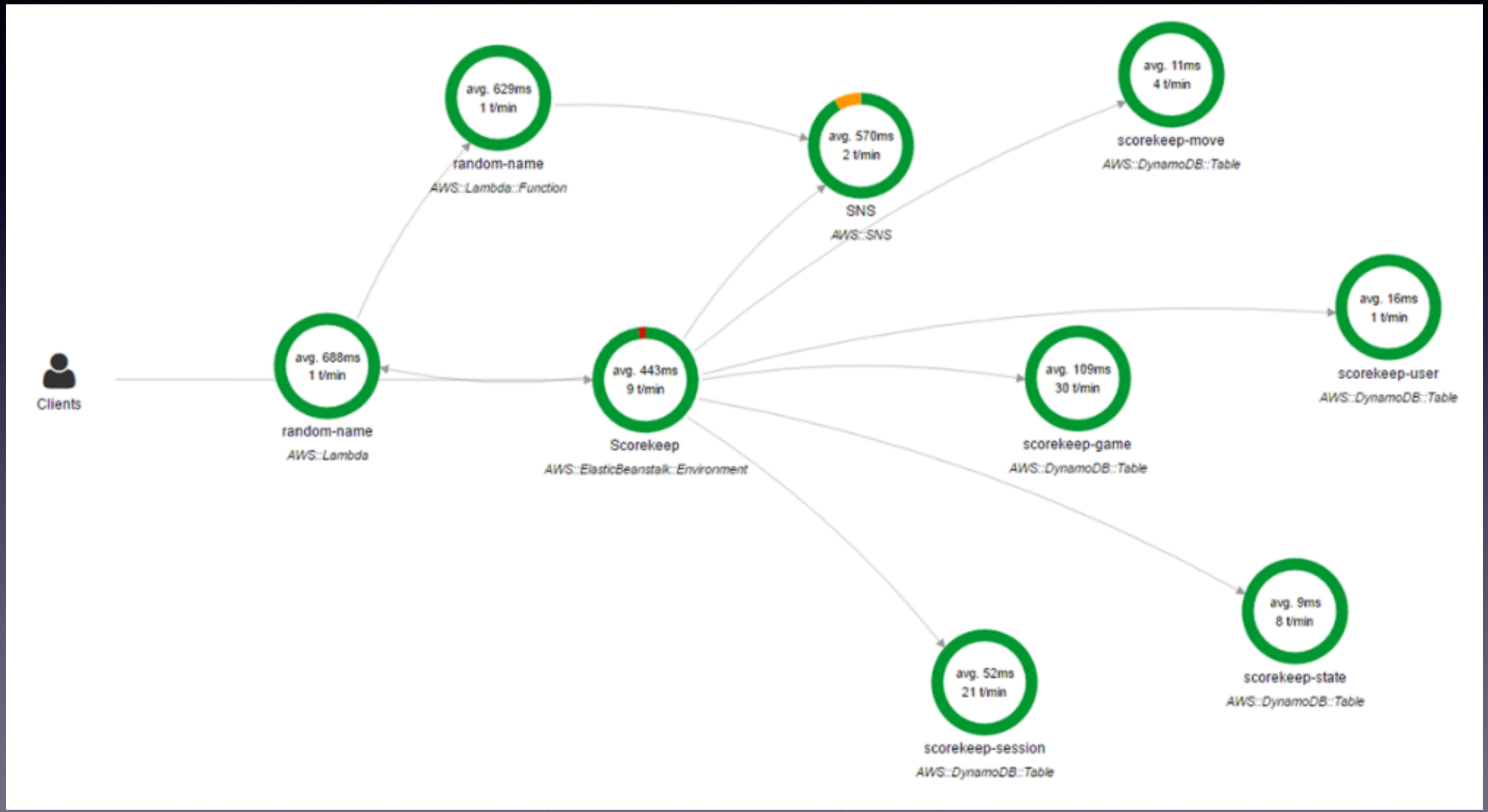
SES

SQS



IoT Button

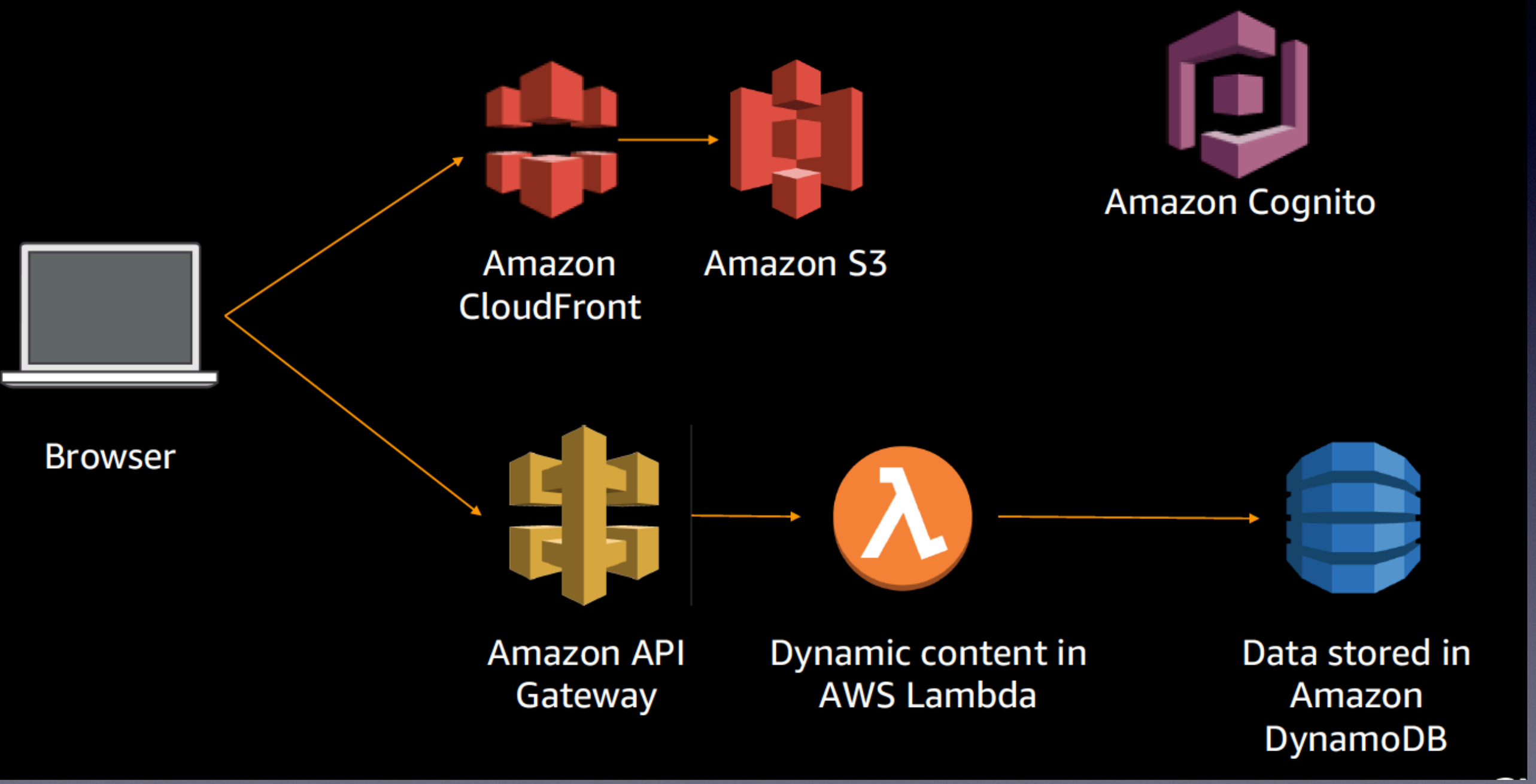
Monitoring with XRay



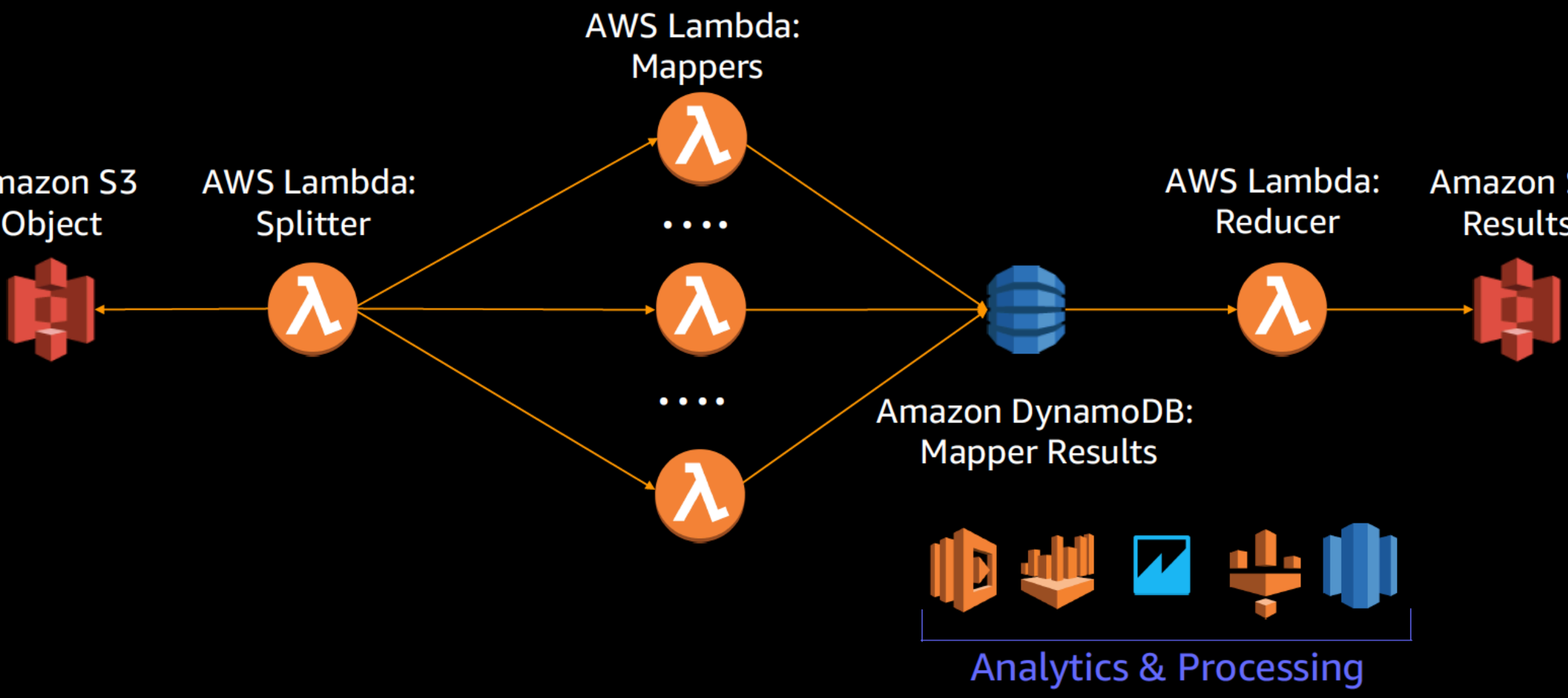
Key Use Cases

- Web Application
- Batch Processing
- Stream Processing
- Data Lake
- Machine Learning
- Mobile Backend
- Control Plane / Security

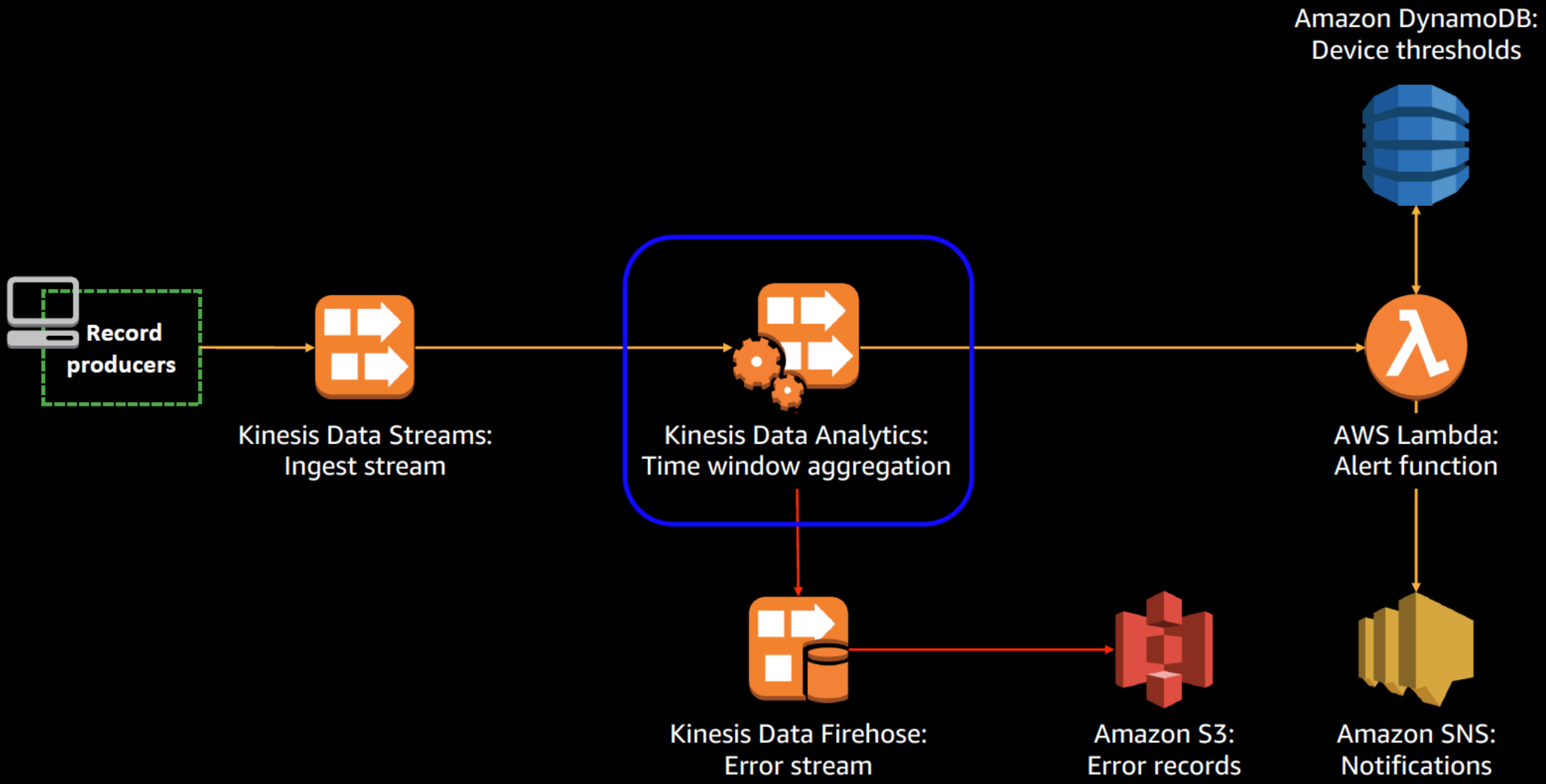
Web Application



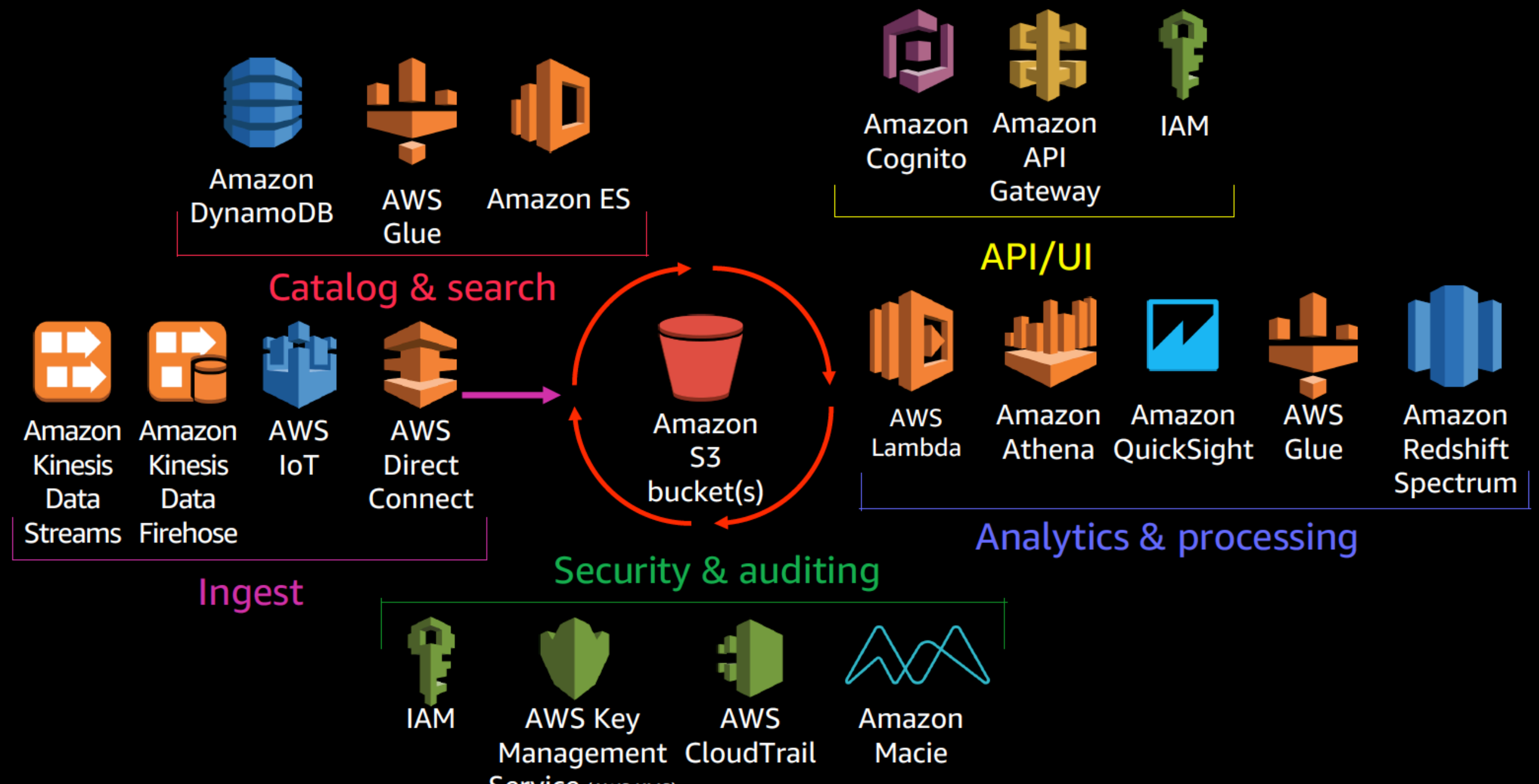
Batch Processing



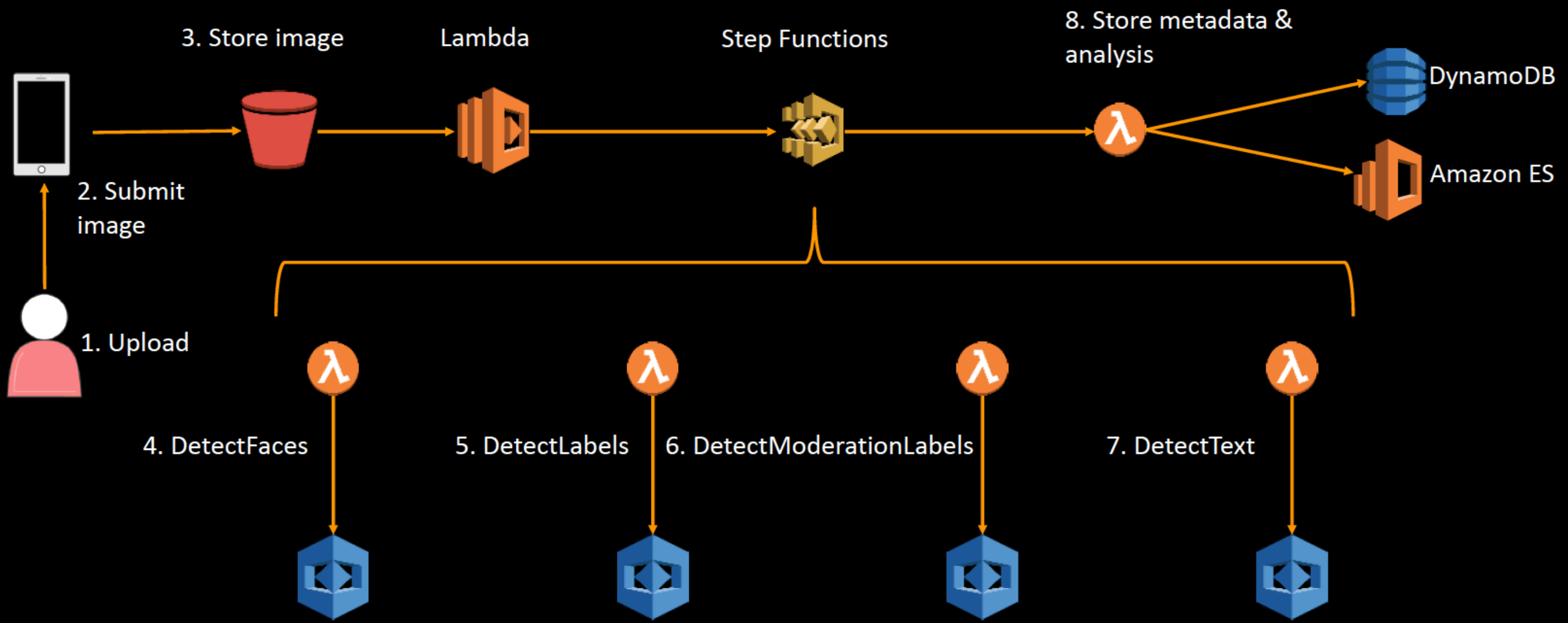
Stream Processing



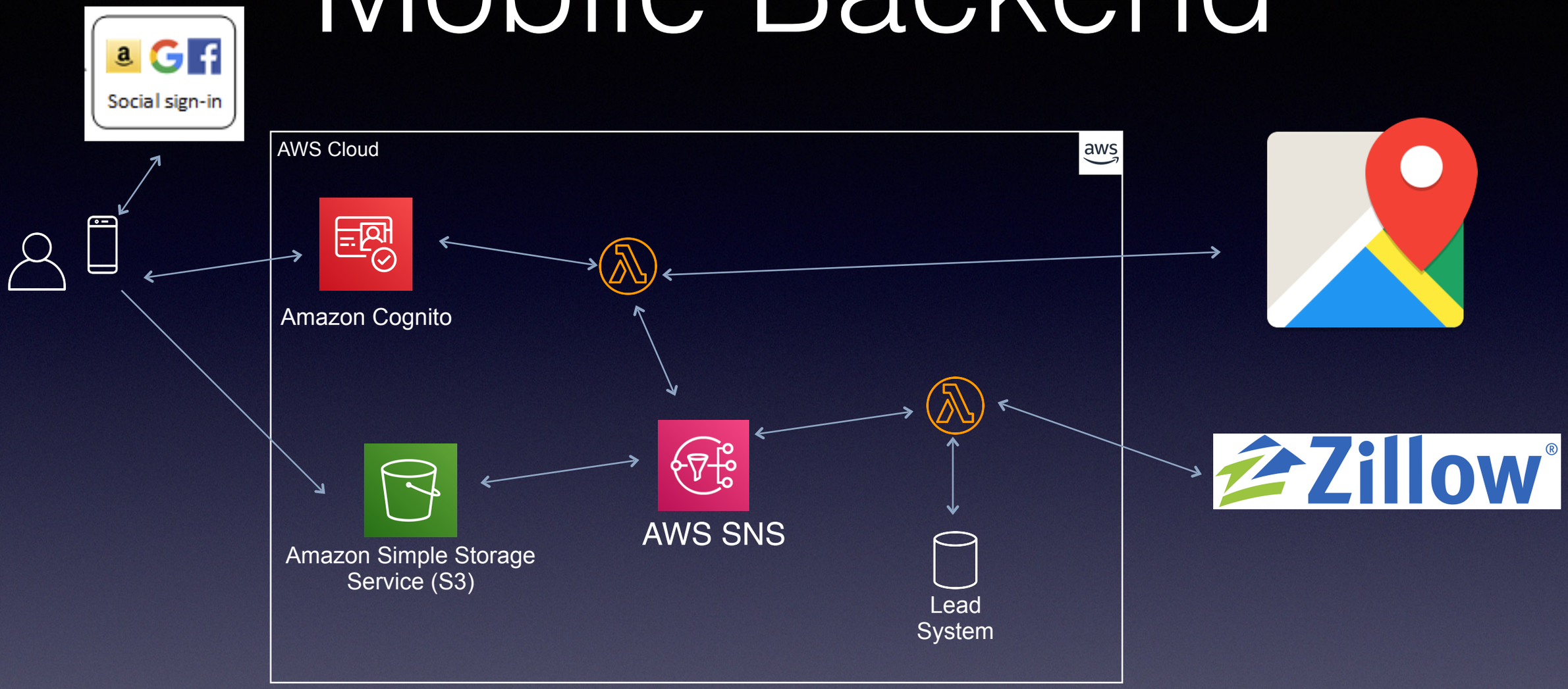
Data Lake


Amazon DynamoDB
AWS Glue
Amazon ES
Catalog & search
Amazon Cognito
Amazon API Gateway
IAM
API/UI
Amazon Kinesis Data Streams
Amazon Kinesis Data Firehose
AWS IoT
AWS Direct Connect
Amazon S3 bucket(s)
AWS Lambda
Amazon Athena
Amazon QuickSight
AWS Glue
Amazon Redshift Spectrum
Analytics & processing
Ingest
Security & auditing
IAM
AWS Key Management Service (AWS KMS)
AWS CloudTrail
Amazon Macie

Machine Learning



Mobile Backend



Login (Social Auth)
Take Picture
Get GPS
Get Address
Get Property Comp

Control Plane / Security

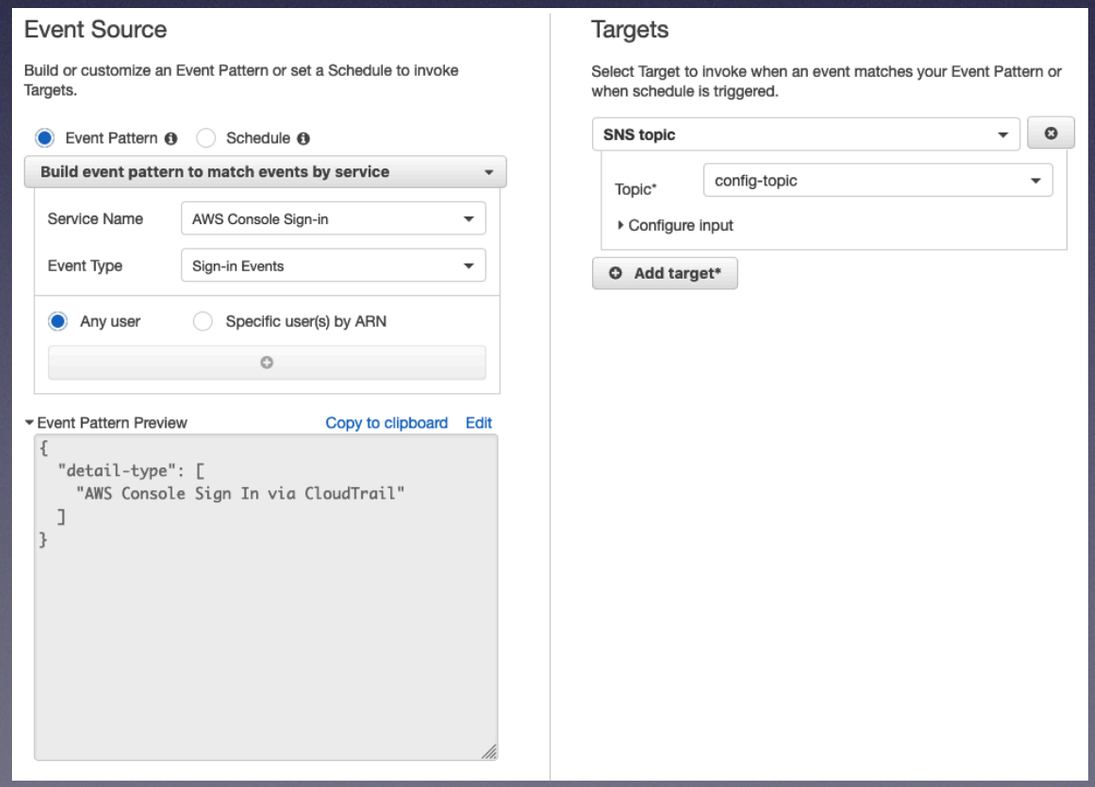
https://github.com/awslabs/aws-lambda-security-controls

ec2-approved-regions
ec2-asg-require-instance-profile
ec2-instance-profile-no-wildcards
ec2-no-public-amis
ec2-security-group-open
security-group-ingress-not-single-host
amazon-glacier
s3-enable-aes-256-encryption
s3-enforce-encryption-on-bucket
s3-public-bucket-detection-remediation
s3-require-customer-managed-keys

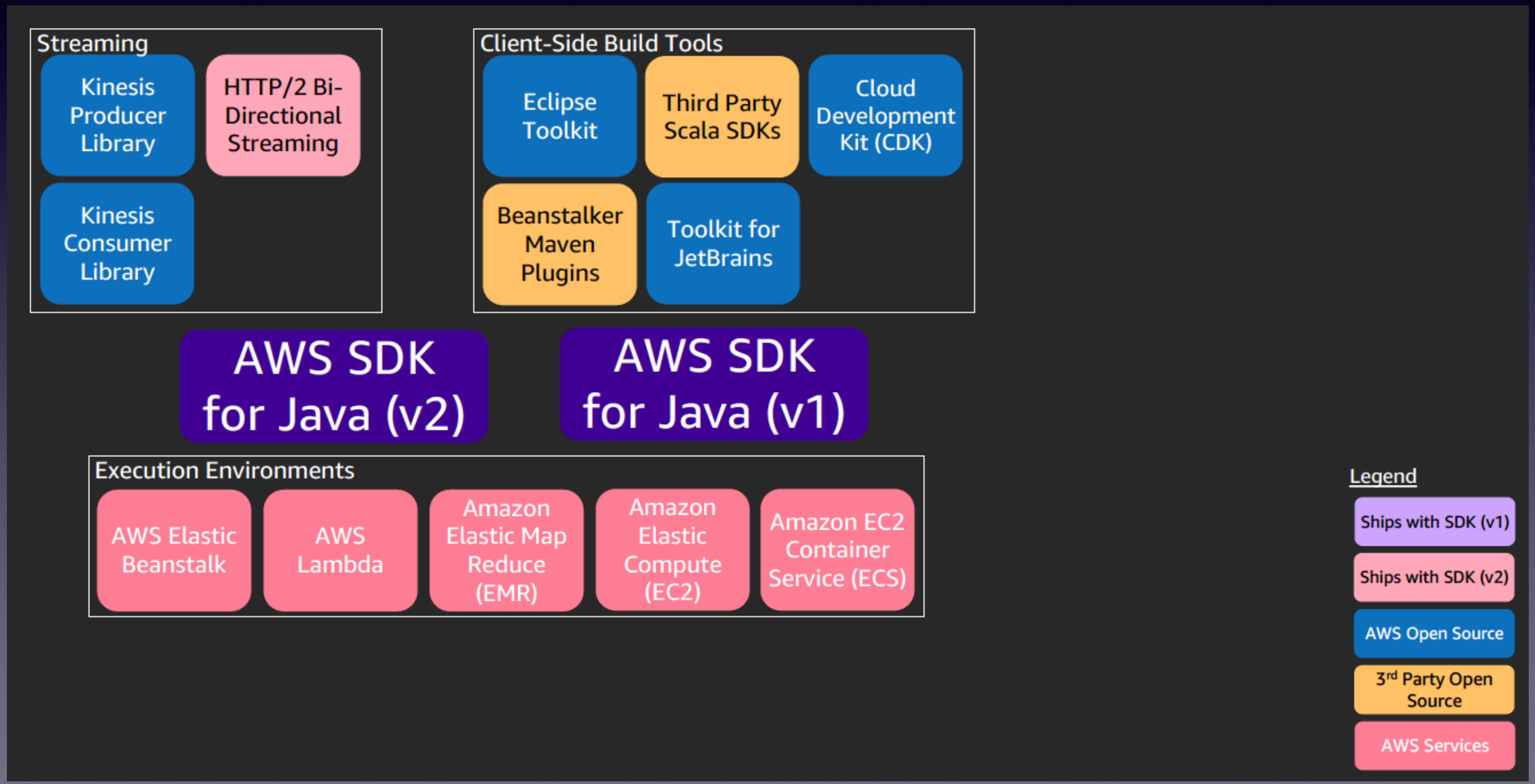
AWS Config

s3-bucket-versioning-enabled
cloud-trail-enabled
root-account-mfa-enabled
encrypted-volumes
eip-attached

CloudWatch Events



Java @ AWS


Streaming
Kinesis Producer Library
HTTP/2 Bi-Directional Streaming
Kinesis Consumer Library
Client-Side Build Tools
Eclipse Toolkit
Third Party Scala SDKs
Cloud Development Kit (CDK)
Beanstalker Maven Plugins
Toolkit for JetBrains
AWS SDK for Java (v2)
AWS SDK for Java (v1)
Execution Environments
AWS Elastic Beanstalk
AWS Lambda
Amazon Elastic Map Reduce (EMR)
Amazon Elastic Compute (EC2)
Amazon EC2 Container Service (ECS)
Legend
Ships with SDK (v1)
Ships with SDK (v2)
AWS Open Source
3rd Party Open Source
AWS Services

Java with Serverless

https://github.com/awslabs/aws-serverless-java-container

- **Spring** - `aws-serverless-java-container-spring`
- **Spring Boot** - `aws-serverless-java-container-spring`
- **Jersey** - `aws-serverless-java-container-jersey`
- **Spark** - `aws-serverless-java-container-spark`
- **Struts2** - `aws-serverless-java-container-struts2`
-

AWS Lambda

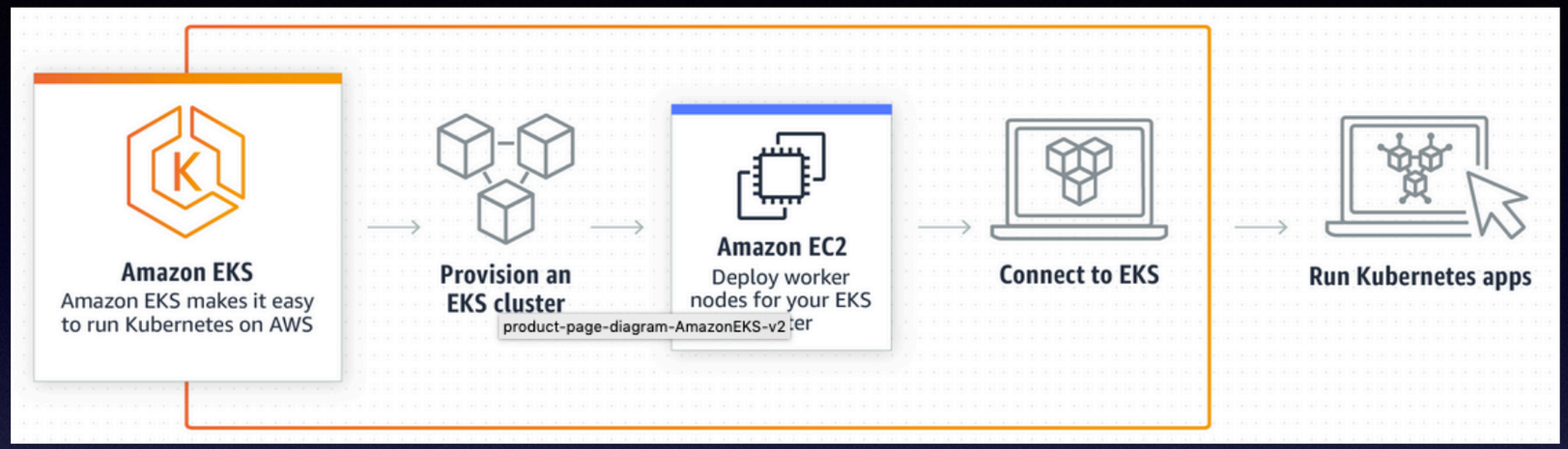
Favor when…

Smaller / infrequent jobs
Can use Dagger2 / Jackson-jr
Can turn code into functions

AWS Fargate

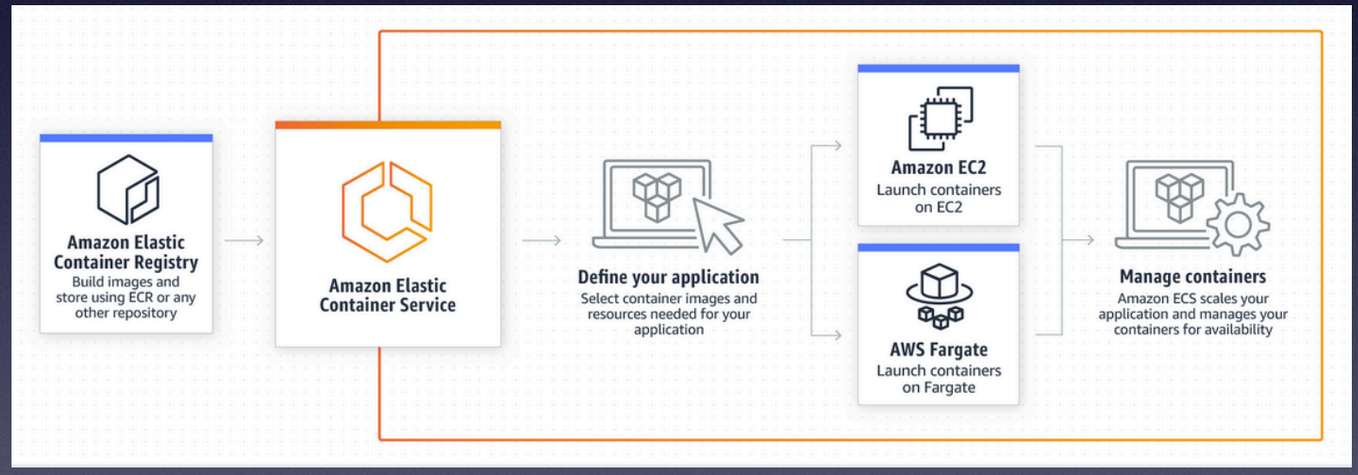
You have containers
You use spring / spring boot
You need something constantly running
You have larger memory requirement
You can’t fit code into “function”
You aren’t ready for AWS Lambda

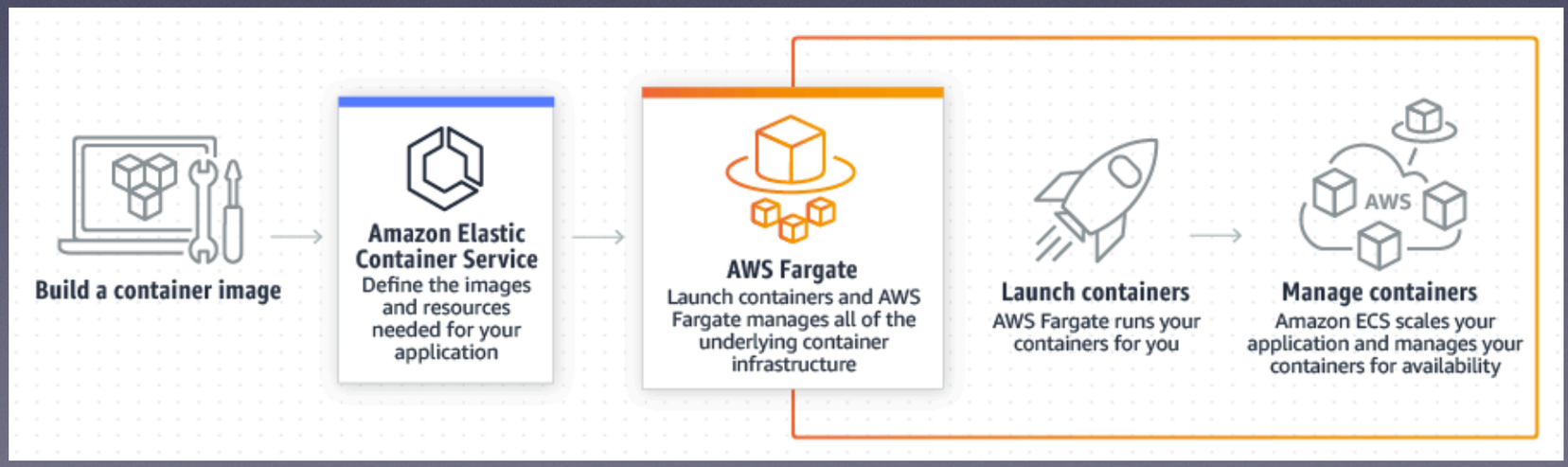
Containers Quick View



Amazon EKS

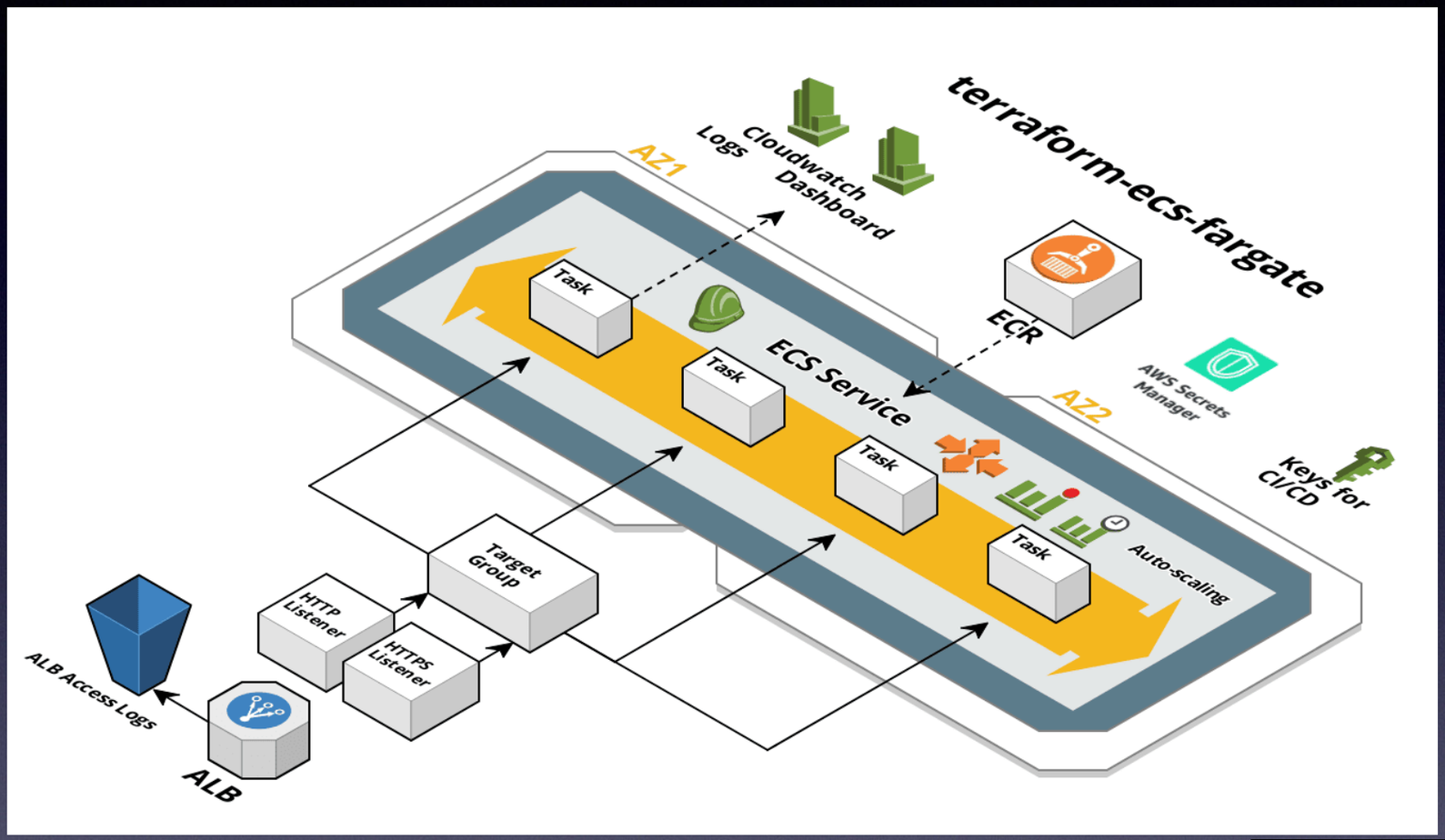
Amazon ECS





AWS Fargate

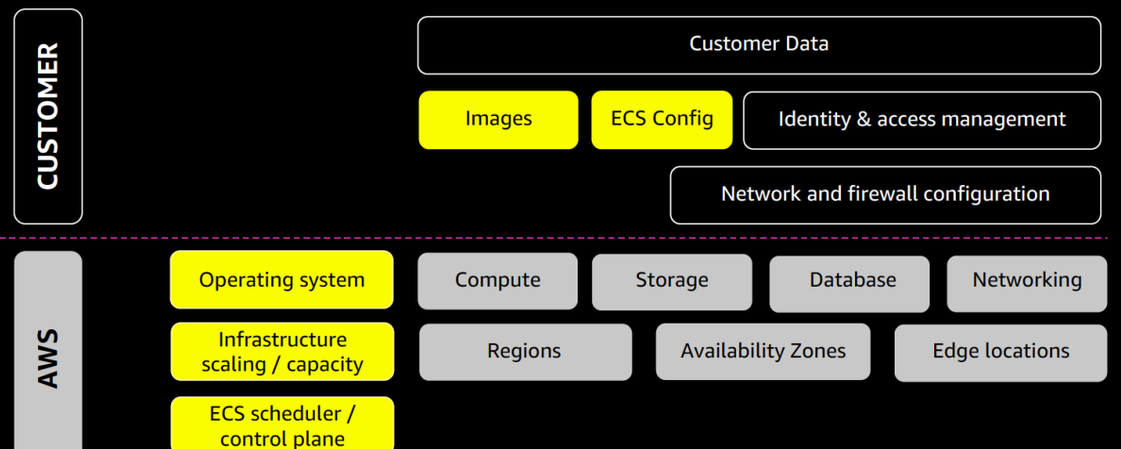
Fargate



No cluster mgmt
Awsvpc networking
Task Roles
Task metadata

Run continuously

Fargate shared security responsibility model



Not Quite for Fargate

- GPUs
- Running windows native container
- Spot Fleet / no RI pricing
- Specific Instance type required (c5 etc)
- Required privileged containers, daemon-sets or co-location of task
- 10 GB disk limit / no EBS attaching

New Features 2018

AWS Cloud9

AWS Toolkit
for PyCharm

AWS Toolkit
for IntelliJ

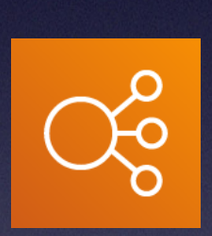
AWS Toolkit
for VS Code

Firecracker

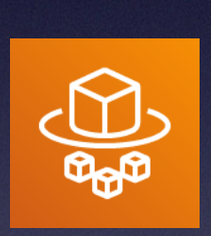
AWS Lambda layers
AWS Nested apps
Websocket support for API Gateway
ALB support for Lambda

Serverless Programming

API Gateway



Application LB



AWS Fargate

AWS Lambda

AWS CloudMap

Amazon MQ

AWS SQS

AWS SNS

DynamoDb

AWS S3

ecosystem & impact

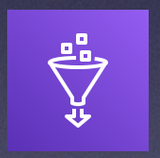
Amazon Athena

Amazon Redshift

AWS Step Functions

AWS Batch

AWS Snowball Edge



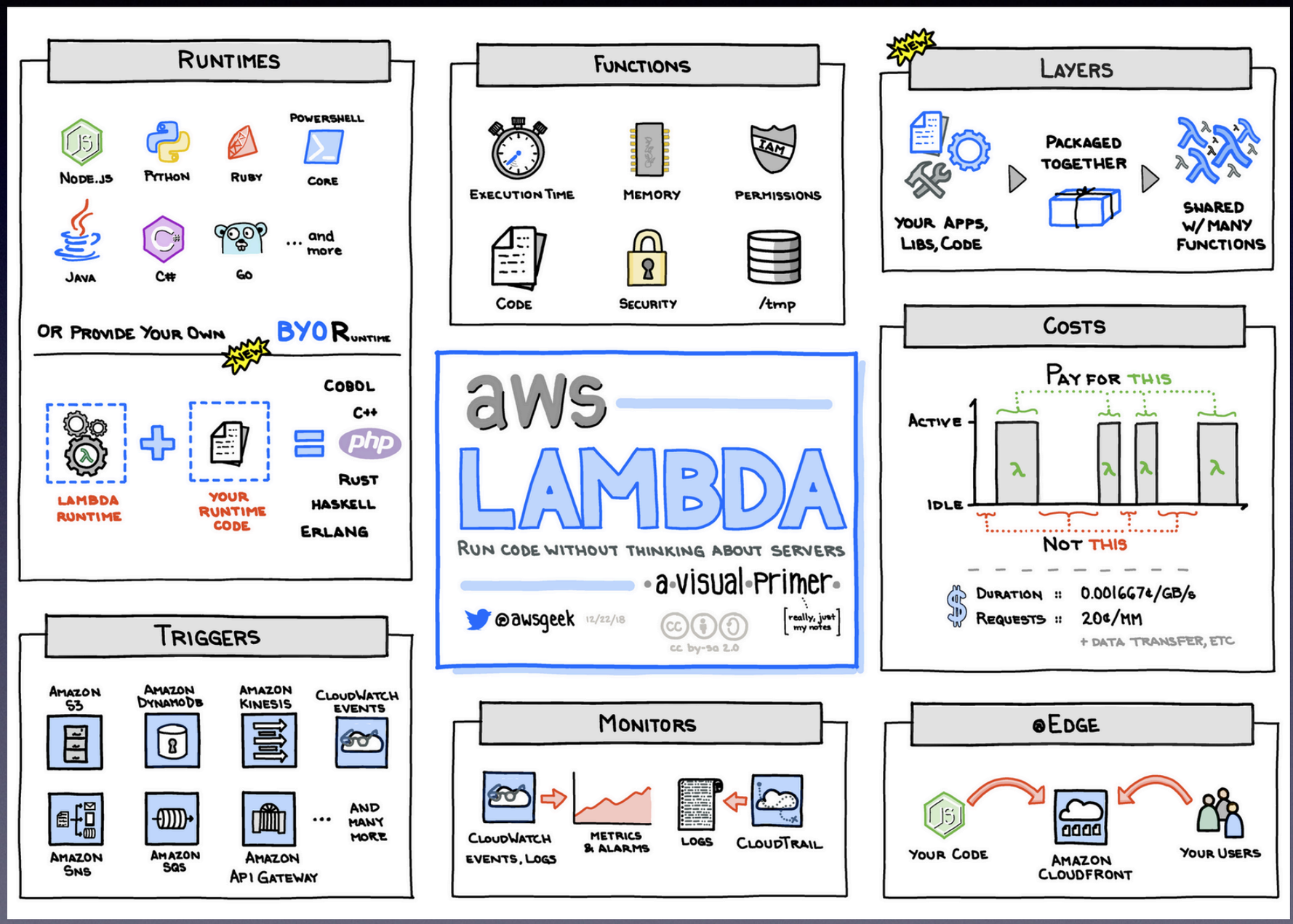
AWS Glue

AWS Serverless Application Repository

AWS X-Ray

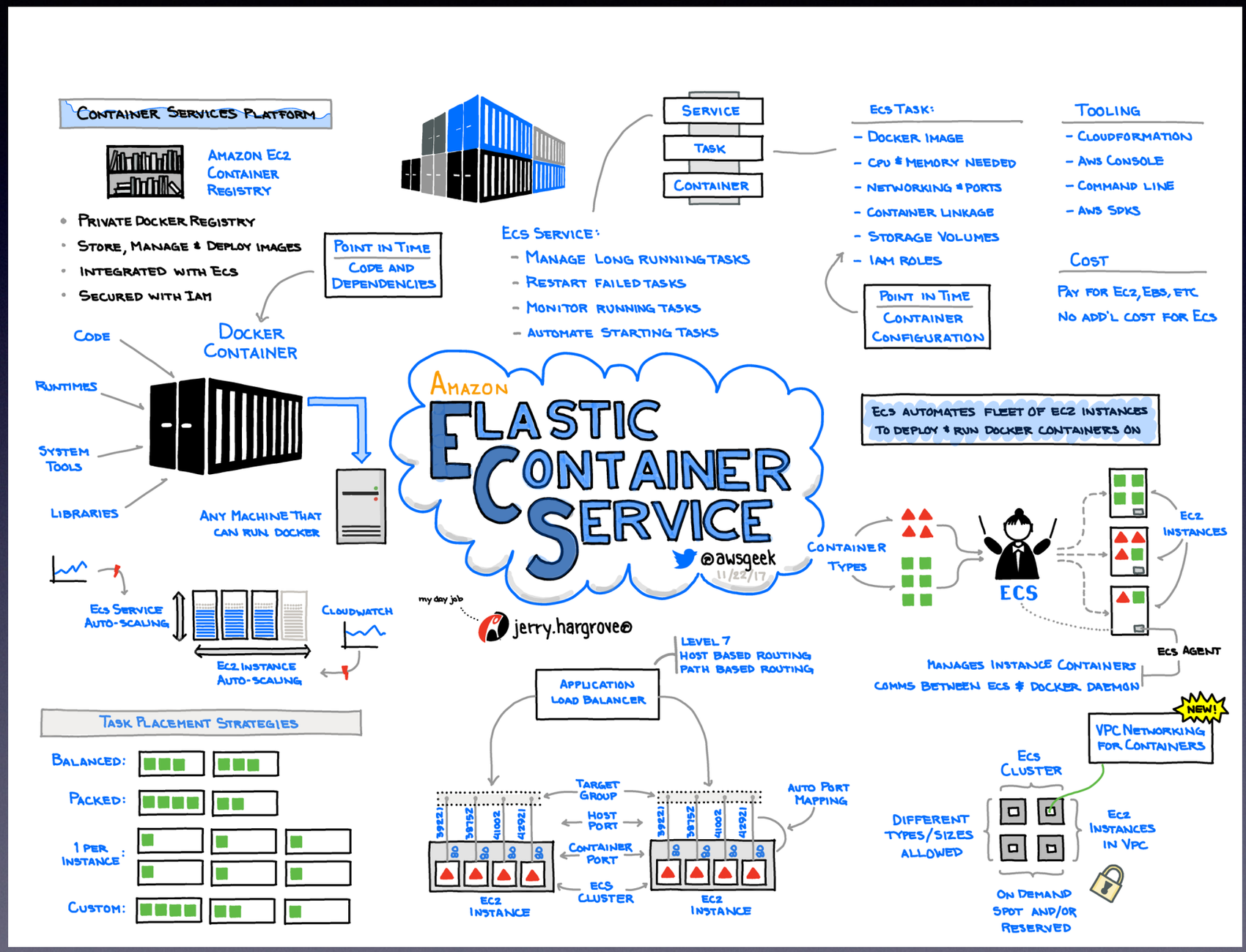
AWS Cloud9

AWS Geek (re)View



https://www.awsgeek.com/

AWS Geek (re)View 2



https://www.awsgeek.com/

Lambda Layers Demo

https://github.com/awslabs/aws-lambda-container-image-converter

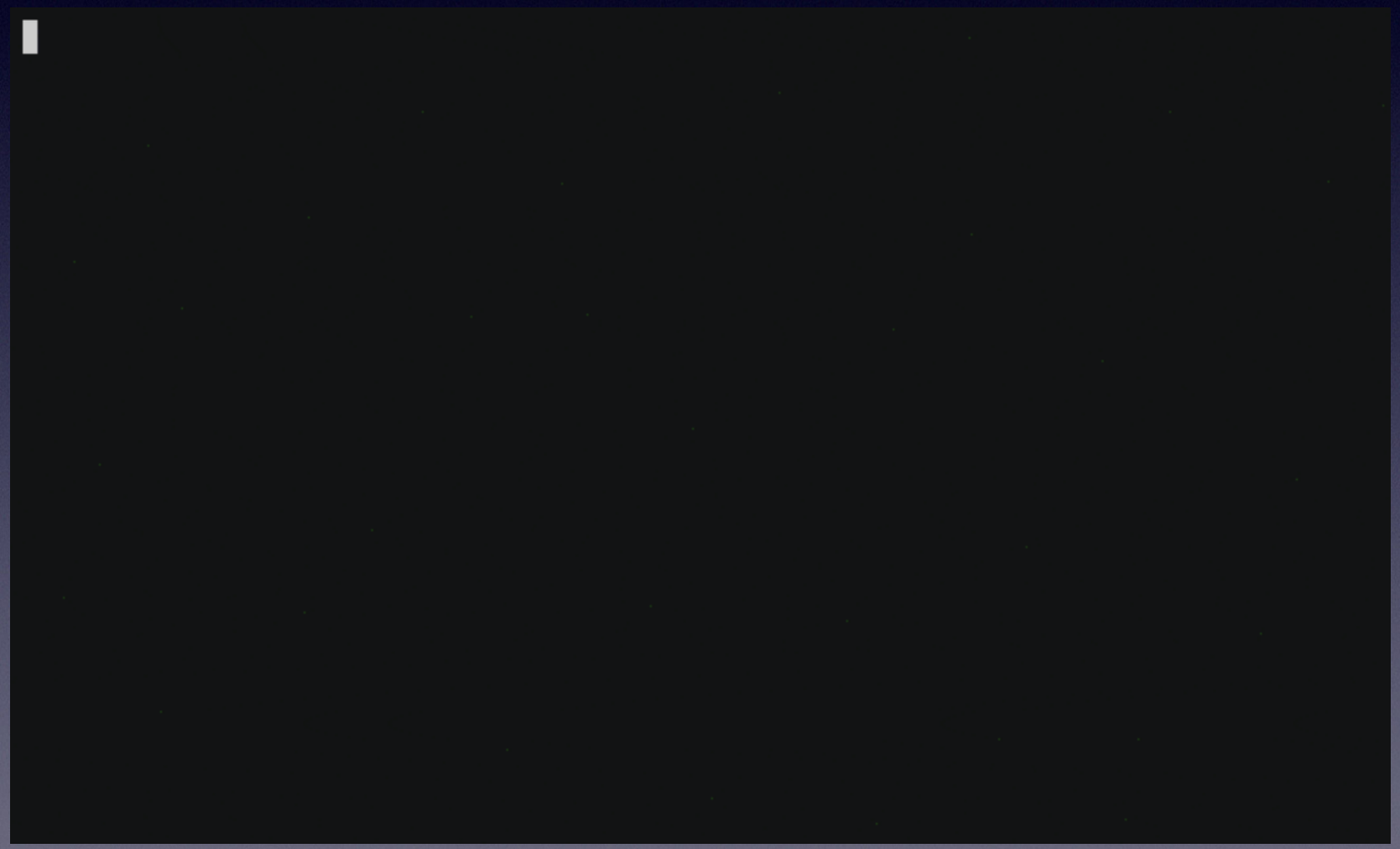
```
$ # Welcome to img2lambda!
```

https://github.com/clareliguori

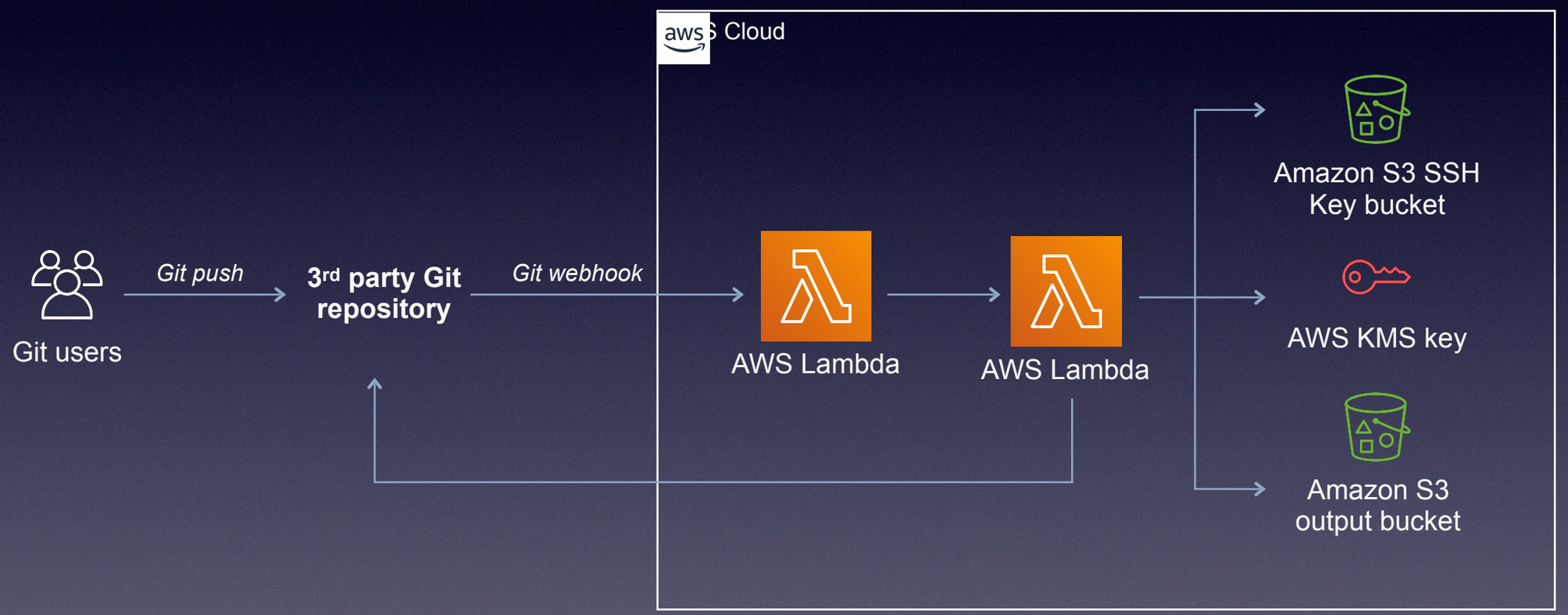
https://twitter.com/clare_liguori

CDK Demo

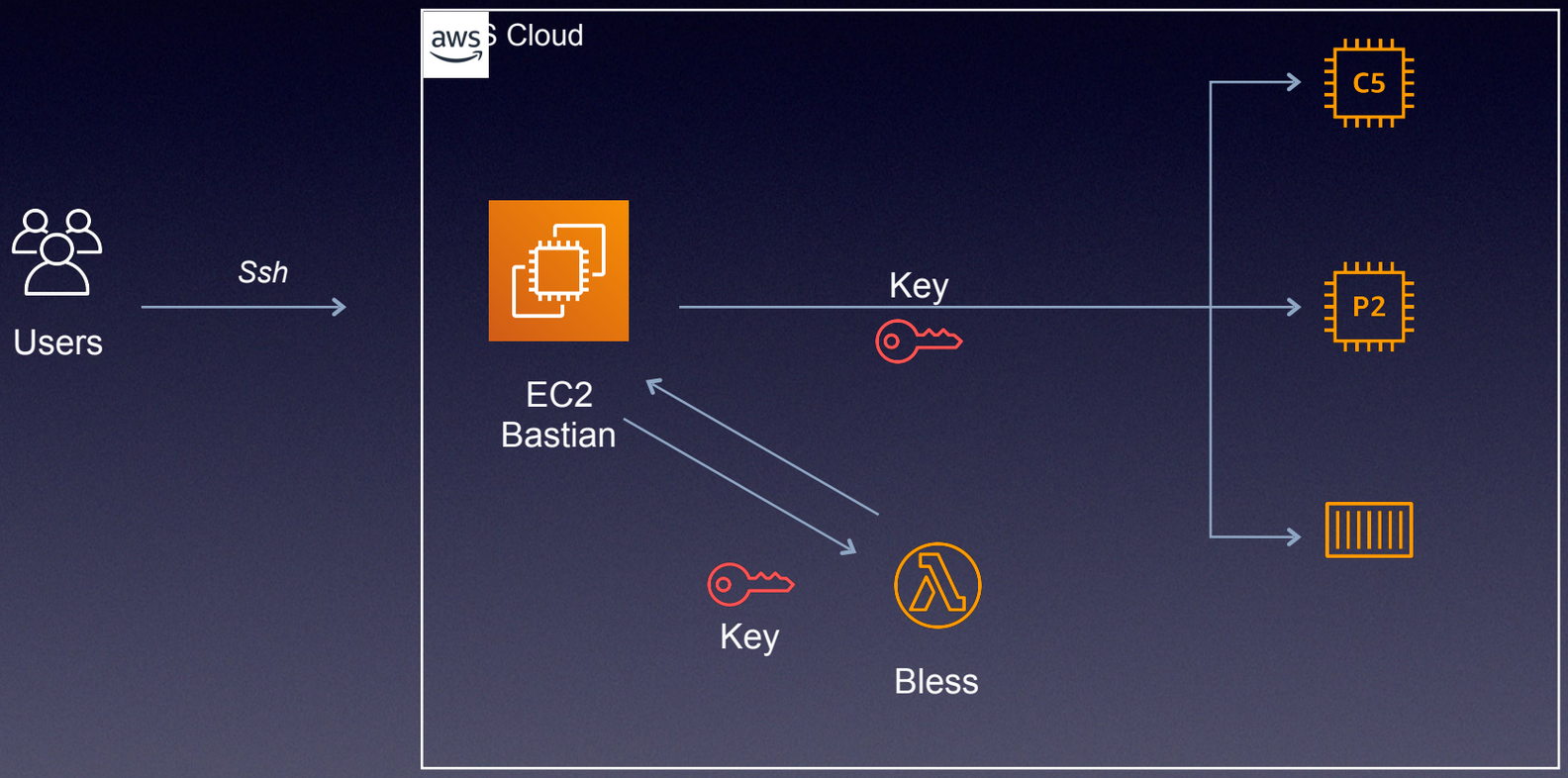
https://github.com/awslabs/aws-cdk



Git hook example



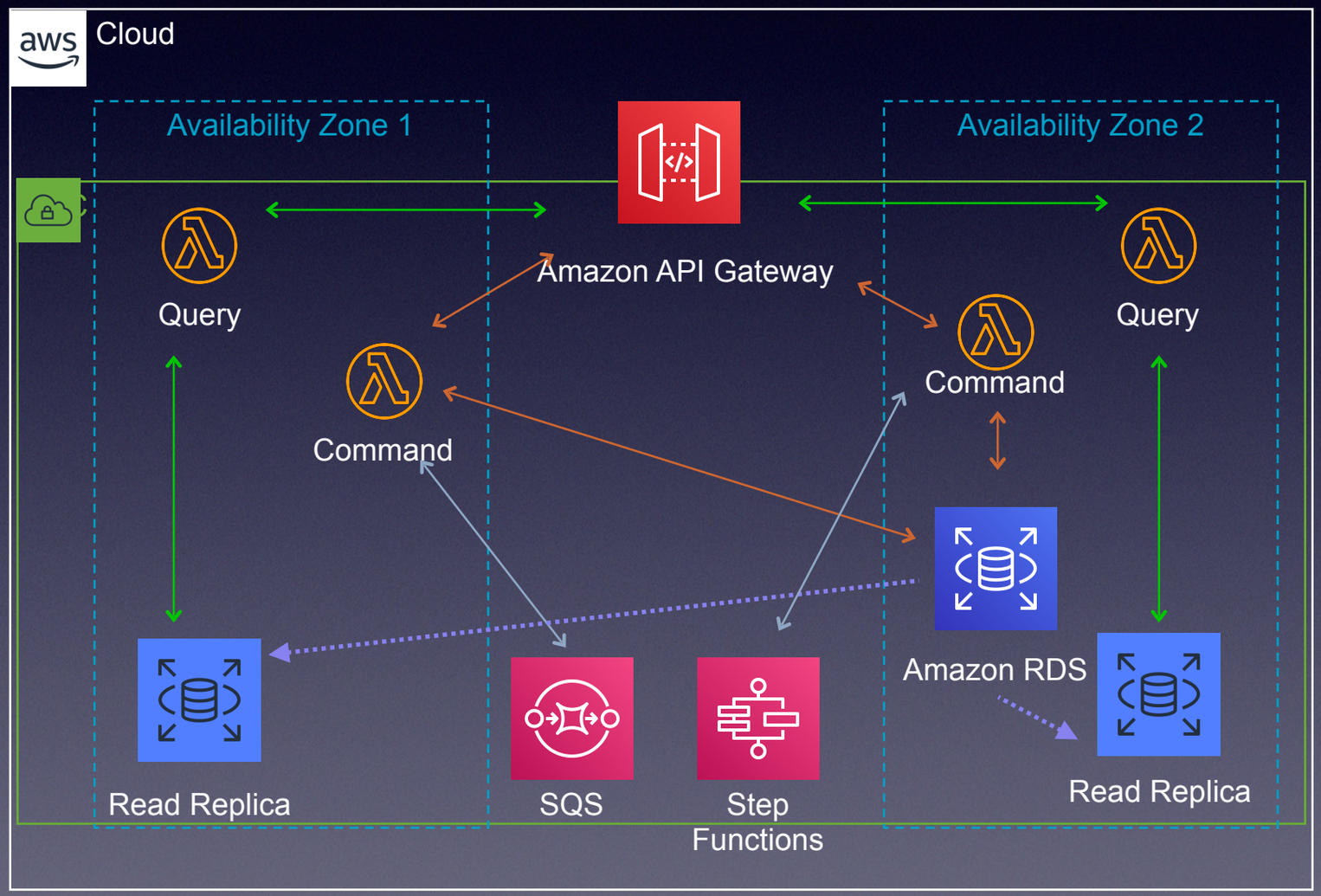
Netflix BLESS



Certificate Authority for temporary SSH auth tokens

https://github.com/Netflix/bless

CQRS Example



https://github.com/jpoley/serverless-cqrs

help contribute…

Resources

https://github.com/awesome-lambda/awesome-lambda

https://github.com/aws/aws-xray-sdk-java

https://github.com/turnerlabs/terraform-ecs-fargate

https://github.com/turnerlabs/fargate

https://github.com/turnerlabs/fargate-create

http://pywren.io/

https://ecsworkshop.com/