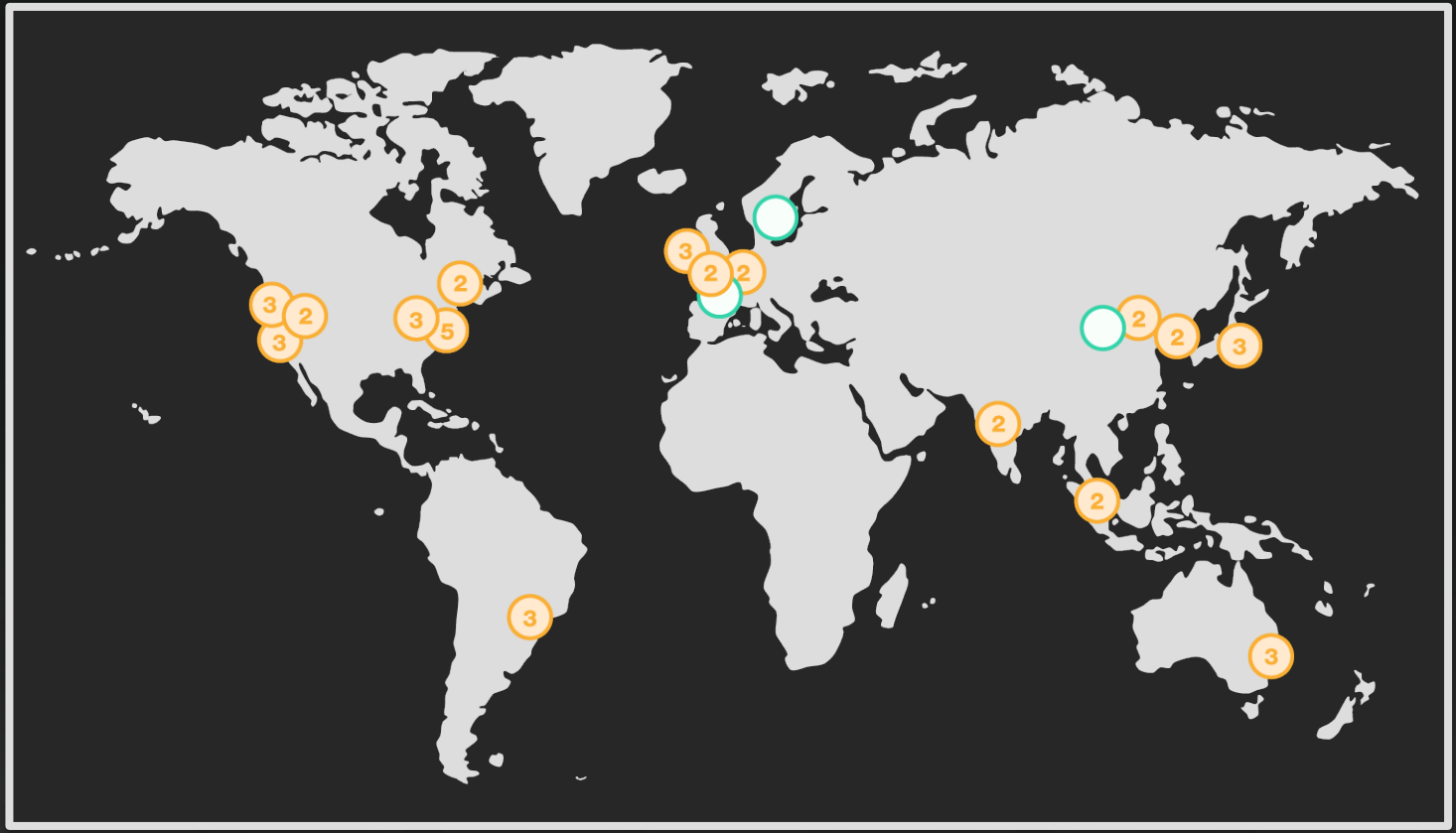
From Cloud Computing to Edge Computing

Julien Simon, Principal Technical Evangelist
Amazon Web Services

julsimon@amazon.com
@julsimon

The AWS Cloud: 16 Regions, 42 Availability Zones



https://aws.amazon.com/about-aws/global-infrastructure/

amazon web services

The AWS Edge: 74 Locations

Ashburn, VA (3), Atlanta GA (3), Chicago, IL, Dallas/Fort Worth, TX (2), Hayward, CA, Jacksonville, FL, Los Angeles, CA (2), Miami, FL, Minneapolis, MN, New York, NY (3), Newark, NJ, Palo Alto, CA, Philadelphia, PA, San Jose, CA, Seattle, WA, South Bend, IN, St. Louis, MO, Montreal, QC, Toronto, ON

Rio de Janeiro, Brazil, São Paulo, Brazil (2)

Amsterdam, The Netherlands (2), Berlin, Germany, Dublin, Ireland, Frankfurt, Germany (5), London, England (4), Madrid, Spain, Marseille, France, Milan, Italy, Munich, Germany, Paris, France (2), Prague, Czech Republic, Stockholm, Sweden, Vienna, Austria, Warsaw, Poland Zurich, Switzerland.

Chennai, India, Hong Kong, China (3), Manila, the Philippines, Melbourne, Australia, Mumbai, India (2), New Delhi, India, Osaka, Japan, Seoul, Korea (3), Singapore (2), Sydney, Australia, Taipei, Taiwan, Tokyo, Japan (3)

https://aws.amazon.com/about-aws/global-infrastructure/

amazon web services

Amazon CloudFront delivers ALL types of content

SSL

Dynamic

Static

User Input

Video

https://www.amazon.com/gp/css/homepage.html/ref=gno_yam_ya

amazon Prime

Chris's Amazon.com | Today's Deals | Gift Cards | Sell | Help

Shop by Department

Search All Go

Buy a Kindle | Kindle eBooks | Advanced Search | Daily Deals | Free Reading Apps | Kindle Singles | Newsstand | Accessories

Revolutionary on-device tech support

Exclusively on Kindle Fire HDX tablets—live on-device tech support from an Amazon expert is just a tap away with the new "Mayday" button

Live Support with Mayday

NEW—Simply tap the "Mayday" button to be connected for free to an Amazon expert who can co-pilot you through any feature by drawing on your screen, walking you through how to do something yourself, or doing it for you—whatever works best. Mayday is available 24x7, 365 days a year, and it's free. Throughout the process, you will be able to see your Amazon Tech advisor live on your screen, but they won't see you. 15 seconds or less is the Mayday response time goal.

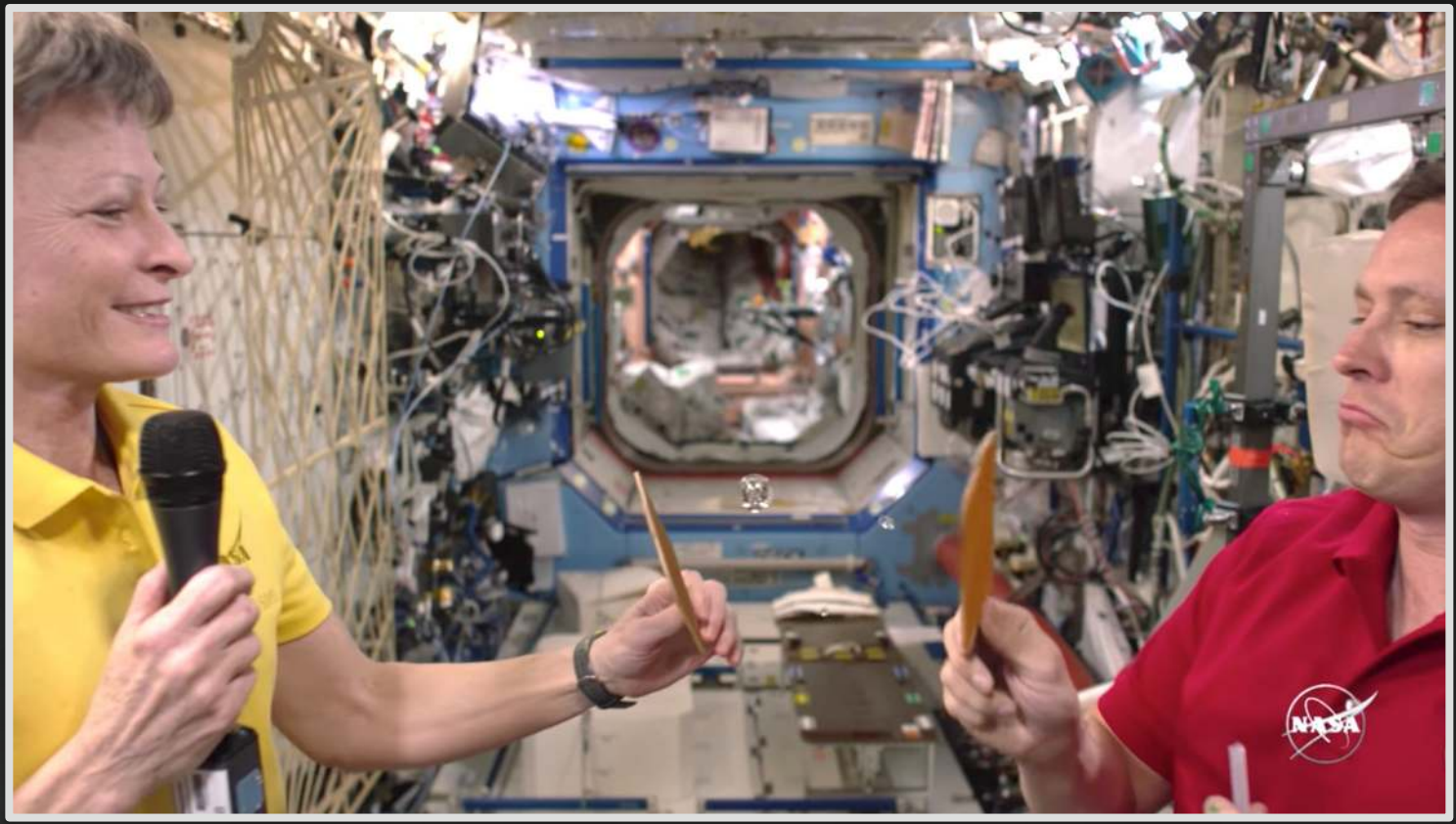
Watch it in Action

Kindle Fire HDX is easy to use right out of the box. But when you need extra assistance, the "Mayday" button is there. See how it works in these short commercials.

https://aws.amazon.com/cloudfront/

amazon web services

NASA's First-Ever 4K Live Stream from Space



https://live.awsevents.com/nasa4k

amazon web services

Edge Locations help secure your platform

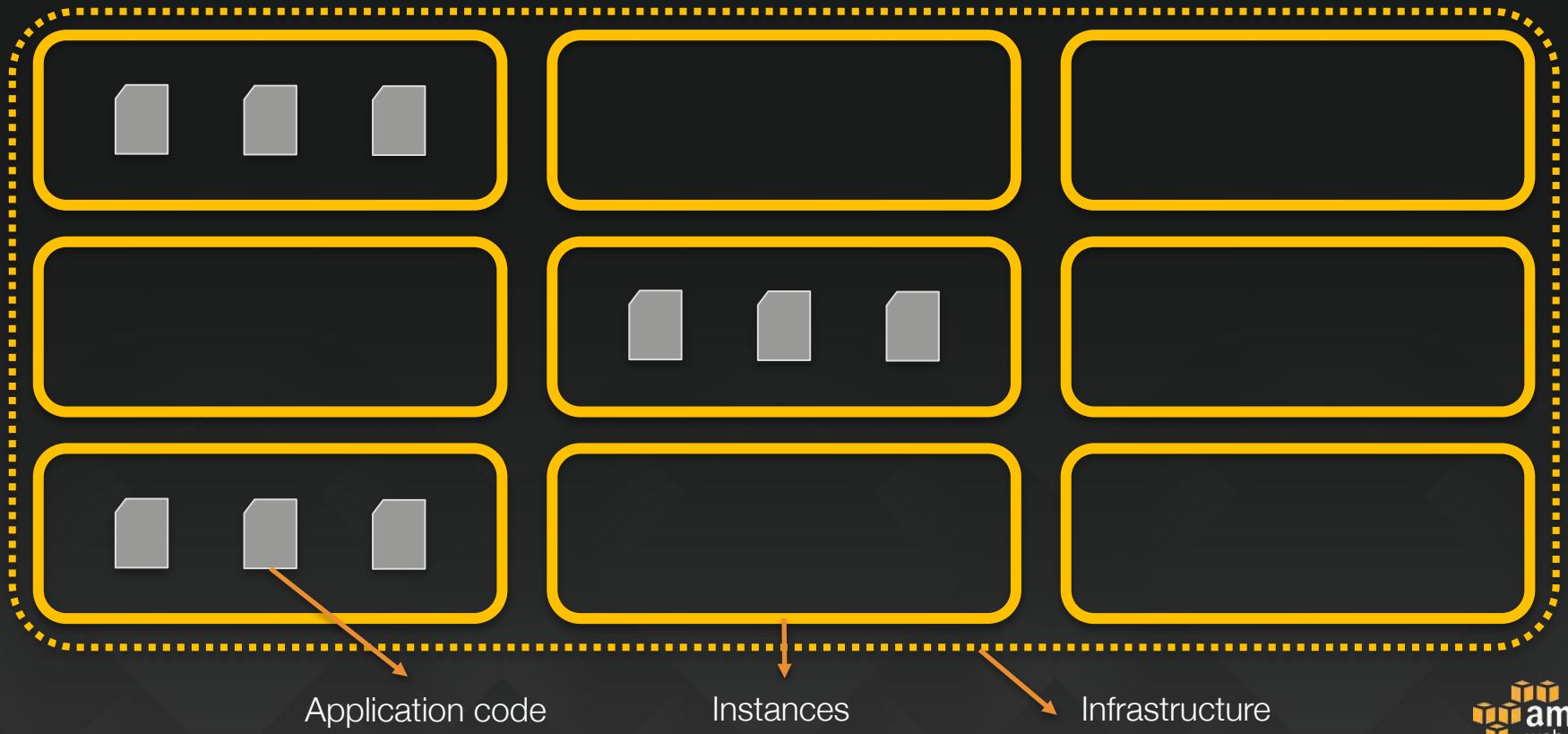
https://aws.amazon.com/waf
https://aws.amazon.com/shield/

amazon web services

What about code?

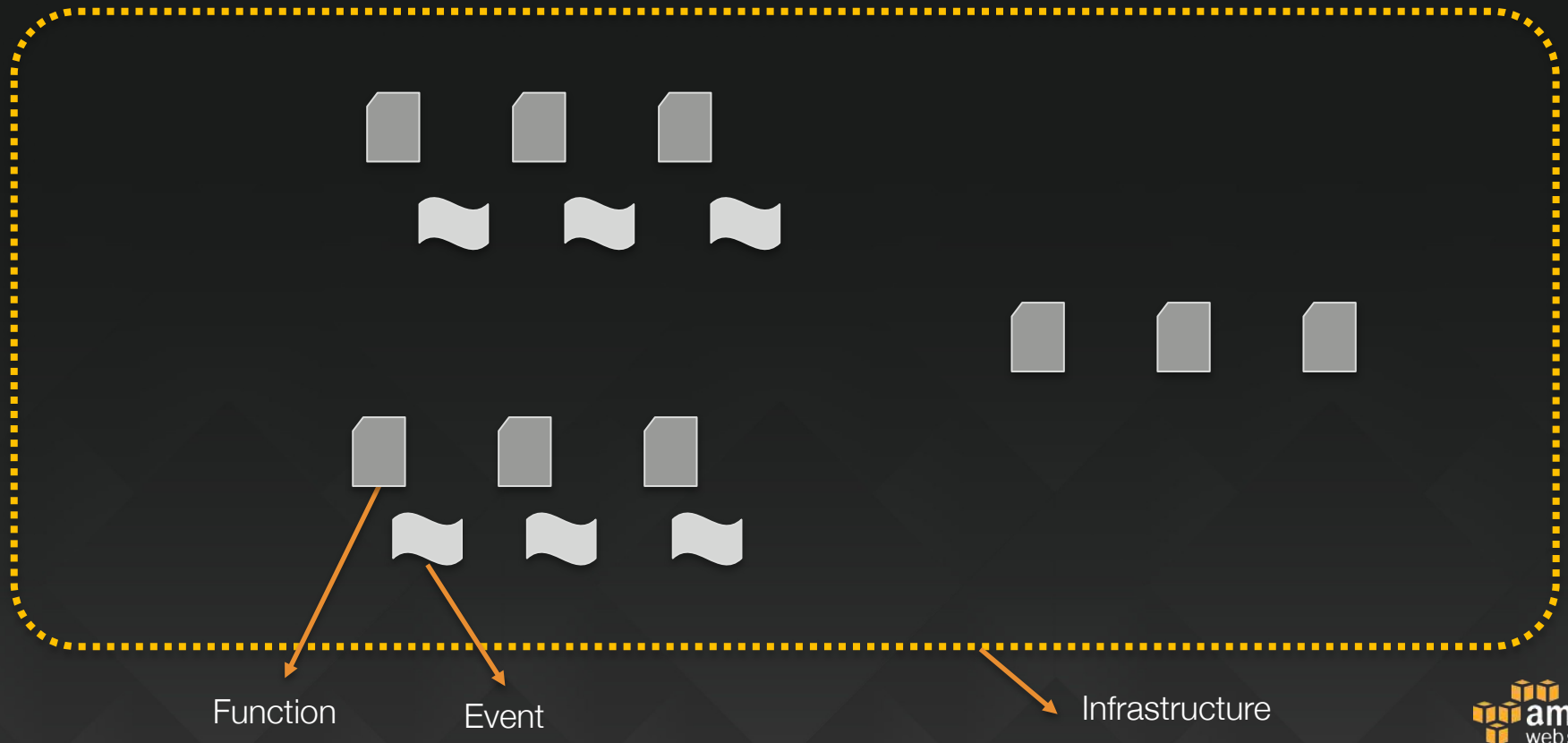
amazon
web services

Evolution of Compute – Public Cloud



amazon web services

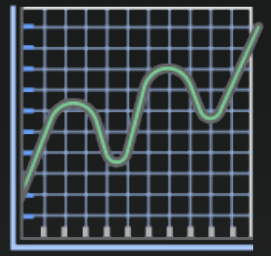
Evolution of Compute – Serverless



amazon web services

Benefits of Serverless

No servers to manage

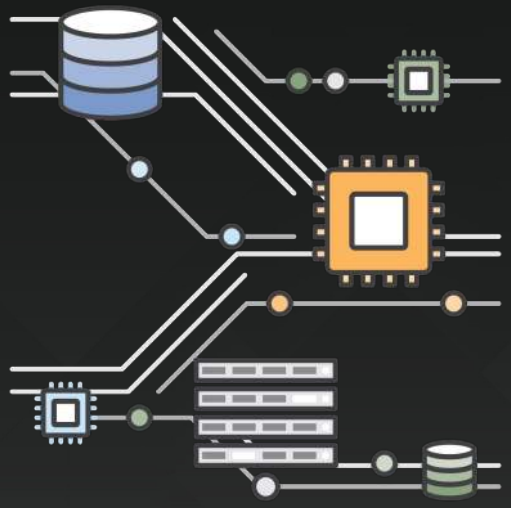


Continuous scaling

Never pay for idle – no cold servers

AWS Lambda: Serverless computing

- Run code without servers: Node.js, Python, Java, C#
- Triggered by events or called from APIs:
  - PUT to an Amazon S3 bucket
  - Updates to Amazon DynamoDB table
  - Call to an Amazon API Gateway endpoint
  - Mobile app back-end call
  - CloudFront requests
  - And many more…
- Makes it easy to:
  - Perform real-time data processing
  - Build scalable back-end services
  - Glue pieces of AWS infrastructure

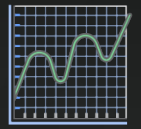


amazon web services

Running code at Edge Locations: Lambda@Edge

- Lambda@Edge is an extension of AWS Lambda that allows you to run your Node.js code at AWS Edge Locations.
- Customize your content very close to your users, improving the end-user experience.

No servers to manage



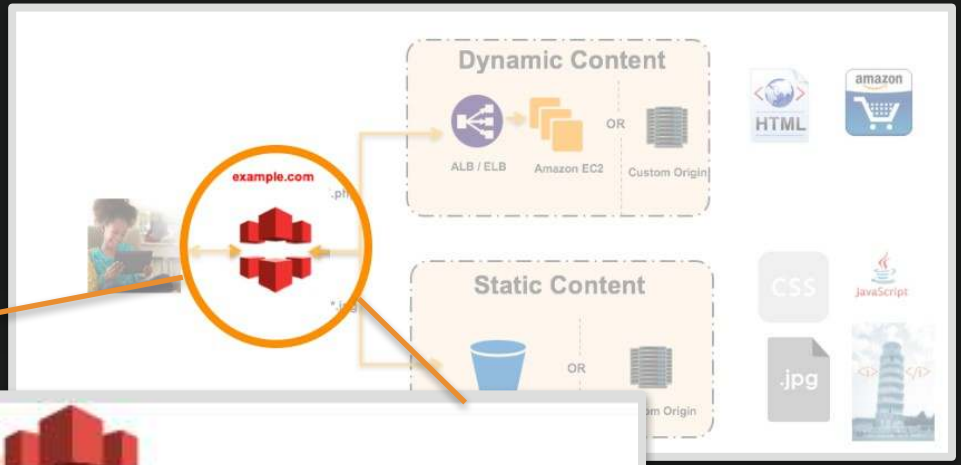
Continuous scaling

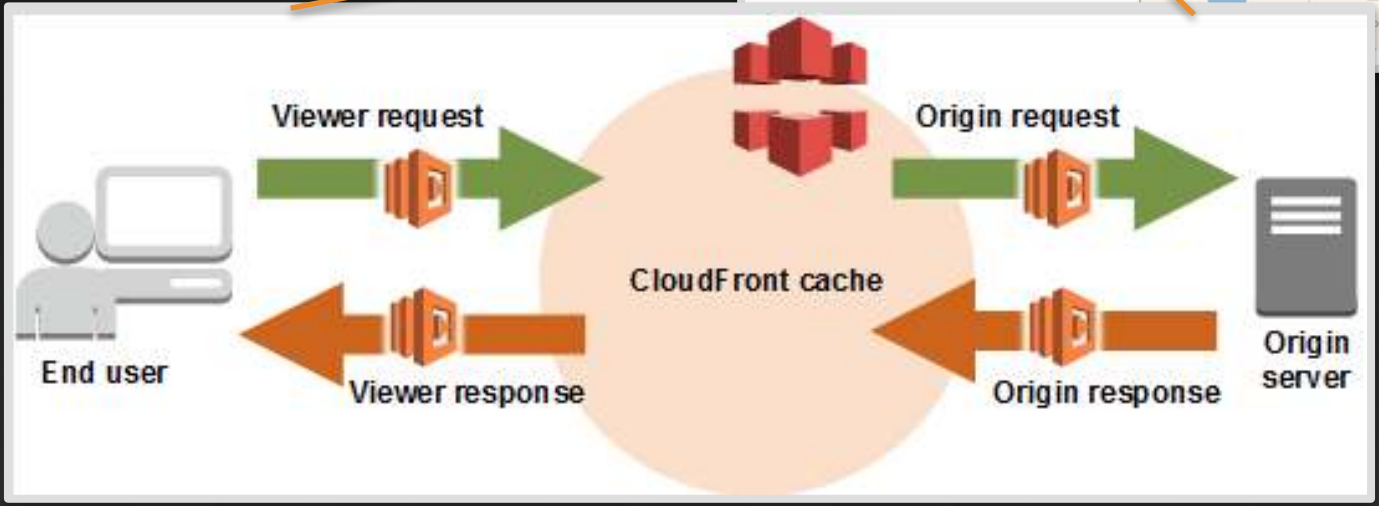
Globally distributed

Never pay for idle – no cold servers

amazon web services

CloudFront Triggers for Lambda@Edge Functions





amazon web services

Write Once, Deploy Everywhere

Ashburn, VA (3), Atlanta GA (3), Chicago, IL, Dallas/Fort Worth, TX (2), Hayward, CA, Jacksonville, FL, Los Angeles, CA (2), Miami, FL, Minneapolis, MN, New York, NY (3), Newark, NJ, Palo Alto, CA, Philadelphia, PA, San Jose, CA, Seattle, WA, South Bend, IN, St. Louis, MO, Montreal, QC, Toronto, ON

Rio de Janeiro, Brazil, São Paulo, Brazil (2)

Amsterdam, The Netherlands (2), Berlin, Germany, Dublin, Ireland, Frankfurt, Germany (5), London, England (4), Madrid, Spain, Marseille, France, Milan, Italy, Munich, Germany, Paris, France (2), Prague, Czech Republic, Stockholm, Sweden, Vienna, Austria, Warsaw, Poland Zurich, Switzerland.

Chennai, India, Hong Kong, China (3), Manila, the Philippines, Melbourne, Australia, Mumbai, India (2), New Delhi, India, Osaka, Japan, Seoul, Korea (3), Singapore (2), Sydney, Australia, Taipei, Taiwan, Tokyo, Japan (3)

https://aws.amazon.com/about-aws/global-infrastructure/

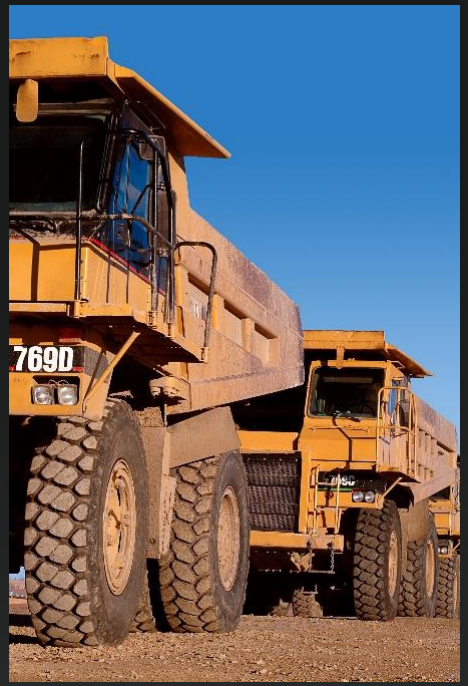
amazon web services

What about everywhere else?

amazon
web services

Most machine-generated data never reaches the cloud

Medical equipment

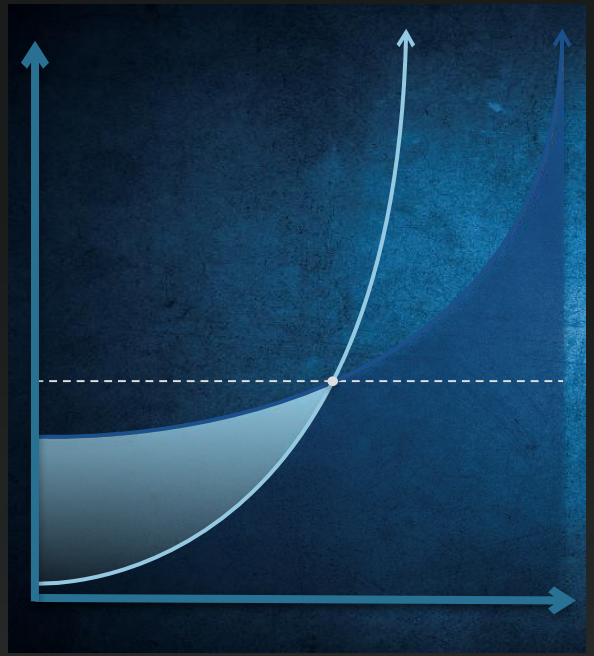


Industrial machinery

Extreme environments

This problem isn’t going away

Law of physics



Law of economics

Law of the land

Our customers need to…

Extend their data center

Write data directly when it's generated

Process data

Encrypted, secure, and embedded compute

Expedite move

A fast and cost effective way to ensure data can be quickly transferred to and from the cloud

Simplify data transfer

Use standard and familiar tools for the data transfer process

amazon web services

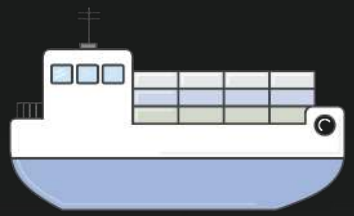
AWS Snowball Edge

Petabyte-scale hybrid device with onboard compute and storage

- 100 TB local storage
- Local compute equivalent to an Amazon EC2 m4.4xlarge instance
- 10GBase-T, 10/25Gb SFP28, and 40Gb QSFP+ copper, and optical networking
- Ruggedized and rack-mountable

amazon web services

AWS Snowball Edge use cases



Offline Staging

IoT

Local Tiering and Compute

Local Transformation

The Philips IntelliSpace Console relies on Snowball Edge

Philips HealthSuite

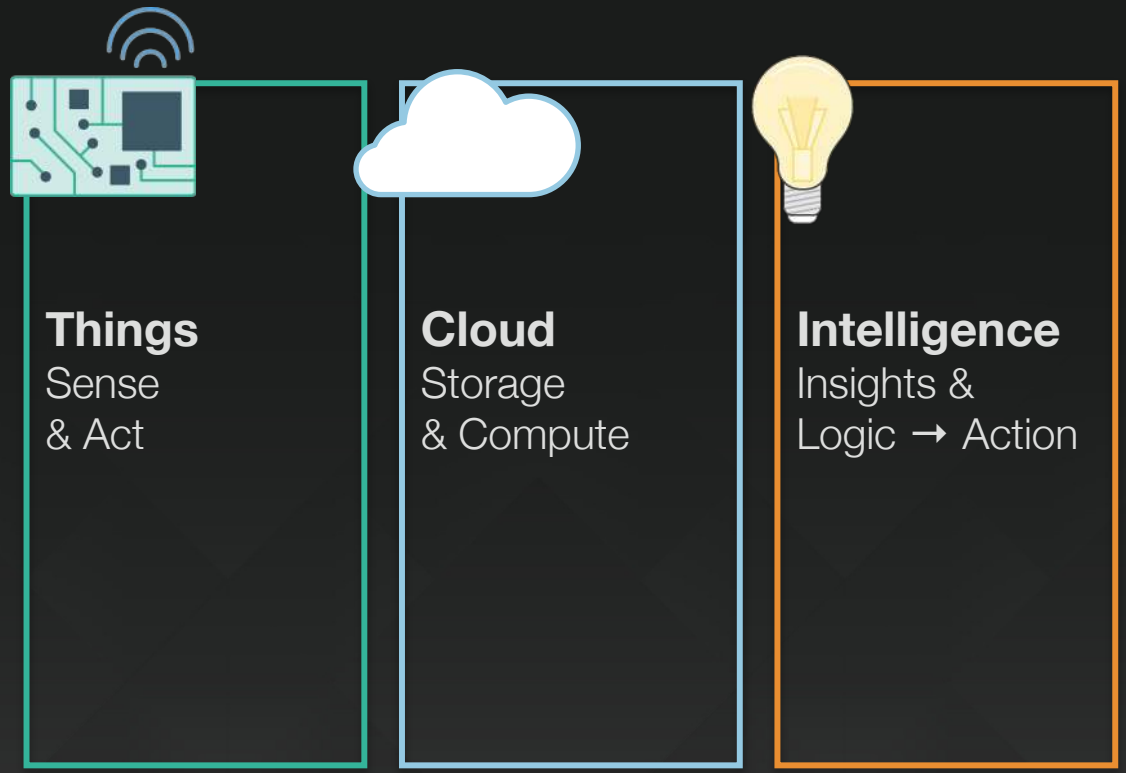
- Aggregates and stores 1200+ ICU patient data points per day
- Uses Lambda for data transformation
- Performs real-time analysis
- Keeps running even if hospital faces an IT / network outage

http://www.usa.philips.com/healthcare/product/HCNOCTN501

amazon web services

What about constrained devices?

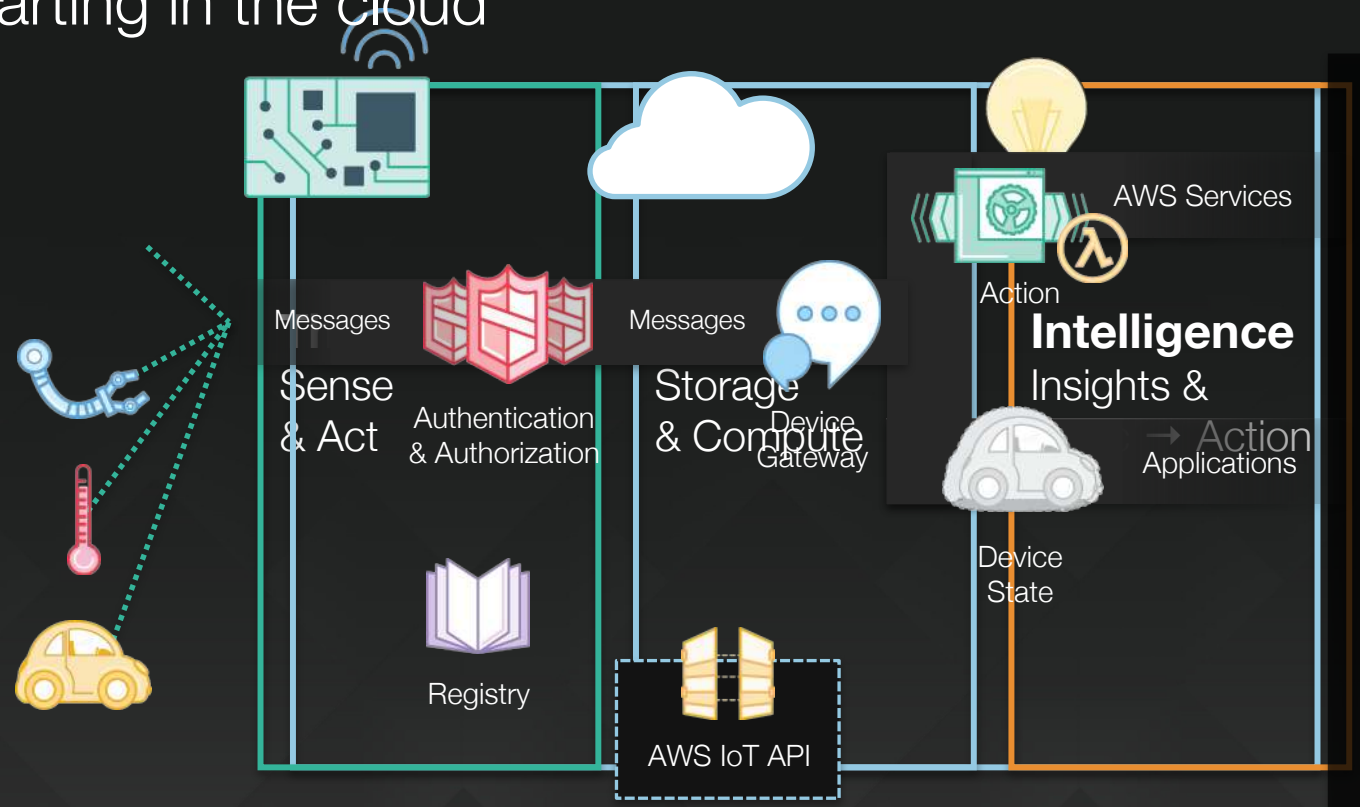
Three pillars of IoT



amazon web services

AWS IoT

Starting in the cloud



https://aws.amazon.com/iot/

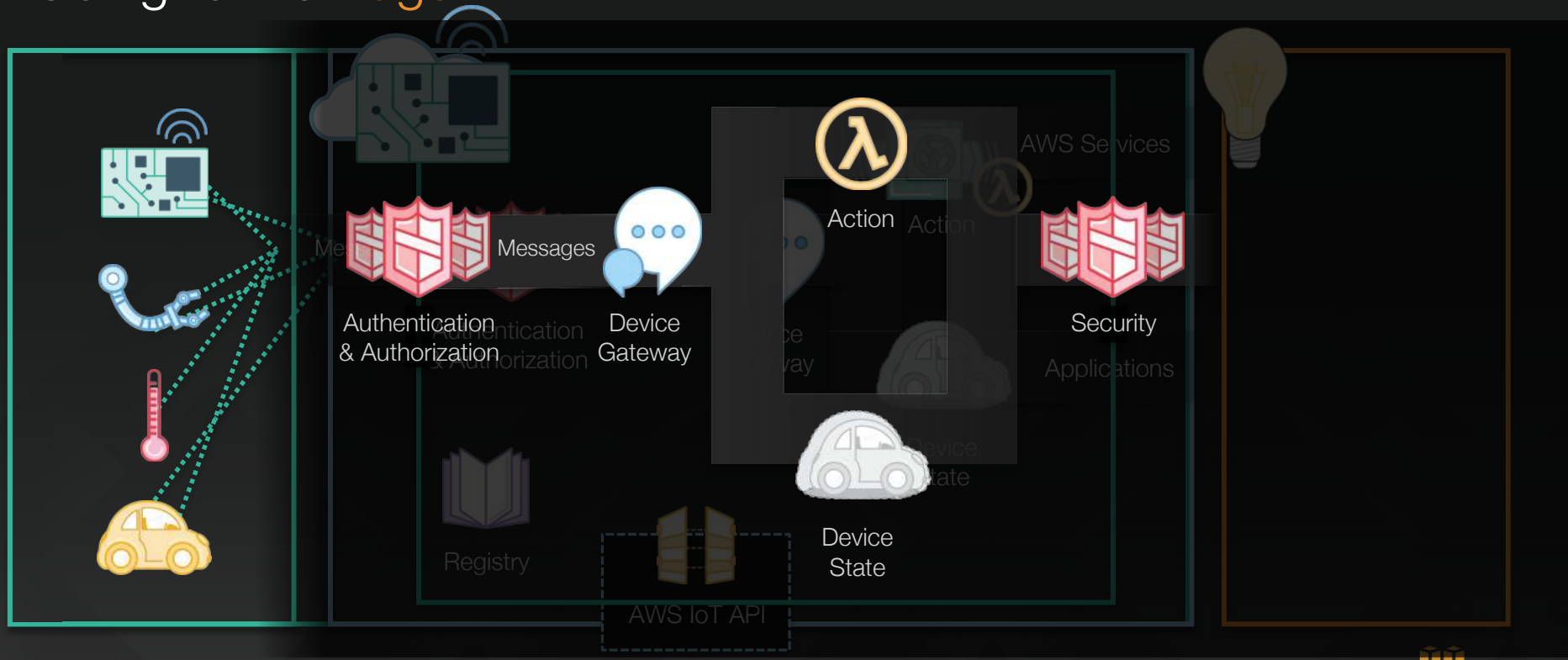
AWS Greengrass

amazon web services

*Note: Greengrass is NOT Hardware (You bring your own)

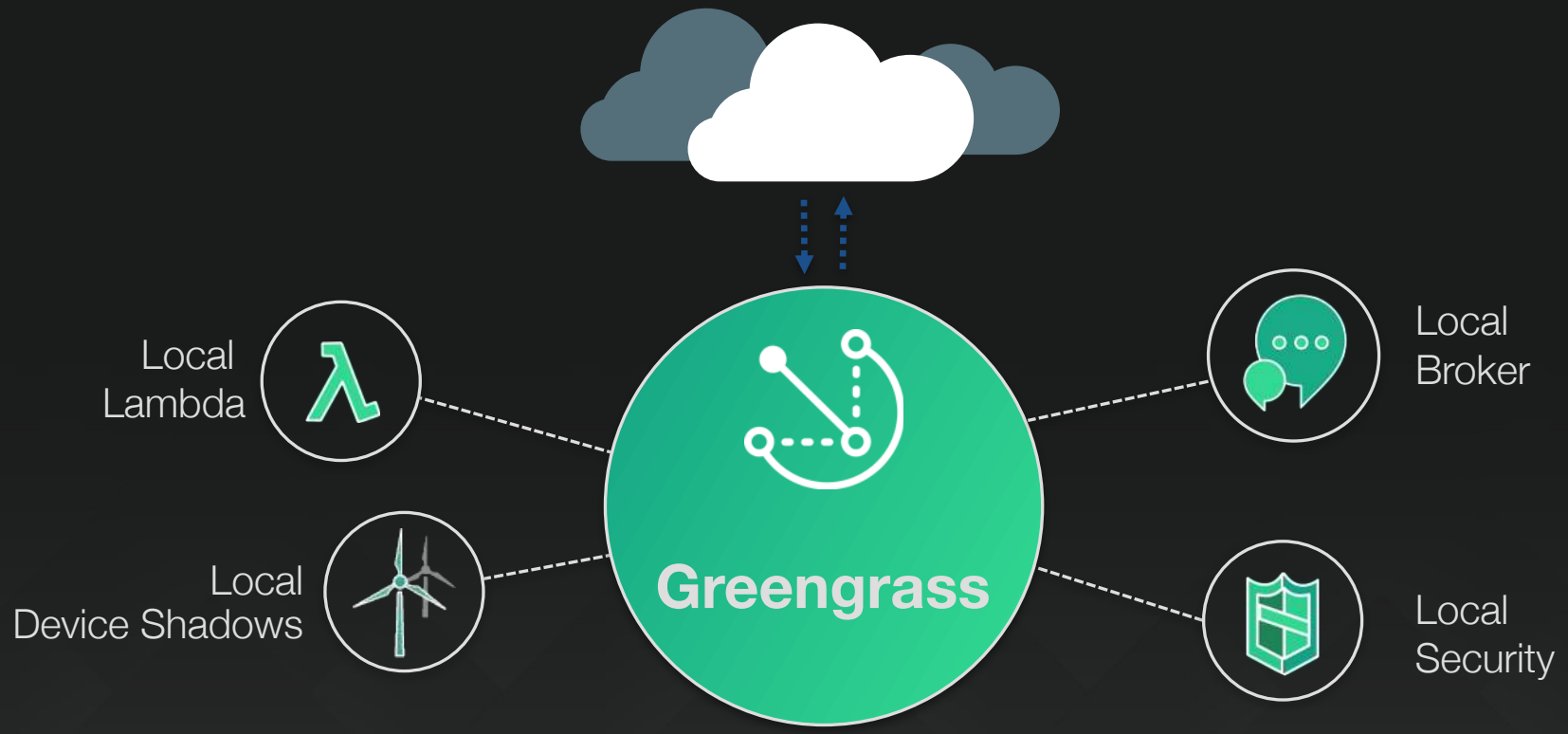
AWS Greengrass

Going to the Edge



https://aws.amazon.com/greengrass/

amazon web services



amazon
web services

Benefits of AWS Greengrass

Respond to local events quickly

Operate **offline**

Simplified device programming

Reduce the cost of IoT applications

amazon web services

What about AI at the Edge?

amazon
web services

Amazon Echo

amazon web services

Deep Learning challenges

- Training Deep Learning models requires a lot of resources (compute & storage)
- Robots or autonomous cars can't exclusively rely on the Cloud
- #1 issue: network availability, throughput and latency
- Other issues: memory footprint, power consumption, form factor
- Need the best of both worlds
  - Elasticity and scalability in the Cloud to train models
  - Local, real-time inference on the device

amazon web services

MXNet

mxnet | Apache incubator | English | Fork me on GitHub

Flexible

Supports both imperative and symbolic programming

Portable

Runs on CPUs or GPUs, on clusters, servers, desktops, or mobile phones

Multiple Languages

Supports over 7 programming languages, including C++, Python, R, Scala, Julia, Matlab, and Javascript

Auto-Differentiation

Calculates the gradient automatically for training a model

Distributed on Cloud

Supports distributed training on multiple CPU/GPU machines, including AWS, GCE, Azure, and Yarn clusters

Performance

Optimized C++ backend engine parallelizes both I/O and computation

Resources

http://mxnet.io/
https://github.com/dmlc/mxnet
https://github.com/dmlc/mxnet-notebooks

http://www.allthingsdistributed.com/2016/11/mxnet-default-framework-deep-learning-aws.html

https://github.com/awslabs/deeplearning-cfn

amazon web services

Lambda@Edge - Content customization

Snowball Edge - Portable compute and storage

Greengrass - Local compute for IoT

MXNet - Edge-friendly Deep Learning

http://aws.amazon.com

amazon web services

Thank you!

Julien Simon, Principal Technical Evangelist
Amazon Web Services

julsimon@amazon.com
@julsimon