



# What's New from re:Invent 2019

**Vladimir Simek**

Senior Solutions Architect  
Amazon Web Services

# A Bit of re:Invent Statistics

- 8<sup>th</sup> re:Invent since 2012
- 65.000+ visitors
- 3.800+ sessions, breakouts, bootcamps, hands-on-labs
- Social program: Midnight Madness, 20+ receptions, 4K/8K run, Bingo Night, Board Game Night, Movie Night, Ping-pong Tournament, Broomball Tournament, Biggest Night Party – re:Play
- Partners Expo, Builders Fair, and many more ...
- 4 executive keynotes, with live broadcast
- 70+ product announcements

# Compute

# AWS container services landscape

## Management

Deployment, Scheduling,  
Scaling & Management of  
containerized applications



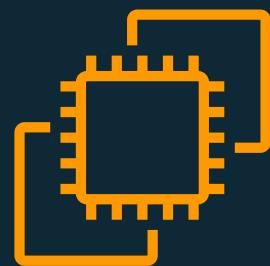
**Amazon Elastic Container Service**



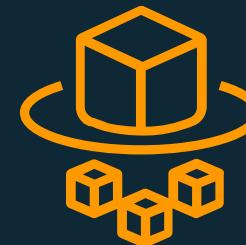
**Amazon Elastic Kubernetes Service**

## Hosting

Where the containers run



**Amazon EC2**



**AWS Fargate**

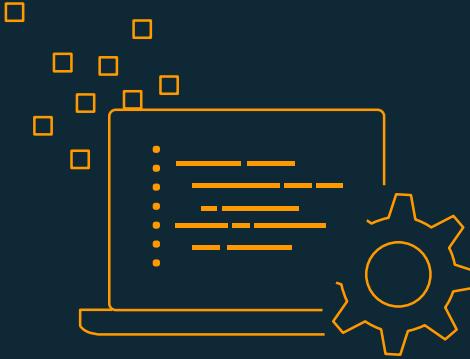
## Image Registry

Container Image Repository



**Amazon Elastic Container Registry**

# Make Kubernetes apps serverless with Amazon EKS + Fargate



## Bring existing pods

You don't need to change your existing pods. Fargate works with existing workflows and services that run on Kubernetes.



## Production ready

Launch ten or tens of thousands of pods in seconds. Easily run pods across multiple AZs for high-availability.



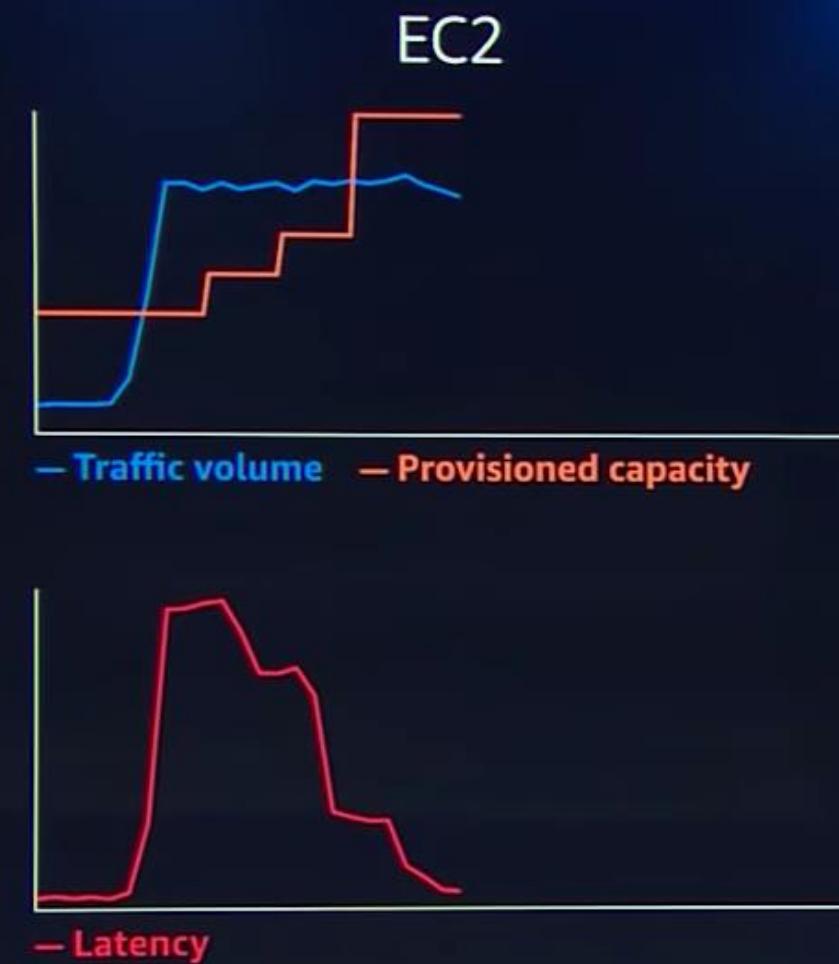
## Right-Sized and Integrated

Only pay for the resources you need to run your pods. Includes native AWS integrations for networking, and security.

---

Fargate runs tens of millions of containers for AWS customers every week

# EKS with EC2



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential



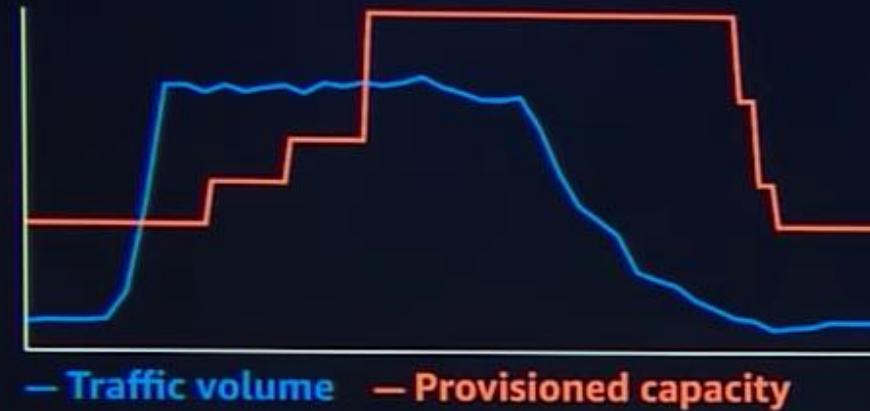
# EKS with Fargate



A full-body shot of a woman with long brown hair, wearing a vibrant purple knee-length dress with a wide collar and short sleeves. She is standing on a stage, gesturing with her hands as she speaks. The background is dark, and there are blue lights visible at the bottom of the frame.

# Comparison

EC2



Fargate



# Broadest choice of processors



---

Second generation of  
Intel® Xeon processor



---

AMD Rome



---

Graviton

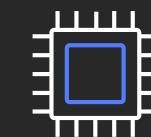
# Announcing AWS Graviton2 Processor

Enabling the best price/performance for your cloud workloads

## Graviton Processor



First Arm-based processor available in major cloud



Built on 64-bit Arm Neoverse cores with AWS-designed silicon using 16 nm manufacturing technology



Up to 16 vCPUs, 10 Gbps enhanced networking, 3.5 Gbps EBS bandwidth

## Graviton2 Processor



7x performance, 4x compute cores, and 5x faster memory



Built with 64-bit Arm Neoverse cores with AWS-designed silicon using 7 nm manufacturing technology



Up to 64 vCPUs, 25 Gbps enhanced networking, 18 Gbps EBS bandwidth

# Announcing Graviton2 based instances

Up to 40% better price-performance for general purpose, compute intensive, and memory intensive workloads.

## M6g

Built for: General-purpose workloads such as application servers, mid-size data stores, and microservices.

## C6g

Built for: Compute intensive applications such as HPC, video encoding, gaming, and simulation workloads.

## R6g

Built for: Memory intensive workloads such as open-source databases, or in-memory caches.

---

Available  
in Preview

Coming in 2020

Local NVMe-based SSD storage options will also be available in general purpose (M6gd), compute-optimized (C6gd), and memory-optimized (R6gd) instances

# ML inference deployment options on Amazon EC2

## Custom chip EC2 Inf1 instances

Applications that leverage common ML frameworks

Powered by AWS Inferentia

Best price/performance for ML inferencing in the cloud

Up to 40% lower cost per inference and up to 3x higher throughput than G4 instances

Available today!

Featuring  
AWS Inferentia

## GPU based EC2 G4 instances

Applications that require access to CUDA, CuDNN or TensorRT libraries

Amazon EC2 G4 instances based on NVIDIA T4 GPUs

Launched!

## CPU based EC2 C5 instances

Small models and low sensitivity to performance

Intel Skylake CPUs  
Support for AVX-512/  
VNNI instruction set

Launched!

# EC2 Inf1 instances feature the AWS Inferentia chip

## High performance

- Up to 2000 TOPS of throughput

## Powered by AWS Inferentia and AWS Neuron

- CNN (image/video)
- Classic LSTM/RNN
- Modern and attention-based neural network (NLP, Translation, Search and Voice)

## Flexible datatype support

- Run models trained in FP32, FP16, BF16, INT8

## Flexible framework support

- Native integration with TensorFlow, MXNet, and PyTorch

# EC2 Inf1 instances are optimized for ML inferencing



Object detection



Natural language processing



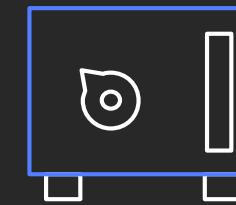
Personalization



Speech recognition



Image processing



Fraud detection



Evolution of EC2 and  
the Nitro Platform

# Journey from then to now

## 2006 “Instance”

**1.7 GHz Xeon Processor**

**1.75 GB of RAM**

**160 GB of local disk**

**250 Mbps network bandwidth**

AWS News Blog

### Amazon EC2 Beta

by Jeff Barr | on 25 AUG 2006 | [Permalink](#) | [Share](#)

Innovation never takes a break, and neither do I. From the steaming hot beaches of Cabo San Lucas I would like to tell you about the Amazon Elastic Compute Cloud, or Amazon EC2, now open for limited beta testing, with more beta slots to open soon.

Amazon EC2 gives you access to a virtual computing environment. Your applications run on a “virtual CPU”, the equivalent of a 1.7 GHz Xeon processor, 1.75 GB of RAM, 160 GB of local disk and 250 Mb/second of network bandwidth. You pay just 10 cents per clock hour (billed to your Amazon Web Services account), and you can get as many virtual CPUs as you need. You can learn more on the [EC2 Detail Page](#). We built Amazon EC2 using a virtual machine monitor by the name of [Xen](#).

Amazon EC2 works in terms of AMIs, or Amazon Machine Images. Each AMI is a pre-configured boot disk — just a packaged-up operating system stored as an [Amazon S3](#) object. There are web service calls to create images, and to assign them to virtual CPUs to run your application. If your application consists of the usual web server, business logic, and database tiers, you can build distinct AMIs for each tier, and then spawn one or more instances of each type based on the load.

In a previous post, [Sometimes You Need Just a Little...](#), I alluded to the new world of scalable, on-demand web services. In that post I talked about the fact that sometimes a little bit of storage is all you need.

Sometimes you need a lot of processing power, and sometimes you need just a little. Sometimes you need a lot, but you only need it for a limited amount of time. Perhaps you are doing some number crunching, some in-depth text processing, some scientific research, or your end-of-month accounting. Or perhaps you want to experiment with some radical new



“Your applications run on a “virtual CPU”, the equivalent of a 1.7 GHz Xeon processor, 1.75 GB of RAM, 160 GB of local disk and 250 Mbps of network bandwidth.”

# Journey from then to now

## 2006 “Instance”

**1.7 GHz** Xeon Processor

**1.75 GB** of RAM

**160 GB** of local disk

**250 Mbps** network bandwidth

## 2019

**4.0 GHz** Xeon Processor

z1d instance

**24 TiB** of RAM

High Memory instances

**60 TB** of NVMe local storage

I3en.metal instances

**48 TB** of local disk

d2.8xlarge

**100 Gbps** network bandwidth



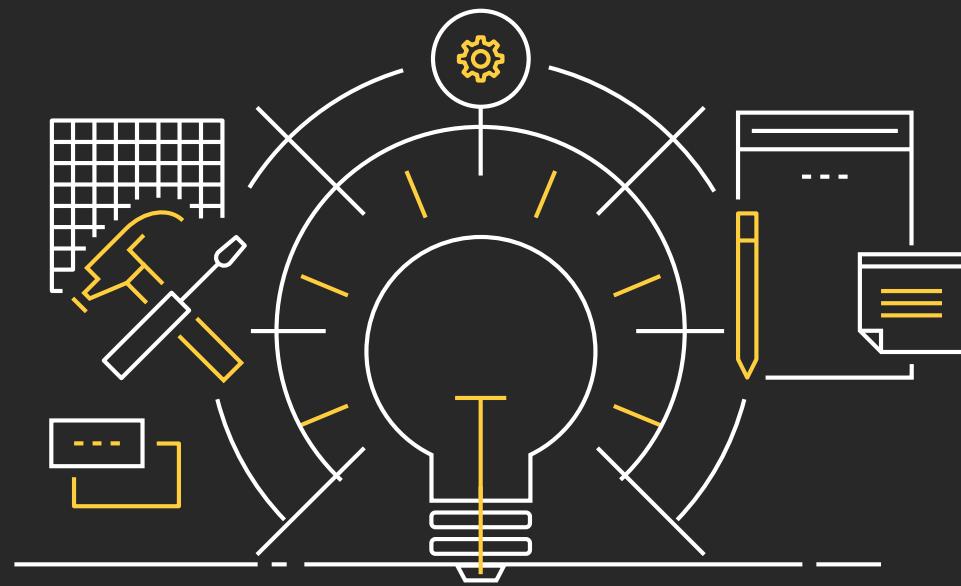
# Broadest and deepest platform choice

Categories	Capabilities	Options
General purpose	Choice of processor (AWS, Intel, AMD)	
Burstable	Fast processors (up to 4.0 GHz)	
Compute intensive	High memory footprint (up to 128 GiB)	
Memory intensive	Instance storage (HDD and NVMe)	
Storage (High I/O)	Accelerated computing (GPUs and FPGAs)	
Dense storage	Networking (up to 100 Gbps)	
GPU compute	Bare Metal	
Graphics intensive	Size (Nano to 32xlarge)	

How do you select the right instance to launch and optimize?

270 +

instance types  
for virtually every  
workload and  
business need



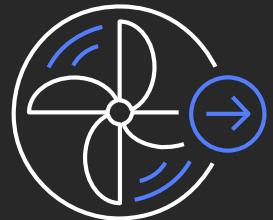
# Introducing AWS Compute Optimizer

# AWS Compute Optimizer

Recommends optimal instances for EC2 and EC2 Auto Scaling groups from 140+ instances from M, C, R, T, and X families



Lower **costs** and  
improve workload  
**performance**

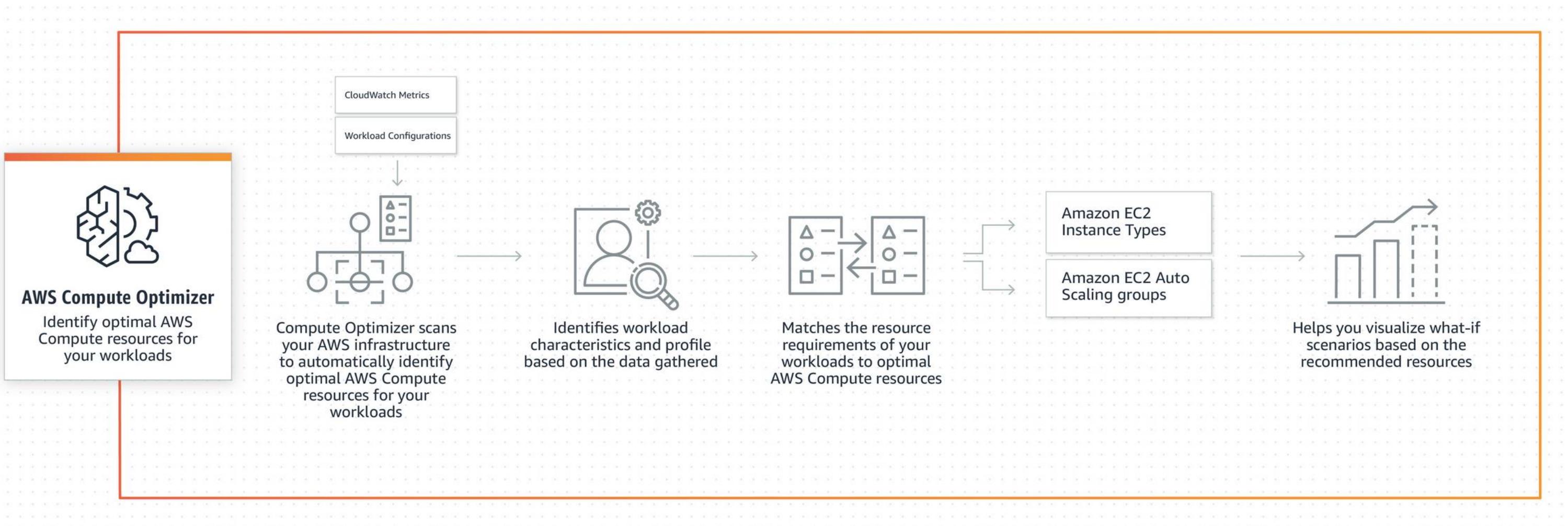


Applies **insights from millions of workloads** to make recommendations



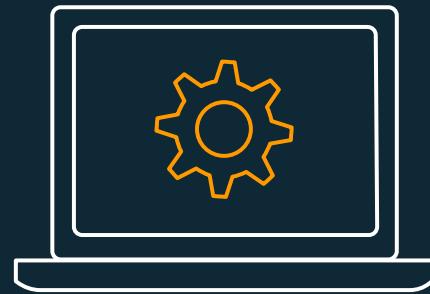
**Saves time comparing** and selecting optimal resources for your workload

# How it works



# Applications that need to remain on-premises

Applications that are sensitive to latency and variability in latency



- Need for near real time responses to end user applications
- Need to control on-site equipment
- Need to communicate with other on-premises systems

Applications that process data locally



- Need to ensure integrity of ingested signal (e.g. at live events before broadcasting)
- Need to reliably process messages from industrial equipment to monitor production
- Need for managing local data stores

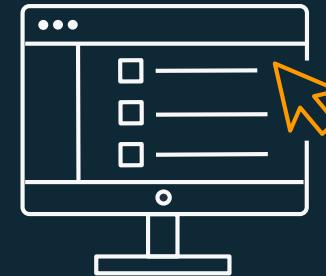
# AWS Outposts: Bringing AWS on-premises



Same AWS-designed  
**infrastructure** as in  
AWS data centers  
(built on AWS Nitro System)



Fully managed, monitored,  
and operated by AWS  
as if in AWS Regions



Single pane of management  
in the cloud providing the  
**same APIs and tools** as  
in AWS Regions

# AWS Outposts Rack

- Industry standard **42U rack**
- **Fully assembled**, ready to be rolled into final position
- **Installed by AWS**, simply plugged into power and network
- **Centralized redundant power conversion unit** and DC distribution system for higher reliability, energy efficiency, easier serviceability
- **Redundant active components** including top of rack switches and hot spare hosts



## Dimensions

- 24" Wide
- 48" Deep
- 80" Tall

# Available in 2 Variants

## Native AWS

---

AWS APIs, services, and features  
as in the AWS cloud

EC2 and EBS with support for  
services including RDS, ECS, EKS,  
EMR, ALB, others

## VMware Cloud on AWS

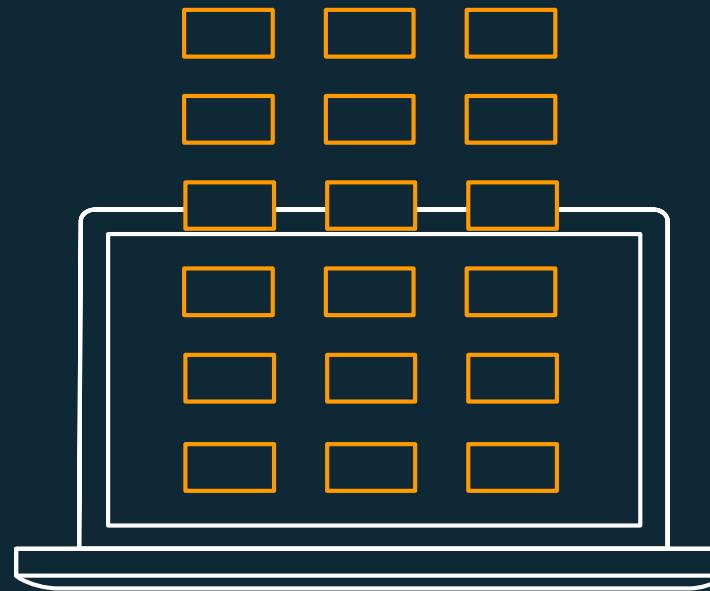
---

VMware APIs and services to  
leverage existing skills, automation,  
and governance policies

For customers running VMware  
SDDC on-premises

# Run AWS services locally

Available at GA or soon after



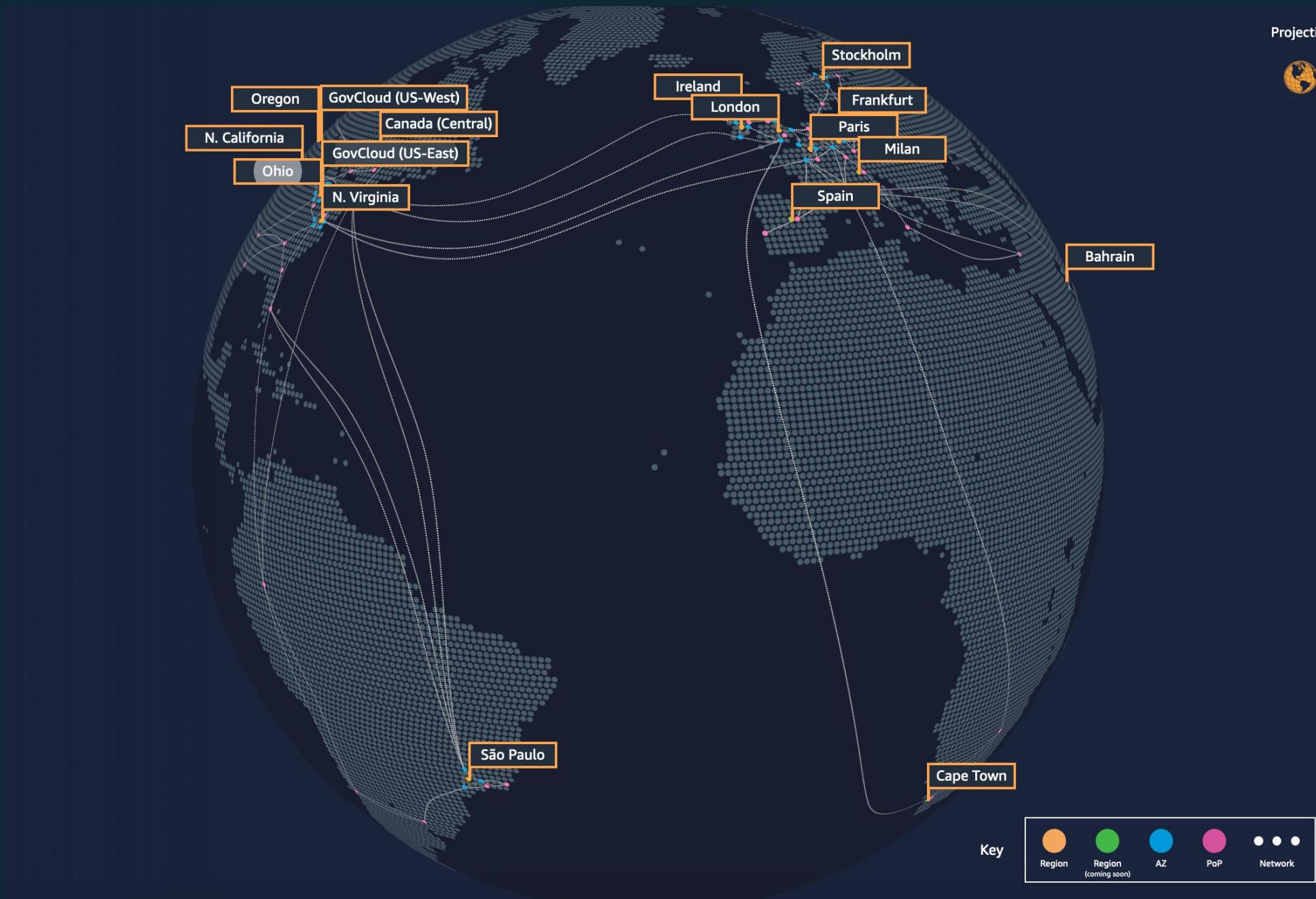
- **Compute & Storage**—Amazon EC2 instances and EBS volumes
- **Networking**—Amazon VPC
- **Database**—Amazon Relational Database Service (RDS)
- **Containers**—Amazon Elastic Container Service (ECS) & Amazon Elastic Kubernetes Service (**EKS**)
- **Data Processing**—Amazon Elastic Map Reduce (EMR)

# AWS Global Infrastructure



22 Regions  
69 Availability Zones

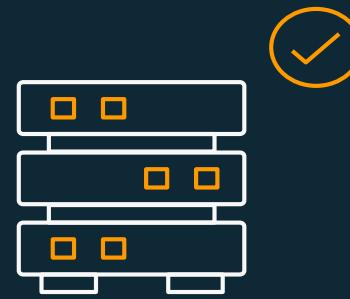
# Latency-sensitive Workloads



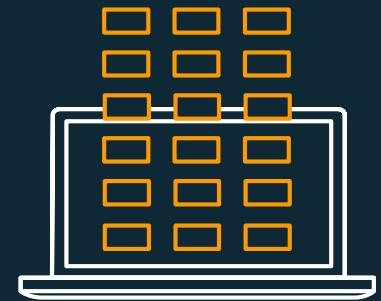
Need increasing amounts of AWS Infrastructure and services closer to end-users

# AWS Local Zones

- New type of AWS infrastructure deployment
- Places compute, storage, database, and other services closer to customers
- For demanding applications that require single-digit latencies



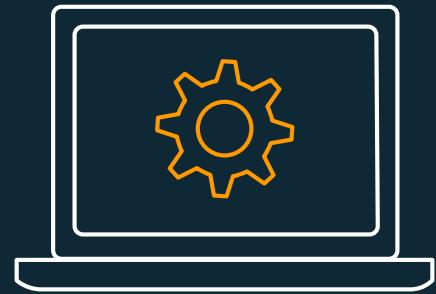
AWS  
infrastructure at  
the edge



Local compute,  
storage,  
database, and  
other services



Connect to  
services in AWS  
Regions



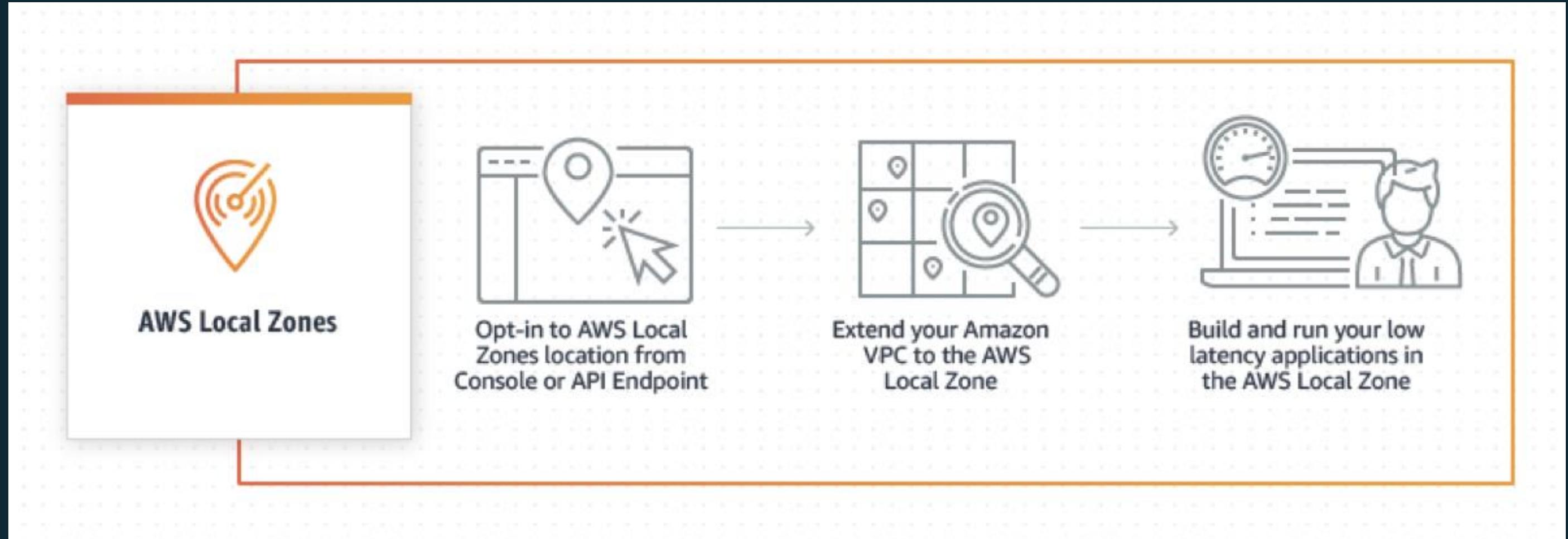
Deliver new low  
latency apps

# AWS Local Zones Use Cases

**Highly-demanding applications that require single-digit millisecond latencies to end-users, including:**

- **Media & Entertainment Content Creation**
- **Real-time Multiplayer Gaming**
- **Reservoir Simulations**
- **Electronic Design Automation**
- **Machine Learning Inference**

# How It Works



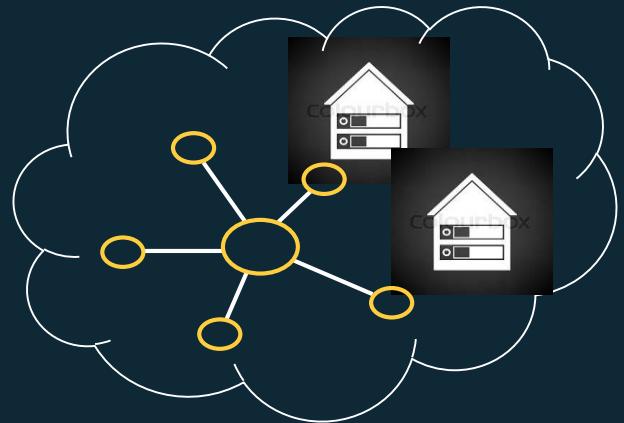
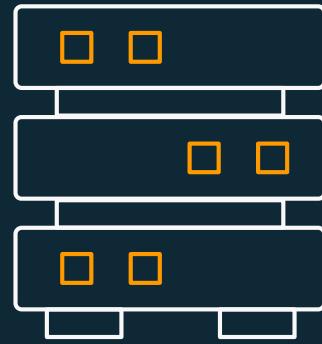
# Connectivity over mobile networks today



# Connectivity over mobile networks with Wavelength



# Wavelength Zone



Same AWS-designed **infrastructure** as in AWS data centers

Hosted in a site within a CSP partner network



Managed and monitored from an AWS region



Integrated into the CSP 5G Network

# AWS Wavelength Benefits



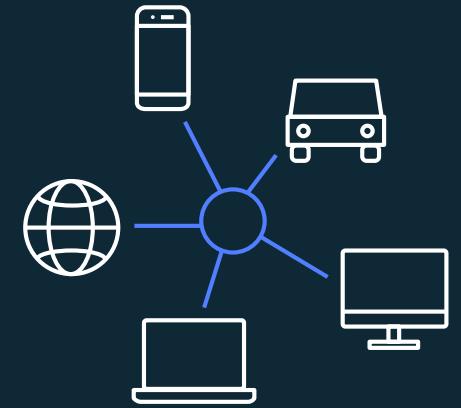
**Low latency/  
high bandwidth**



**Consistent  
development  
experience**



**Same AWS  
benefits**



**Ubiquity**

# AWS Wavelength

- Extends AWS infrastructure to 5G networks
- Run latency-sensitive portions of applications in “Wavelength Zones,” and seamlessly connect to the rest of your applications and the full breadth of services in AWS
- Same AWS APIs, tools, and functionality
- Global partner network



# Databases and Analytics

# Amazon Managed Apache Cassandra Service

A **scalable, highly available, and managed**  
*Apache Cassandra-compatible database service.*



# Benefits

## Apache Cassandra–compatible

Amazon Managed Cassandra Service implements the Apache Cassandra CQL API, allowing you to use the Cassandra Query Language (CQL) and drivers that you already use. Updating your application is as easy as changing the endpoint to the new Amazon Managed Cassandra Service table.

## No servers to manage

You don't need to provision, patch, or manage servers, so you can focus on building better applications. Capacity is on demand—you pay for only the resources you use, and you don't have to plan for peak workloads.

## Performance at scale

Consistent, single-digit-millisecond response times at any scale. Build applications with virtually unlimited throughput and storage that can serve thousands of requests per second without the need for capacity planning.

## Highly available and secure

Tables are encrypted by default and replicated three times in multiple AWS Availability Zones for high availability. Secure your data with access management, and use performance monitoring to keep your applications running smoothly.



Many applications, including serverless apps, have a large number of open DB connections and high connection open/close rate, exhausting DB resources

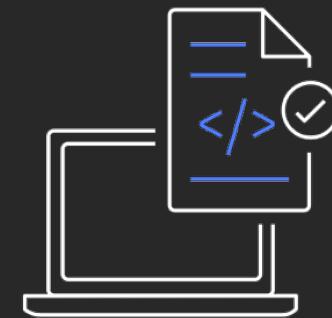


Modern apps can have 1000s of DB connections, exhausting DB resources

Self-managed proxy servers help manage DB load, but are difficult to deploy



Custom failure handling code can contain security risks like DB credentials



INTRODUCING

# Amazon RDS Proxy - Preview

Fully managed, highly available database proxy feature for Amazon RDS. Pools and shares DB connections to make applications more scalable, more resilient to database failures, and more secure.



Pool and share  
app scaling



availability  
DB failover times



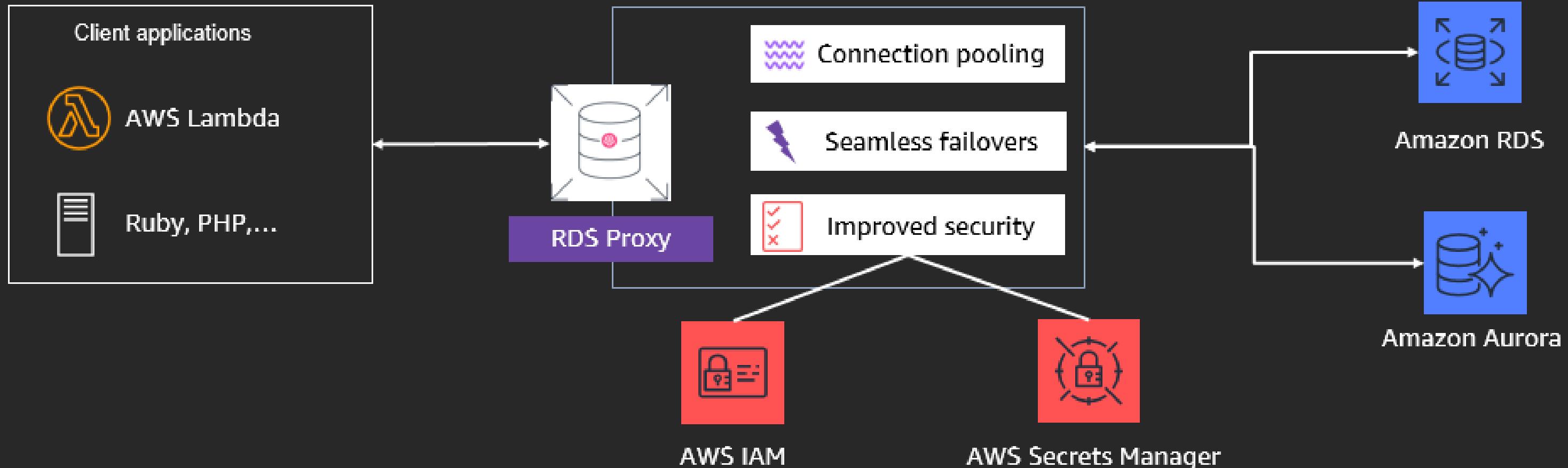
data  
security  
access controls



Fully managed  
compatible



# Amazon RDS Proxy – How it Works



# New 3<sup>rd</sup> generation Redshift compute instance: RA3

Managed storage



High-speed cache

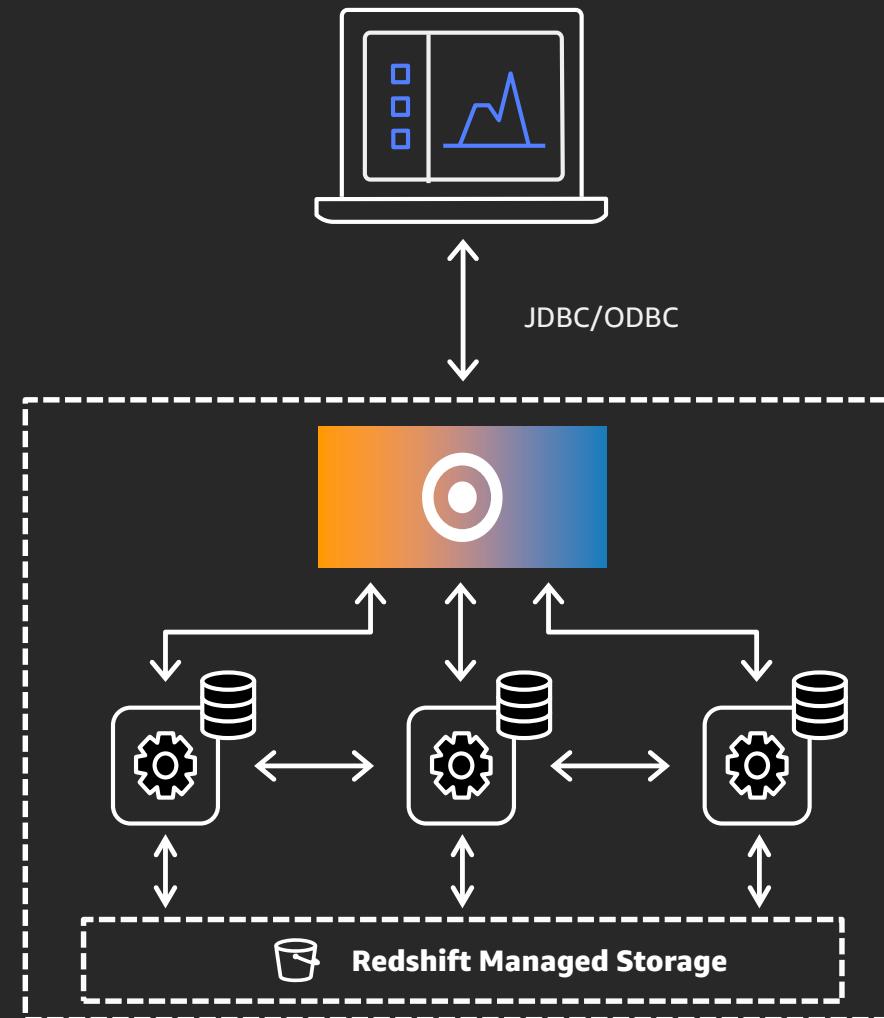


High-bandwidth networking

Scale data warehouse only based on steady state compute needs

Pay separately for storage and compute

Automatic, no changes to any workflows, no need to manage storage

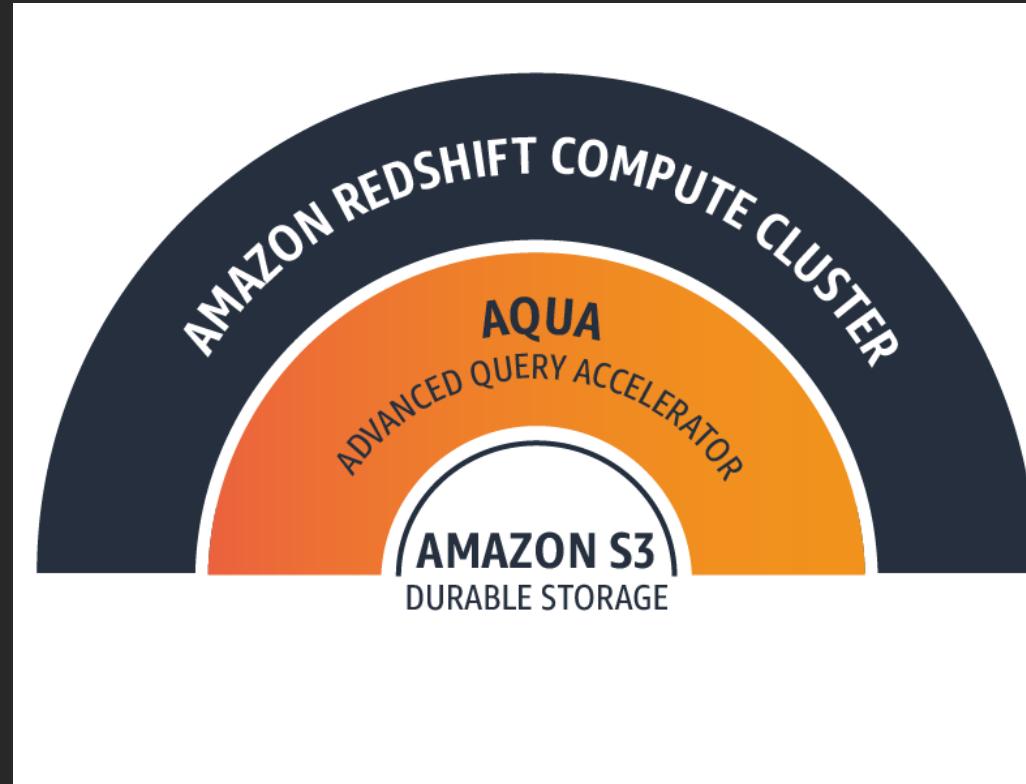


PREVIEW

COMING IN 2020

# AQUA – Advanced Query Accelerator

Redshift runs 10x faster than any other cloud data warehouse without increasing cost



AQUA brings compute to the storage layer so data doesn't have to move back and forth

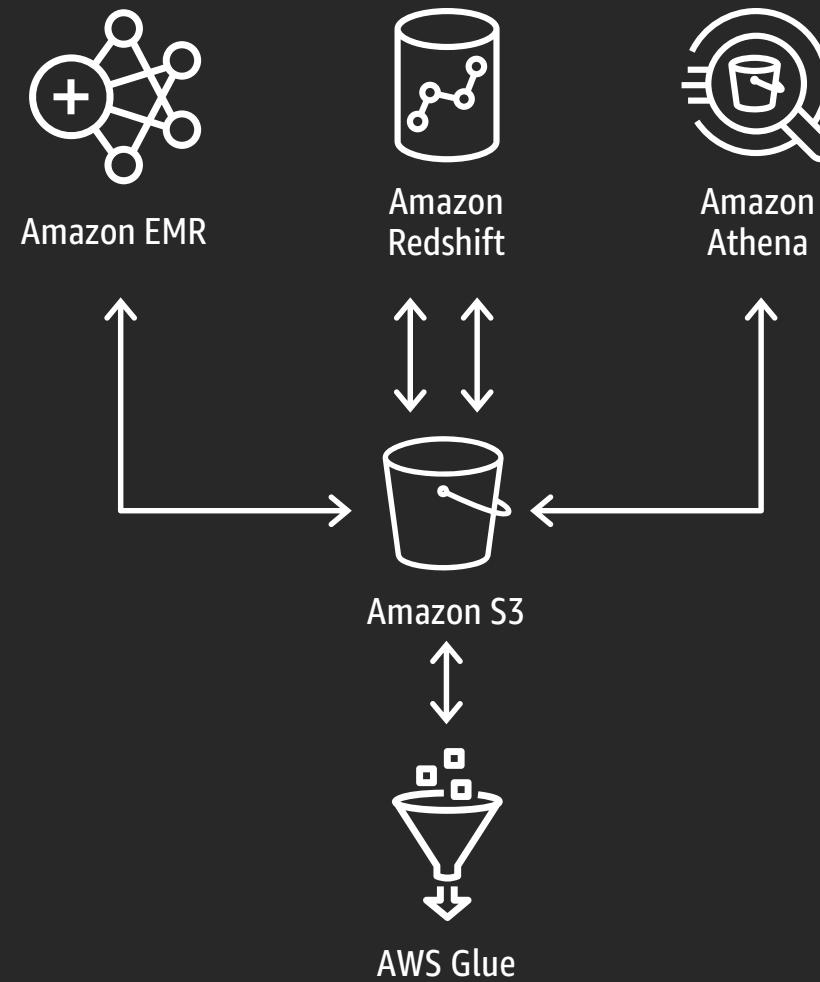
High-speed cache on top of S3 scales out to process data in parallel across many nodes

AWS-designed processors accelerate data compression, encryption, and data processing

100% compatible with the current version of Redshift

# Amazon Redshift Data Lake Export

Export data directly to Amazon S3 in Apache Parquet



Save results of data transformation into S3 data lake

Export with the UNLOAD command and specify Parquet

Redshift formats, partitions, and moves data into S3

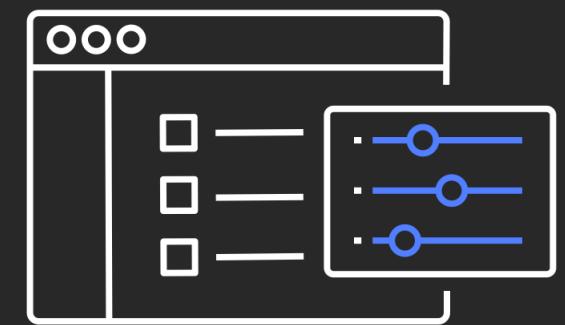
Analyze with Amazon SageMaker, Athena, and EMR



Application usage, click  
stream data, KPIs



Faults and errors, Security



Up time, resource use,  
performance

# Your logs have information

# What is Elasticsearch?

Elasticsearch,  
Logstash,  
and Kibana

Distributed  
search and  
analytics engine

Easy ingestion  
and visualization

Rank			DBMS	Score		
Nov 2019	Oct 2019	Nov 2018		Nov 2019	Oct 2019	Nov 2018
1.	1.	1.	Oracle 	1336.07	-19.81	+34.96
2.	2.	2.	MySQL 	1266.28	-16.78	+106.39
3.	3.	3.	Microsoft SQL Server 	1081.91	-12.81	+30.36
4.	4.	4.	PostgreSQL 	491.07	+7.16	+50.83
5.	5.	5.	MongoDB 	413.18	+1.09	+43.70
6.	6.	6.	IBM Db2 	172.60	+1.83	-7.27
7.	7.	8.	Elasticsearch 	148.40	-1.77	+4.94
8.	8.	7.	Redis 	145.24	+2.32	+1.06
9.	9.	9.	Microsoft Access	130.07	-1.10	-8.36
10.	10.	11.	Cassandra 	123.23	+0.01	+1.48

Source: DB-Engines.com, October 2019

# Amazon Elasticsearch Service



Amazon Elasticsearch Service is a fully managed service that makes it easy to deploy, manage, and scale Elasticsearch and Kibana

```
2.81.55 -- [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245
mp6.unicomp.net -- [01/Jul/1995:00:00:06 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
20.110.21 -- [01/Jul/1995:00:00:09 -0400] "GET /shuttle/missions/sts-73/mission-sts-73.html HTTP/1.0" 200 4085
r.letters.com -- [01/Jul/1995:00:00:11 -0400] "GET /shuttle/countdown/liftoff.html HTTP/1.0" 304 0
20.110.21 -- [01/Jul/1995:00:00:11 -0400] "GET /shuttle/missions/sts-73/sts-73-patch-small.gif HTTP/1.0" 200 4179
r.letters.com -- [01/Jul/1995:00:00:12 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 304 0
r.letters.com -- [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/video/livevideo.gif HTTP/1.0" 200 0
12.115.106 -- [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/countdown.html HTTP/1.0" 200 3985
aa.net -- [01/Jul/1995:00:00:13 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
4.144.152 -- [01/Jul/1995:00:00:13 -0400] "GET / HTTP/1.0" 200 7074
mp6.unicomp.net -- [01/Jul/1995:00:00:14 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
mp6.unicomp.net -- [01/Jul/1995:00:00:14 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
mp6.unicomp.net -- [01/Jul/1995:00:00:14 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
aa.net -- [01/Jul/1995:00:00:15 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
aa.net -- [01/Jul/1995:00:00:15 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
aa.net -- [01/Jul/1995:00:00:15 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
4.144.152 -- [01/Jul/1995:00:00:17 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 304 0
20.110.21 -- [01/Jul/1995:00:00:17 -0400] "GET /images/launch-logo.gif HTTP/1.0" 200 1713
y391.asahi-net.or.jp -- [01/Jul/1995:00:00:18 -0400] "GET /facts/about_ksc.html HTTP/1.0" 200 3977
-141.eden.com -- [01/Jul/1995:00:00:19 -0400] "GET /shuttle/missions/sts-71/images/KSC-95EC-0916.jpg HTTP/1.0" 200 341
y391.asahi-net.or.jp -- [01/Jul/1995:00:00:19 -0400] "GET /images/launchpalms-small.gif HTTP/1.0" 200 11473
89.154.54 -- [01/Jul/1995:00:00:24 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
s-gw.starway.net.au -- [01/Jul/1995:00:00:25 -0400] "GET /shuttle/missions/51-l/mission-51-l.html HTTP/1.0" 200 6723
ia-30.shadow.net -- [01/Jul/1995:00:00:27 -0400] "GET / HTTP/1.0" 200 7074
89.154.54 -- [01/Jul/1995:00:00:29 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
a.prodigy.com -- [01/Jul/1995:00:00:33 -0400] "GET /shuttle/missions/sts-71/sts-71-patch-small.gif HTTP/1.0" 200 1205
ia-30.shadow.net -- [01/Jul/1995:00:00:35 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 200 5866
2.lloyd.com -- [01/Jul/1995:00:00:37 -0400] "GET /shuttle/missions/sts-71/images/KSC-95EC-0613.jpg HTTP/1.0" 200 6171
-pc.moorecap.com -- [01/Jul/1995:00:00:38 -0400] "GET /history/apollo/apollo-13/images/70HC314.GIF HTTP/1.0" 200 1012
89.154.54 -- [01/Jul/1995:00:00:40 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
12-01.ix.netcom.com -- [01/Jul/1995:00:00:41 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
```

Storing data is expensive at scale

Limits the data retained for analysis

Miss out on valuable insights

INTRODUCING

# UltraWarm, a new storage tier for Amazon Elasticsearch Service



Store massive amounts of log data



Run interactive log analytics and visualization

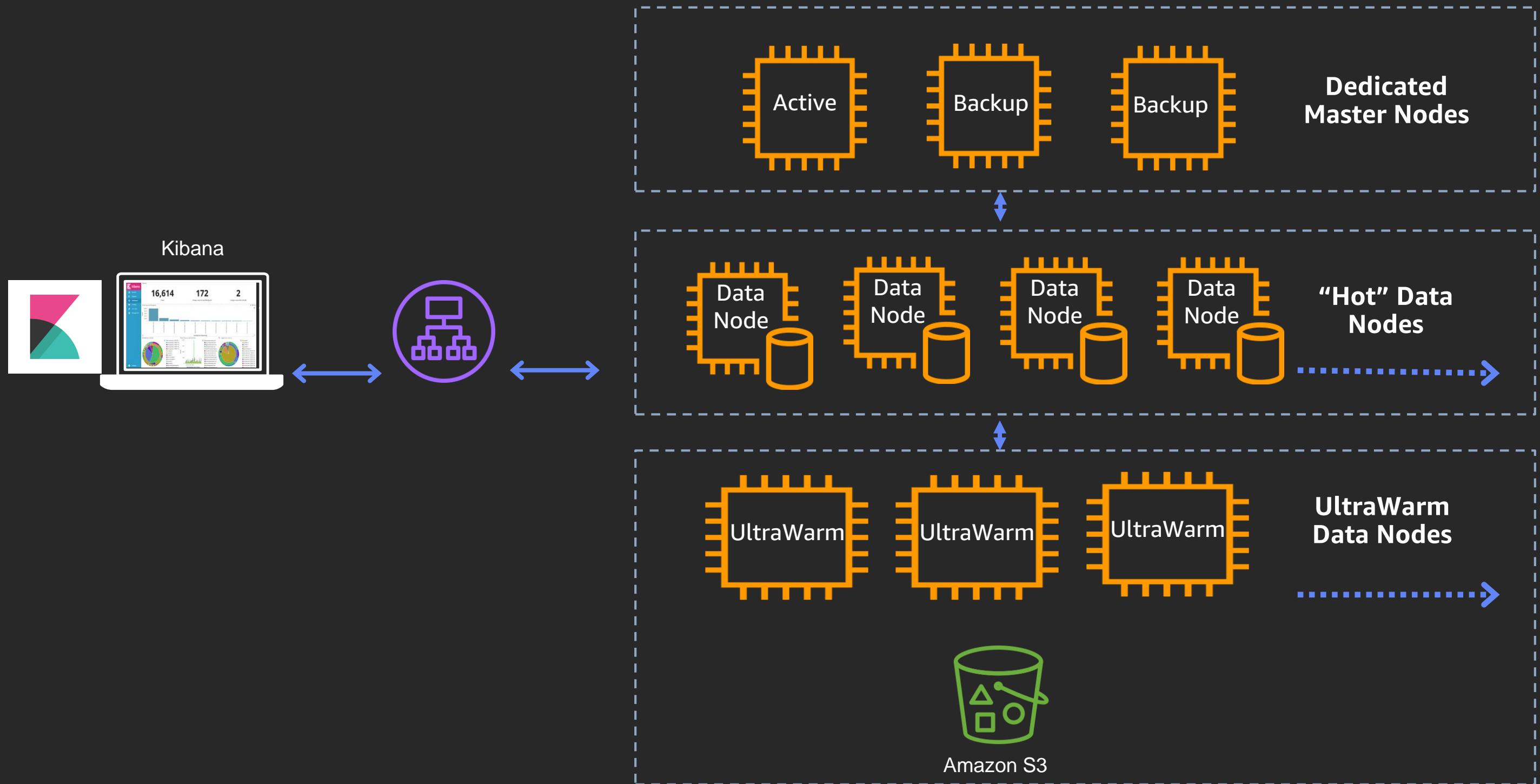


Higher performance and durability



Achieve up to 90% cost savings

# Architecture



# Reduces cost per GB to enable longer retention



Achieve 90%  
cost savings

Comparing to existing service today up to 90% savings  
on \$/GB vs Amazon Elasticsearch Service I3-based hot  
nodes

Enabling you to keep data accessible to  
Elasticsearch/Kibana for months and years

INTRODUCING

# UltraWarm, a new storage tier for Amazon Elasticsearch Service



Store up to 3 PB  
of log data



Run interactive log  
analytics and  
visualization



Higher performance  
and durability



Achieve up to  
90% cost  
savings

AVAILABLE IN PREVIEW TODAY

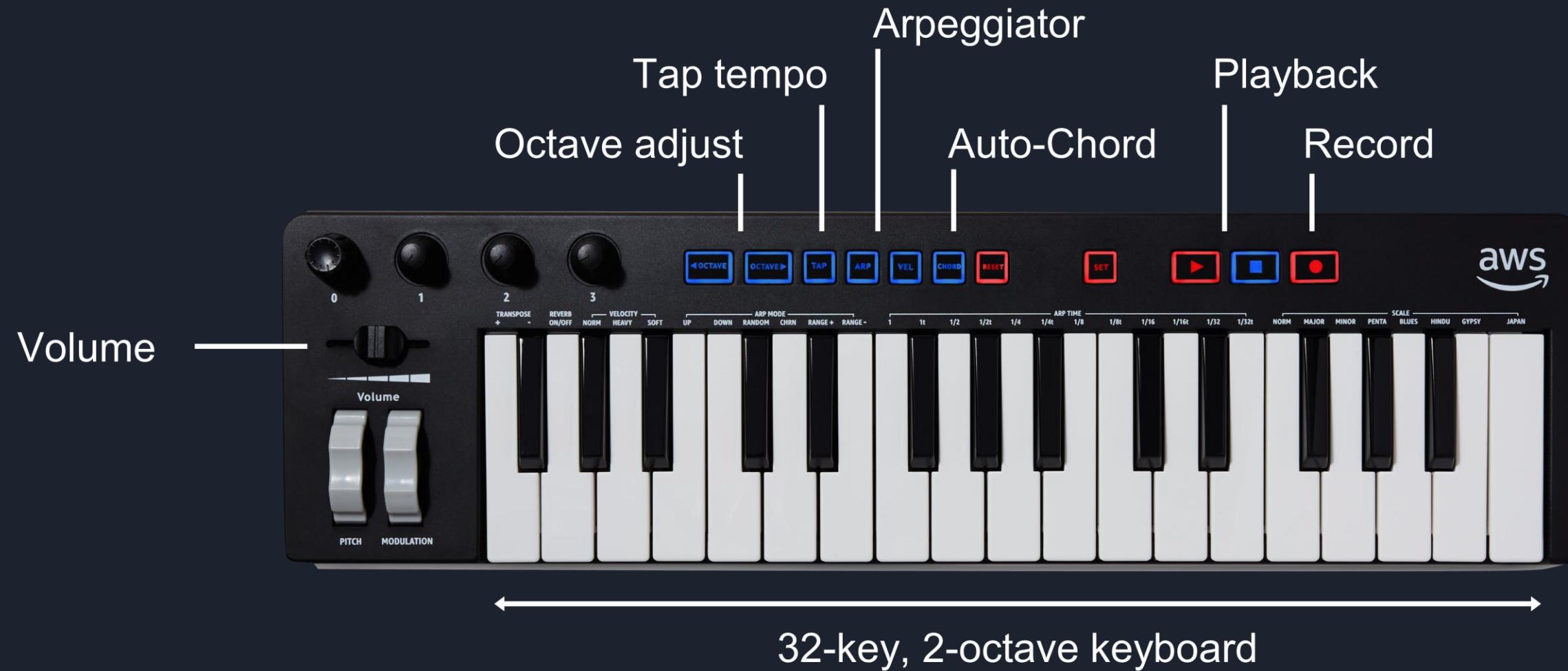
# Artificial Intelligence

AWS

# DeepComposer

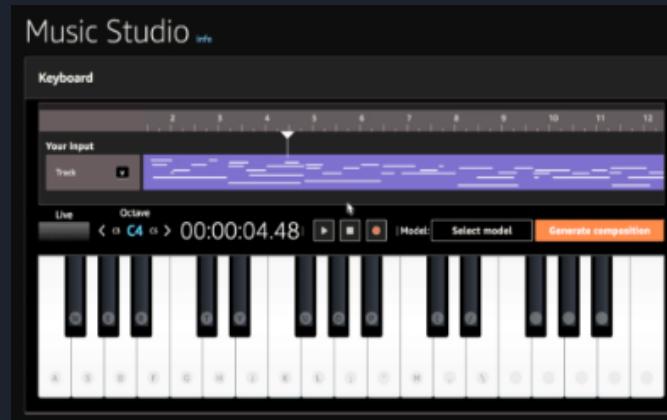
The world's first machine learning-enabled musical keyboard for developers





# Creative Meets Generative

1



Input a melody by connecting the AWS DeepComposer keyboard

2



Choose from jazz, rock, pop, classical, or build your own custom genre model in Amazon SageMaker

3



Publish your tracks to SoundCloud from the console. Export MIDI files to your favorite DAW

# Build your AI with AWS and Intel

AWS and Intel deliver the most comprehensive set of resources, tools, training, and services together



# AWS DeepRacer: An exciting way for developers to get hands-on experience with reinforcement learning (RL)



## Virtual simulator

Build models in Amazon SageMaker and train, test, and iterate quickly and easily on the track in the AWS DeepRacer 3D racing simulator.



## Car

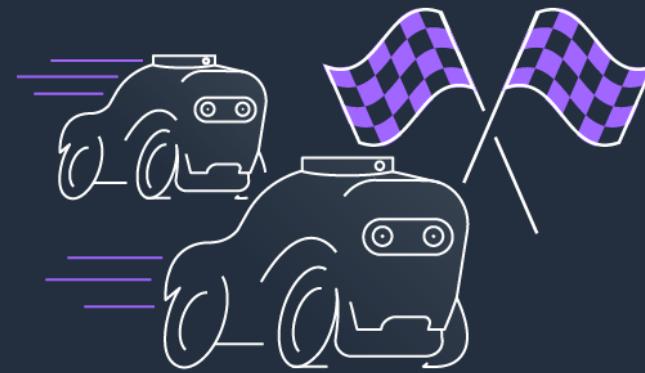
Experience the thrill of the race in the real-world when you deploy your reinforcement learning model onto AWS DeepRacer.



## Global Racing League

Compete in the world's first global, autonomous racing league, to race for prizes and glory and a chance to advance to the Championship Cup, or create your own races and compete against fellow developers in a league of your own.

# New Features



## AWS DeepRacer Evo

AWS DeepRacer Evo comes with stereo cameras and a Light Detection and Ranging (LIDAR) sensor, allowing the car to detect objects on the track and helping your model determine when to overtake another car to beat it to the finish line.

Available on [amazon.com](https://amazon.com) in 2020.  
[Sign up for the interest list](#)



## AWS DeepRacer simulator

Use advanced reinforcement learning algorithms, build models in Amazon SageMaker and train, test, and iterate models quickly and easily on the track in the AWS DeepRacer 3D racing simulator.



## Community Races

Create your own online DeepRacer race in the AWS console using the new Community Races feature. Compete against colleagues and friends in a league of your own. your own leagues.

# Fraud comes in all shapes and forms

## Payment Fraud

- Compromised Payment Instruments (e.g., stolen cards)
- Intentional Non-Payment (e.g., pre-paid cards)

## Account Takeover/Compromise

- Username/Password
- API Key

## Abuse

- Free Tier Misuse
- Premium Phone Number

# Business Rules vs Machine Learning



Prevention



Detection

Business Rules look for specific conditions or behaviors

- Business Rules are easily explained and validated
- Sample New Account Registration rule:

```
If IP_ADDRESS_LOCATION == ['Japan'] and CUST_ADDRESS_COUNTRY == ['JAPAN'] and CUSTOMER_PHONE_LOC == ['Spain'] THEN Investigate
```

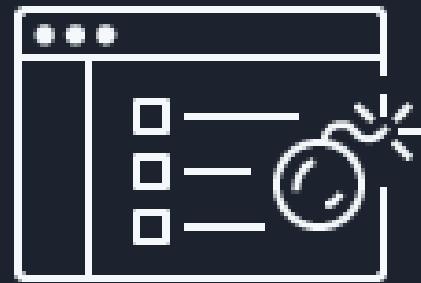
ML Models learn more general patterns by looking at lots of examples

- When fraudsters make small tweaks, the model still recognizes them as suspicious since it's unlike anything it has seen from legitimate customers
- ML models are not just good at finding the risky patterns, they're much less brittle than rules

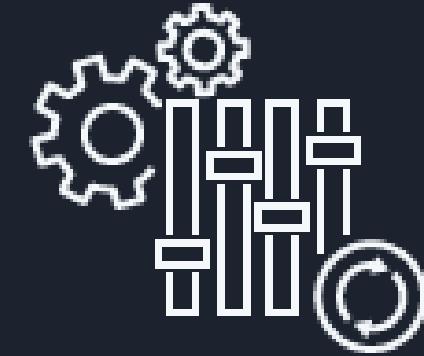
# Fraud detection is difficult



\$\$\$ billions lost to  
fraud each year



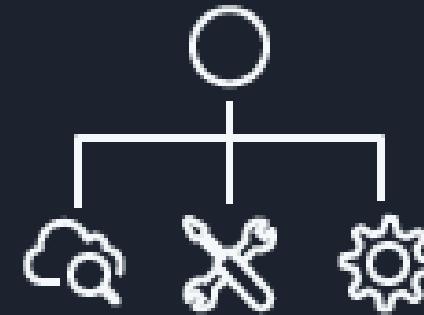
Online business prone  
to fraud attacks



Bad actors change  
tactics often



Rules = more human  
reviews



Dependent on others to  
update detection logic

# Fraud detection with ML is also difficult



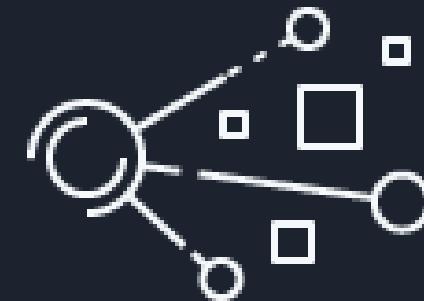
Top data scientists are costly & hard to find



One-size-fits-all models underperform



Often need to supplement data



Data transformation + feature engineering



Fraud imbalance = needle in a haystack

# Introducing Amazon Fraud Detector

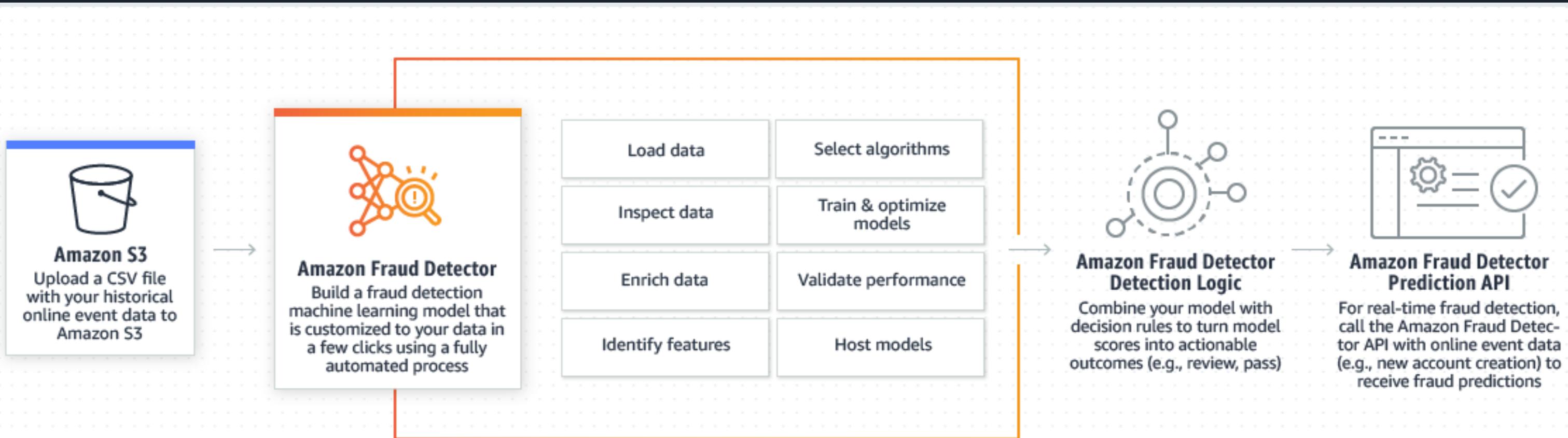
---

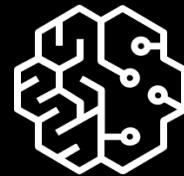
A fraud detection service that makes it easy for businesses to use machine learning to detect online fraud in real-time, at scale.

---



# How it works





# Amazon SageMaker

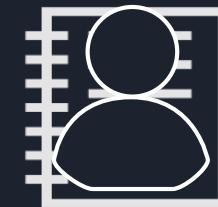
Machine Learning for every developer & data scientist

# Amazon SageMaker

## Addressing challenges to machine learning



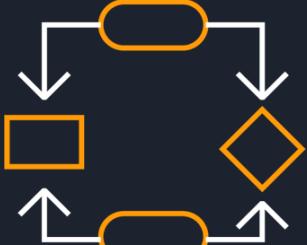
First fully integrated development environment (IDE) for machine learning  
**[Amazon SageMaker Studio](#)**



Enhanced notebook experience with quick-start & easy collaboration  
**[Amazon SageMaker Notebooks \(Preview\)](#)**



Experiment management system to organize, track & compare thousands of experiments  
**[Amazon SageMaker Experiments](#)**



Automatic debugging, analysis, and alerting  
**[Amazon SageMaker Debugger](#)**



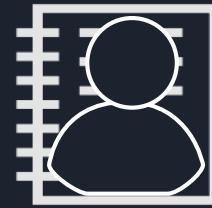
Model monitoring to detect deviation in quality & take corrective actions  
**[Amazon SageMaker Model Monitor](#)**



Automatic generation of ML models with full visibility & control  
**[Amazon SageMaker Autopilot](#)**

# Introducing Amazon SageMaker Studio

The first fully integrated development environment (IDE) for machine learning



Collaboration at scale

Share notebooks without tracking code dependencies



Easy experiment management

Organize, track, and compare thousands of experiments



Automatic model generation

Get accurate models with full visibility & control without writing code



Higher quality ML models

Automatically debug errors, monitor models, & maintain high quality



Increased productivity

Code, build, train, deploy, & monitor in a unified visual interface

xgboost\_customer\_churn.ipynb X

F + X □ ▶ ■ C Markdown git conda\_amazonei\_mxnet\_p27

- Have the predictor variable in the first column
- Not have a header row

But first, let's convert our categorical features into numeric features.

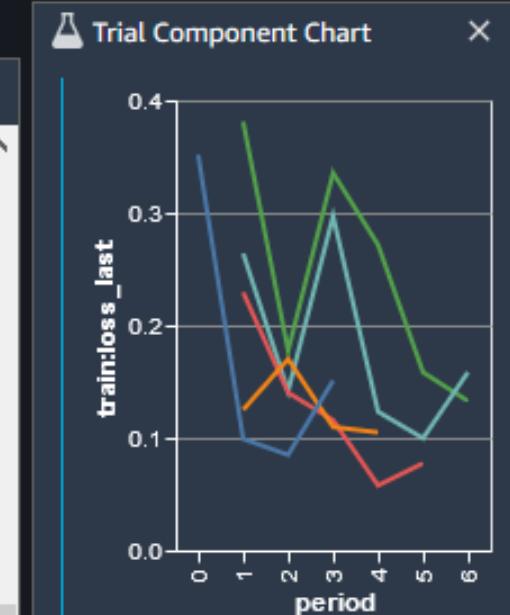
```
[ ]: model_data = pd.get_dummies(churn)
model_data = pd.concat([model_data['Churn?_True.'], model_data.drop(['Chur
< ...
***
```

And now let's split the data into training, validation, and test sets. This will help prevent us from overfitting the model, and allow us to test the models accuracy on data it hasn't already seen.

```
[ ]: train_data, validation_data, test_data = np.split(model_data.sample(frac=1
train_data.to_csv('train.csv', header=False, index=False)
validation_data.to_csv('validation.csv', header=False, index=False)
< ...
***
```

Now we'll upload these files to S3.

```
[ ]: boto3.Session().resource('s3').Bucket(bucket).Object(os.path.join(prefix,
boto3.Session().resource('s3').Bucket(bucket).Object(os.path.join(prefix,
< ...
***
```



Trial Component List X

C TRIAL COMPONENTS

10 rows selected

Status	Experiment	Type	Trial	...
Completed	customer-churn-pred...	Training job	Trial-3	Training job
Completed	customer-churn-pred...	Training job	Trial-2	Training job
Completed	customer-churn-pred...	Training job	Trial-1	Training job
Completed	customer-churn-pred...	Training job	Trial-0	Training job
...	...	...	...	...

# Introducing Amazon SageMaker Autopilot

## Automatic model creation with full visibility & control



### Quick to start

Provide your data in a tabular form & specify target prediction



### Automatic model creation

Get ML models with feature engineering & automatic model tuning automatically done



### Visibility & control

Get notebooks for your models with source code



### Recommendations & Optimization

Get a leaderboard & continue to improve your model

aws File Edit View Run Kernel Git Tabs Settings Help

+ 📂 ⬆️ ⬇️ 🔍

/ automl-preview /

Name	Last Modified
bank-additional	12 hours ago
model	12 hours ago
sagemaker_auto...	3 hours ago

EXPERIMENT: MY-SAGEMAKER-AUTOPILOT

Trials Job profile

Open candidate generation notebook Open data exploration notebook

TRIALS

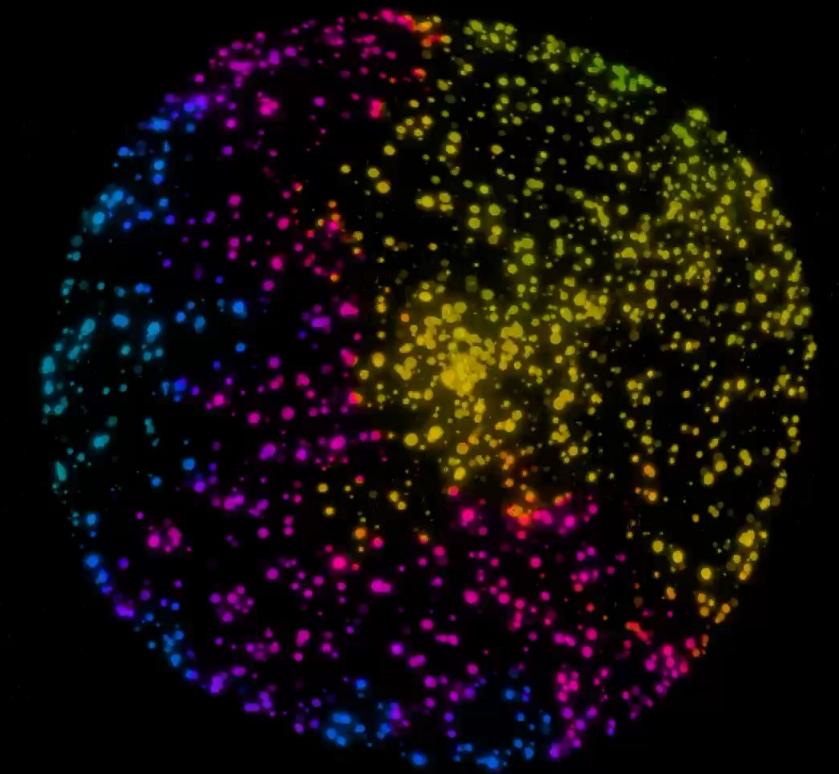
0 row selected Deploy model

Trial name	Status	Start time	End time	Objective
my-sagemaker-tuning-job-...	Completed	9 hours ago		0.9206119775772095
my-sagemaker-tuning-job-...	Completed	9 hours ago		0.9202479720115662
my-sagemaker-tuning-job-...	Completed	7 hours ago		0.9200050234794617
my-sagemaker-tuning-job-...	Completed	7 hours ago		0.9195190072059631
my-sagemaker-tuning-job-...	Completed	9 hours ago		0.9191550016403198
my-sagemaker-tuning-job-...	Completed	7 hours ago		0.9190340042114258
my-sagemaker-tuning-job-...	Completed	8 hours ago		0.9189119935035706
my-sagemaker-tuning-job-...	Completed	8 hours ago		0.9186699986457825
my-sagemaker-tuning-job-...	Completed	8 hours ago		0.9186699986457825

0 \$ 1 my-sagemaker-autopilot

# Developer Tools

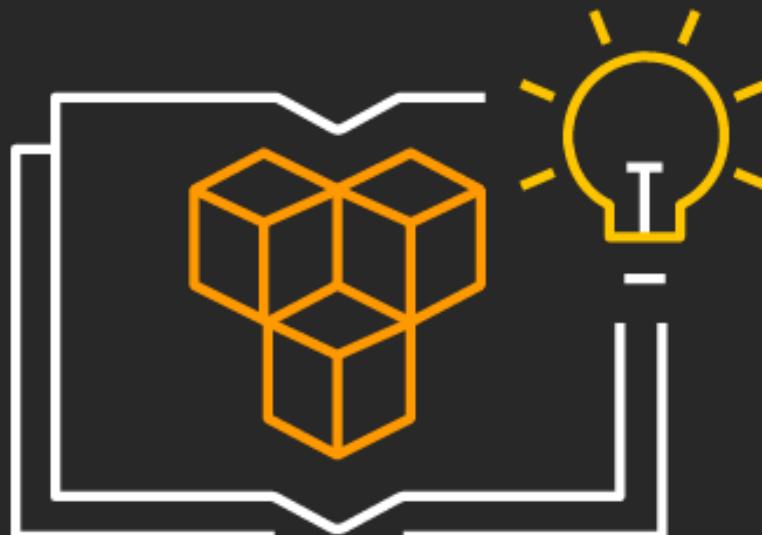
# How does Amazon...



**build resilient, high-quality distributed systems?**

# The Amazon Builders' Library

How Amazon builds and operates software



Architecture, software delivery, and operations

---

By Amazon's senior technical executives and engineers

---

Real-world practices with detailed explanations

---

Content available for free on the website

# The Amazon Builders' Library

## ARCHITECTURE

LEVEL 300



### Leader election in distributed systems

Author: Marc Brooker

Improving efficiency, reducing coordination, and simplifying architectures by using leader election.



## SOFTWARE DELIVERY AND OPERATIONS

LEVEL 300



### Going faster with continuous delivery

Author: Mark Mansour

Automating the software testing and deployment process for speed and reliability



## SOFTWARE DELIVERY AND OPERATIONS

LEVEL 400



### Implementing health checks

Author: David Yanacek

Automatically detecting and mitigating server failures without unintended consequences from fleet-wide false positives.



## ARCHITECTURE

LEVEL 400



### Workload isolation using shuffle-sharding

Author: Colm MacCarthaigh

Shuffle Sharding is one of our core techniques for drastically limiting the scope of impact of operational issues

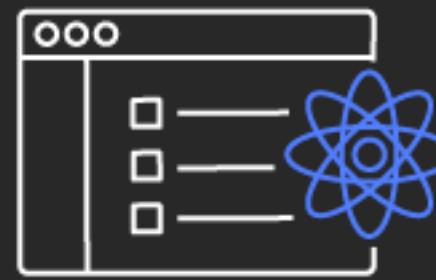


Many more at [aws.amazon.com/builders-library](http://aws.amazon.com/builders-library)

# Quantum Computing

# Amazon Braket

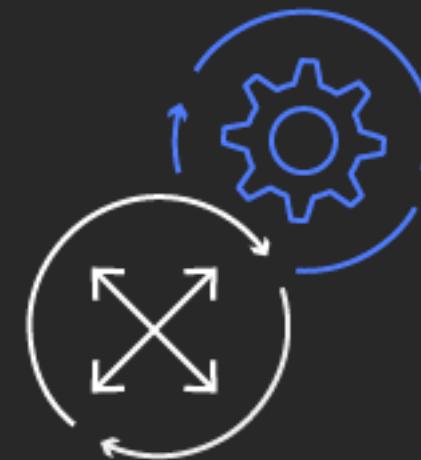
Fully managed service that makes it easy for scientists and developers to explore and experiment with quantum computing.



Single environment to design, test, and run quantum algorithms



Experiment with a variety of quantum hardware technologies



Run hybrid quantum and classical algorithms



Get Expert Help

# Summary

# What we skipped?

- AWS Nitro Enclaves
- AWS Lambda Provisioned Concurrency
- HTTP APIs (Preview) for Amazon API Gateway
- Amazon Kendra
- Amazon CodeGuru
- End of Support Migration Program (EMP) for Windows Server
- Transcribe Medical
- AWS Step Functions Express Workflows
- + Security, Storage & Networking Announcements

# Resources

re:Invent 2019 Blogposts:

<https://aws.amazon.com/blogs/aws/category/events/reinvent/>

AWS re:Invent 2019 – Keynote with Andy Jassy

<https://www.youtube.com/watch?v=7-31KgImGgU>

AWS re:Invent 2019 – Keynote with Dr. Werner Vogels

<https://www.youtube.com/watch?v=OdzaTbaQwTq>

AWS re:Invent 2019 – Monday Night Live with Peter deSantis

<https://www.youtube.com/watch?v=GPUWATKe15E>

# Questions?

# Thank you!

**Vladimir Simek, AWS**

vladsim@amazon.com