

Le webinaire commencera
dans quelques minutes

AWS re:Invent 2017

New services & features

Julien Simon, Principal AI/ML Evangelist, EMEA
@julsimon

Agenda

- Compute: Instances, containers, serverless
- Databases, Analytics, Storage
- Network & Security
- Developer Tools
- Artificial Intelligence
- Internet of Things (IoT)
- Augmented Reality (AR) and Virtual Reality (VR)

Not covered today

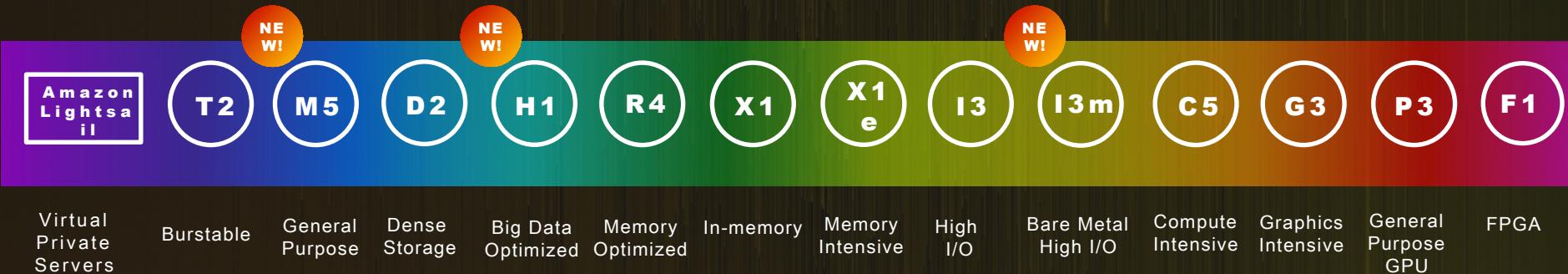
- *Media Services (Amazon Elemental Technologies)*
- *Alexa for Business*



Compute – Instances

NEW!

BROADEST SPECTRUM OF COMPUTE INSTANCES



EC2 Elastic GPUs
Graphics acceleration for EC2 instances



EC2 Spot Instances
• Hibernation
• No Bid Pricing



EC2 Dedicated Instances

P3.16XLARGE

BY THE NUMBERS

8

Tesla V100 GPUs

128 GiB

GPU memory

40,960

CUDA cores

5,120

Tensor cores

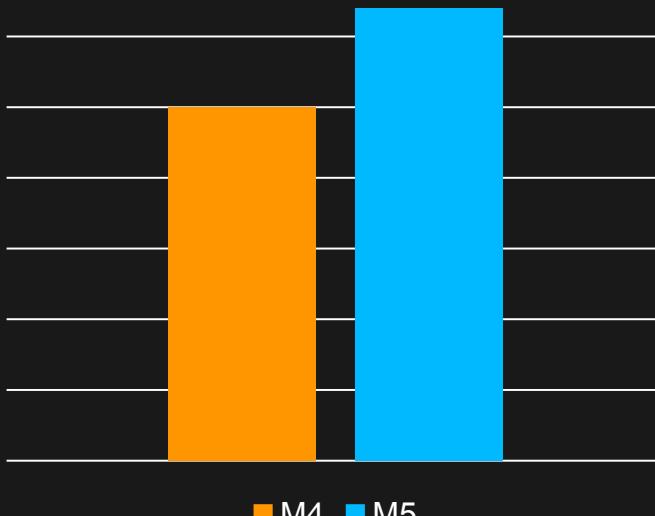
300

GB/s

Amazon EC2 M5 Instances (**GA**)

Next generation of EC2 general purpose instances

14% Price / Performance Improvement With M5



- Powered by 2.5 GHz Intel **Xeon Platinum** 8000-series "Skylake" Processor
- New larger instance size – m5.24xlarge with **96 vCPUs** and **384 GiB of memory**
- Improved network and EBS performance on smaller instance sizes
- Support for **Intel AVX-512**
- Powered by new lightweight Nitro Hypervisor

Amazon EC2 H1 Instances (GA)

New dense storage instance family for big data workloads

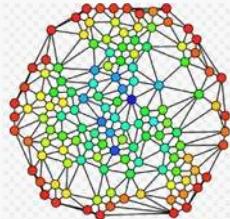
H1

New Storage-optimized instance

Up to 16TB of locally attached HDD storage

Up to 25 Gbps network bandwidth with ENA

Big Data Clusters



Kafka Streaming



MapReduce

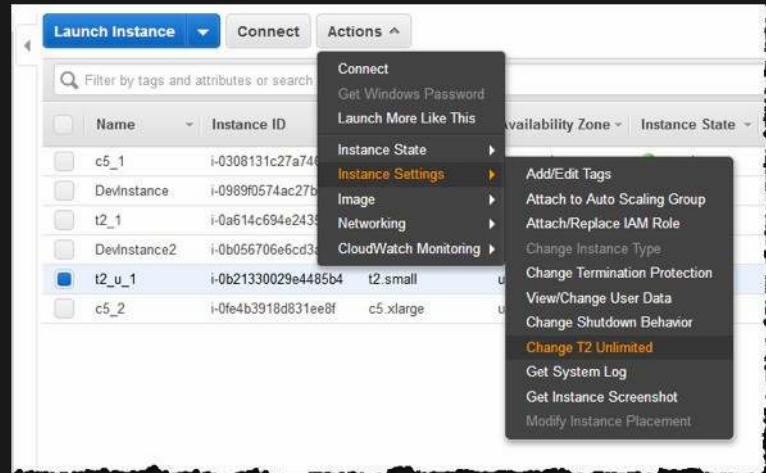


I3.metal (Bare Metal)

- Two Intel Xeon E5-2686v4 2.3GHz (72 vCPUs)
- 512GiB Memory
- 15.2 TB local SSD NVMe storage
- 25 Gbps ENA-based enhanced networking

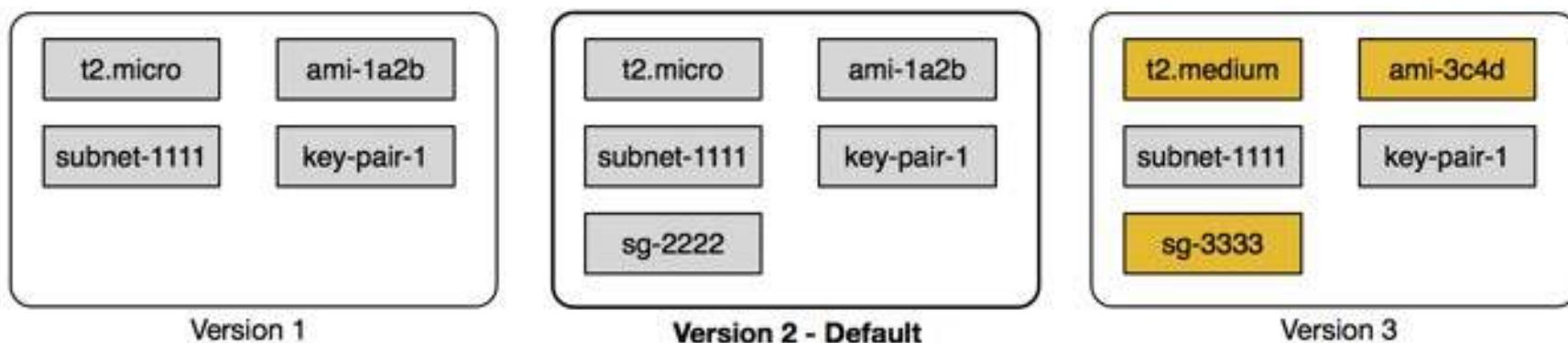
T2 Unlimited

- Continue bursting beyond your stored burst limit
- Track in CloudWatch
CPUSurplusCreditBalance
- \$0.05 per vCPU hour (charged per-millisecond)
- Simple console or API call
- Allows you to deal with spiky loads without changing to a more expensive instance type



Launch Templates

- Versioned launch parameterization (saving time in API and console)
- Template any parameter you would use in the *RunInstances* API



Launch template

Spot Enhancements

- Save up to 90% over on-demand costs
- Synchronous launch of spot instances via RunInstances API (immediately get instance IDs)
- Previously had to deal with: Spot markets, bidding, asynchronous API
- Price changes in spot market happen gradually based on longer-term trends in supply / demand.
- Instance Hibernation:
 - EC2 hibernation agent write in-memory state of instance to root EBS volume
 - Amazon Linux, Ubuntu, Windows Server
 - Not all instance types supported
 - Preserves private IPs across stop/start cycles

Amazon Time Sync Service

- Access NTP over 169.254.169.123
- Fleet of redundant satellite-connected and Atomic clocks

```
[ec2-user@ip-10-0-0-193 ~]$ chronyc tracking
Reference ID      : A9FEA97B (169.254.169.123)
Stratum          : 4
Ref time (UTC)   : Tue Nov 21 05:57:49 2017
System time      : 0.000001344 seconds slow of NTP time
Last offset      : -0.000004561 seconds
RMS offset       : 0.000605875 seconds
Frequency        : 10.421 ppm fast
Residual freq    : -0.001 ppm
Skew              : 0.067 ppm
Root delay       : 0.000519172 seconds
Root dispersion  : 0.000163384 seconds
Update interval  : 64.6 seconds
Leap status       : Normal
[ec2-user@ip-10-0-0-193 ~]$ ]
```

Compute – Containers



Amazon Elastic Container Service (ECS)



The easiest way to
deploy and manage
containers

1

Integration with entire AWS
platform

ALB, Auto Scaling, Batch, Elastic Beanstalk, CloudFormation, CloudTrail, CloudWatch Events, CloudWatch Logs, CloudWatch Metrics, ECR, EC2 Spot, IAM, NLB, Parameter Store, and VPC

2

Scales to support clusters and
applications of any size

3

Service integrations are at
the container level

CUSTOMERS RUNNING CONTAINERS WITH ECS



Over 100,000 active ECS clusters

100s of millions of containers started each week

400% YoY customer growth since 2016

WHAT ABOUT KUBERNETES?



“According to a recent survey, 63% of users were already hosting Kubernetes on Amazon EC2. This was up from 44% a year ago.”

<https://www.cncf.io/blog/2017/06/28/survey-shows-kubernetes-leading-orchestration-platform/>

NEW!

Amazon Elastic Container Service for Kubernetes (EKS)



Available in preview
today



Hybrid cloud compatible



Highly available



Automated
upgrades and
patches

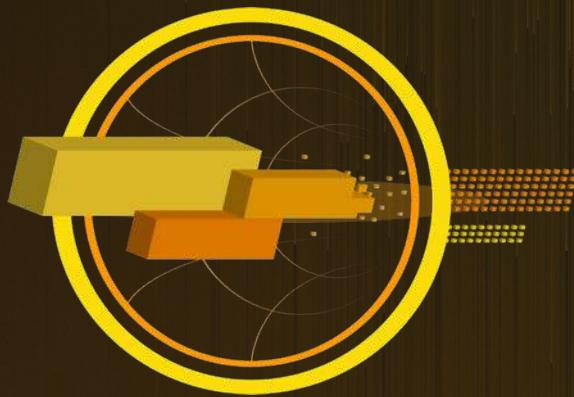


Integrated with
AWS Services

CloudTrail, CloudWatch,
ELB, IAM, VPC,
PrivateLink

NEW!

AWS Fargate - Serverless Container Service



Available for ECS

today

Available for EKS in

2018

Run containers without managing servers or clusters



No clusters to manage



Manages underlying infrastructure



Easy to run, easy to scale

Compute – Serverless

AWS LAMBDA IS EVERYWHERE

Event-driven services

 AWS Lambda

 Amazon API Gateway

 AWS Step Functions

 AWS X-Ray

Event sources

 Amazon S3

 Amazon DynamoDB

 Amazon Kinesis Streams

 Amazon Kinesis Firehose

 Amazon SNS

 Amazon SES

 Amazon Cognito

 Amazon CloudFormation

 Amazon CloudWatch Logs

 Amazon CloudWatch Events

 AWS CodeCommit

 AWS Config

 Amazon Lex

 Amazon CloudFront

Lambda inside

 AWS IoT

 AWS IoT Button

 AWS Greengrass

 AWS Snowball

 Edge
AWS

Lambda@Edge



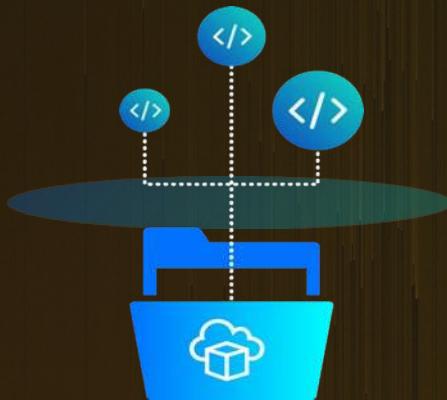
CUSTOMERS LOVE LAMBDA



NEW!

AWS Serverless Application Repository

Discover, deploy, and publish serverless apps



Available in preview
today

4 new powerful
features for AWS
Lambda



Discover a
collection of
serverless
applications



Easily deploy
to your AWS
account in a
few clicks



Publish your
own
applications to
share with the
community



API Gateway VPC
integration



Concurrency controls



3GB memory support



.NET Core 2.0 language
support (pre-announce)

...and Go

New AWS Lambda Runtimes - Preview

.NET 2.0

Go

joining Python, Node.js, Java and C#



CodeDeploy Lambda Support

Lambda deployment configurations						
		Search by name	Configurations per page	10	<< Viewing 1 to 9 of 9 configurations >>	
	Name	Compute platform	Configuration type	Step	Interval	
○ ▶	CodeDeployDefault.LambdaCanary10Percent30Minutes	AWS Lambda	Canary	10%	30 minutes	
○ ▶	CodeDeployDefault.LambdaCanary10Percent5Minutes	AWS Lambda	Canary	10%	5 minutes	
○ ▶	CodeDeployDefault.LambdaCanary10Percent10Minutes	AWS Lambda	Canary	10%	10 minutes	
○ ▶	CodeDeployDefault.LambdaCanary10Percent15Minutes	AWS Lambda	Canary	10%	15 minutes	
○ ▶	CodeDeployDefault.LambdaLinear10PercentEvery1Minute	AWS Lambda	Linear	10%	1 minutes	
○ ▶	CodeDeployDefault.LambdaLinear10PercentEvery2Minutes	AWS Lambda	Linear	10%	2 minutes	
○ ▶	CodeDeployDefault.LambdaLinear10PercentEvery3Minutes	AWS Lambda	Linear	10%	3 minutes	
○ ▶	CodeDeployDefault.LambdaLinear10PercentEvery10Minutes	AWS Lambda	Linear	10%	10 minutes	
○ ▶	CodeDeployDefault.LambdaAllAtOnce	AWS Lambda		N/A	N/A	

API Gateway Canary Deployments

- When deploying to a stage you can now specify “canarySettings”
- Shift a percentage of API traffic between 0.0 and 100.0
- Stage variables for the canary that can override production
- Continuously shift traffic (start with 10% move to 100%)
- Promote a canary release to a production stage

Databases & Analytics

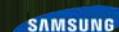
AURORA IS THE FASTEST GROWING SERVICE IN THE HISTORY OF AWS

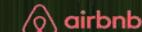
 ticketmaster



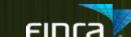
 verizon[®]

 Bristol-Myers Squibb



 airbnb

 Expedia

 FINRA

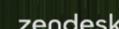
 hulu

 FanDuel

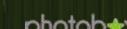


 FUNNY or DIE

 Experian

 zendesk

 THOMSON REUTERS

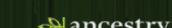
 photobox

 Cloudinary

 Alfresco

 FICO

 INTERCOM

 ancestry

 NASDAQ

 The Mainichi

 SAFE SOFTWARE

 Pearson

 []

 DOW JONES

 UBISOFT

 NETFLIX

 G3 IoT

 GREE

 CBS Interactive

 DHI

 docomo

 EA
 3M
PEOPLEADMIN

 Curves | Jenny Craig

 CyberAgent

 FANTASYDRAFT

 Zillow

 TalentBin
BY MONSTER

 nielsen

 Blackboard
 iret
 CAMEL GAMES
 Lookout
 SCHOLASTIC
 gogo

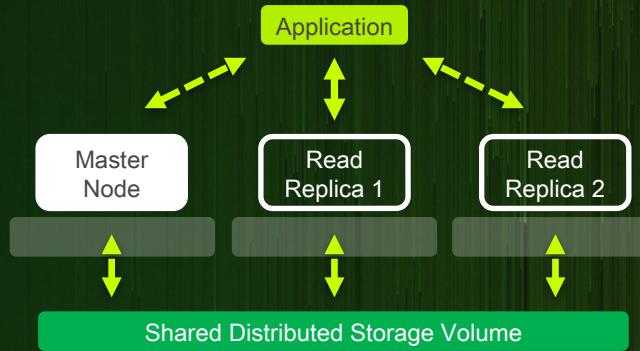
Grown by 2.5X again!

Tens of thousands of customers

NEW!

Aurora Multi-Master (Preview)

First relational database service
with scale-out on both read and write, across multiple datacenters



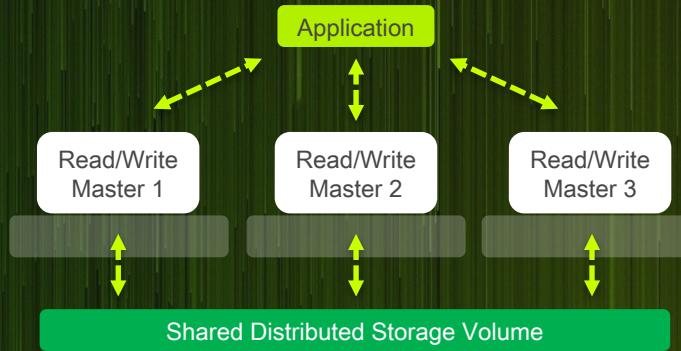
Availability
Zone 1

Availability
Zone 2

Availability
Zone 3

Up to 15 read replicas across 3 availability
zones
Auto-scale new read replicas

Seamless recovery from read replica
failures



Availabilit
y
Zone 1

Availabilit
y
Zone 2

Availabilit
y
Zone 3

Zero application downtime from ANY node
failure
Zero application downtime from ANY AZ
failure
Faster write performance

Multi-region coming in 2018

UNPREDICTABLE WORKLOADS ARE CHALLENGING



NEW!

Aurora Serverless



Sign up for the preview today

On-demand, auto-scaling database for applications with unpredictable or cyclical workloads



No need to provision instances



Automatically scales capacity up and down



Starts up on demand and shuts down when not in use



Pay per second and only for the database capacity you use

AMAZON DYNAMODB: FAST, FLEXIBLE, NOSQL DATABASE SERVICE



NEW!

DYNAMODB GLOBAL TABLES

First fully managed, multi-master, multi-region database



Build high performance, globally distributed applications

Low latency reads and writes to locally available tables

Disaster proof with multi-region redundancy

Easy to setup and no application re-writes required

DYNAMODB BACKUP & RESTORE

Only cloud database to provide on-demand and continuous backups



On-Demand Backups for long term data archival and compliance



Point In Time Restore for short term retention and protection against application errors



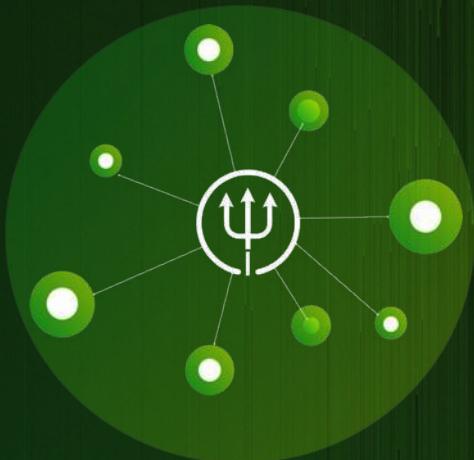
Backup hundreds of TB instantaneously with NO performance impact

On-Demand Backup generally available today

Point In Time Restore coming early

NEW!

Amazon Neptune



Available in preview
today



OPEN

Supports Apache TinkerPop and W3C RDF graph models

Fully managed graph database



FAST AND SCALABLE
Store billions of relationships and query with milliseconds latency



RELIABLE
6 replicas of your data across 3 AZs with full backup and restore

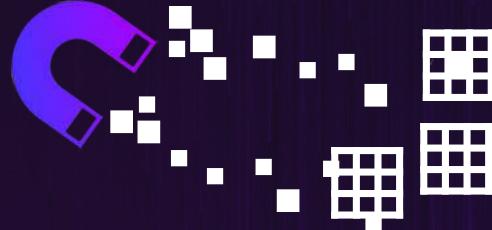


EASY
Build powerful queries easily with Gremlin and SPARQL

NEW!

S3 SELECT

Available in preview today



COMPLEX PRESTO QUERY
Against a standard TPC-DS dataset
6 sub-queries with each containing:

1
table

3
aggregations

4
filters

Powerful new S3 capability to pull out only the object data you need using standard SQL expressions

New API to select and retrieve data within objects

Accelerate any application that processes a subset of object data in S3

Improve data access performance by up to 400%

8 seconds

Without S3 Select

1.8 seconds

With S3 Select

4.5x faster

Glacier SELECT

Run queries directly on data stored in Glacier



Generally available today

Run queries on data stored at rest in Amazon Glacier

Any application can query Glacier data

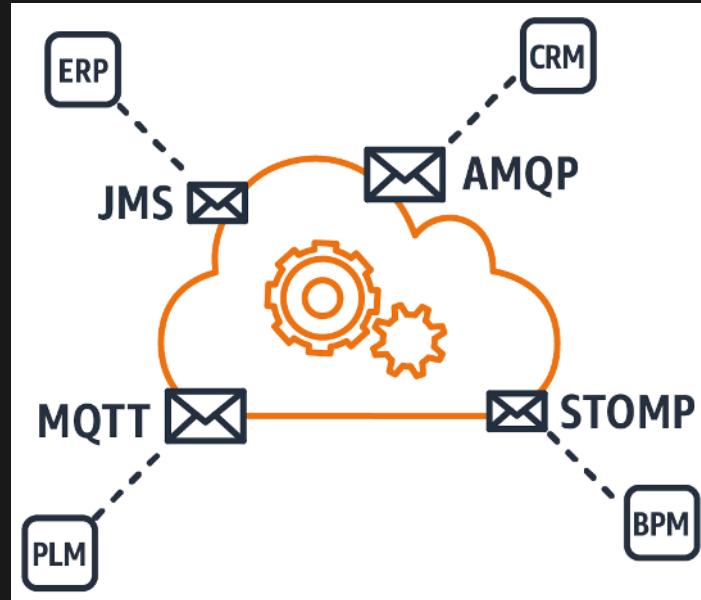
Retrieve only what you need

Makes Glacier part of your data lake

Amazon MQ

(available in GA at re:Invent 2017)

Managed message broker service for Apache ActiveMQ



Offload broker set up and administration

High availability and durability

Easily migrate from standards-based brokers

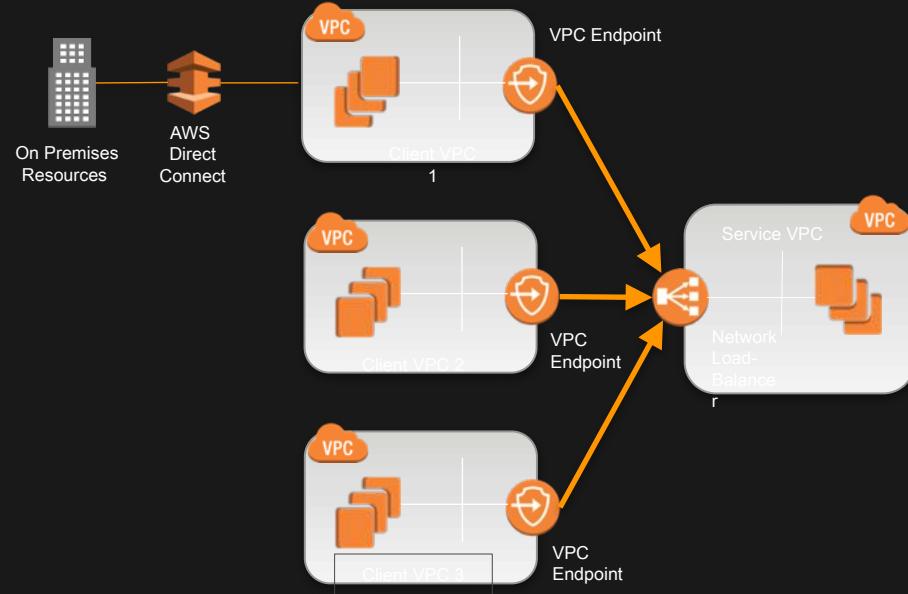
Launch a message broker in minutes

Pay as you go – get started for free

Networking & Security

AWS PrivateLink (GA)

Share AWS, customer and partner services privately between VPCs and on-premises networks



Secure. Scalable. Reliable.

Available on AWS Marketplace

Customers
and Partners



Vanguard®



APPDYNAMICS



Expedia®



cisco
Stealthwatch
Cloud



aqua



Inter-Region VPC Peering

- Traffic is encrypted
- No need for IGW or VPN devices
- Over AWS private backbone
- Does not transit on public internet

Peering Connections > Create Peering Connection

Create Peering Connection

Peering connection name tag ⓘ

Select a local VPC to peer with

VPC (Requester) ⓘ

CIDRs	CIDR	Status	Status Reason
172.31.0.0/16	172.31.0.0/16	associated	
2600:1f14:ddf:8b00::/56	2600:1f14:ddf:8b00::/56	associated	

Select another VPC to peer with

Account My account Another account

Region This region (us-west-2) Another region ⓘ

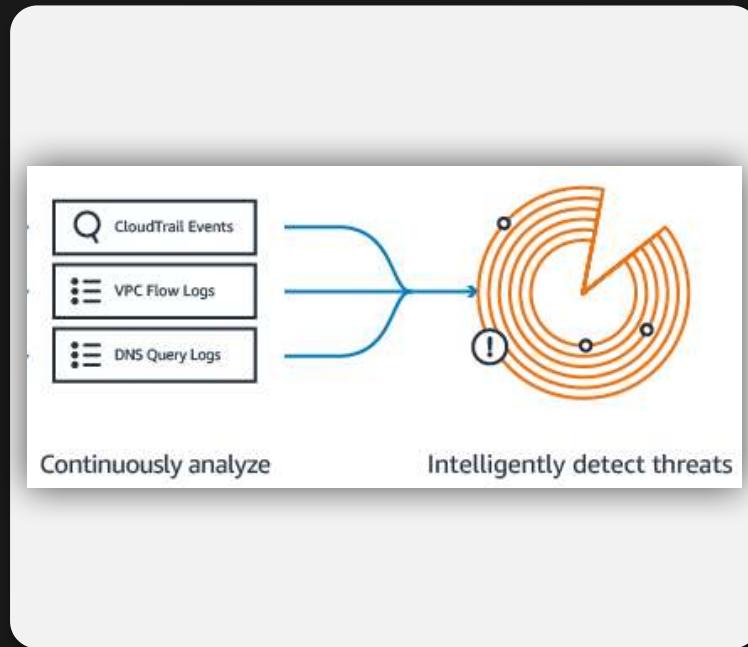
VPC (Acceptor)

* Required Cancel

Amazon GuardDuty

(available in GA at re:Invent 2017)

Intelligent threat detection for AWS accounts and workloads



Continuous analysis of billions of events

Detections maintained by AWS Security

Uses machine learning and threat intelligence

Strengthens security through automation

One-click deployment with no additional software or infrastructure to manage

Developer Tools



NEW!

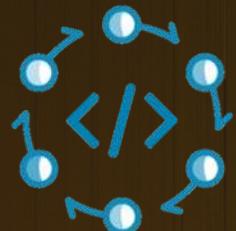
AWS Cloud9



AWS Cloud9

**Generally available
today**

 **Integration with
Serverless App**



A cloud IDE for writing, running,
and debugging code



Zero to
deploy
FAST



Deep
integration



Real-time pair
programming



Broad
debugging
support

**PUBLISH DIRECTLY INTO CODESTAR
TOOLS**



AWS
CodePipeline



AWS
CodeDeploy



AWS
CodeCommit



AWS
CodeBuild

AWS AppSync Preview (preview)

Build data driven mobile and web apps with real time updates and an offline programming model

- Build real-time, collaborative mobile and web apps
- Rapid prototyping and development with GraphQL
- Automatically manages off-line users' data
- Helps keep data secure
- Automatically resolves data conflicts
- Quickly build apps with data from multiple sources

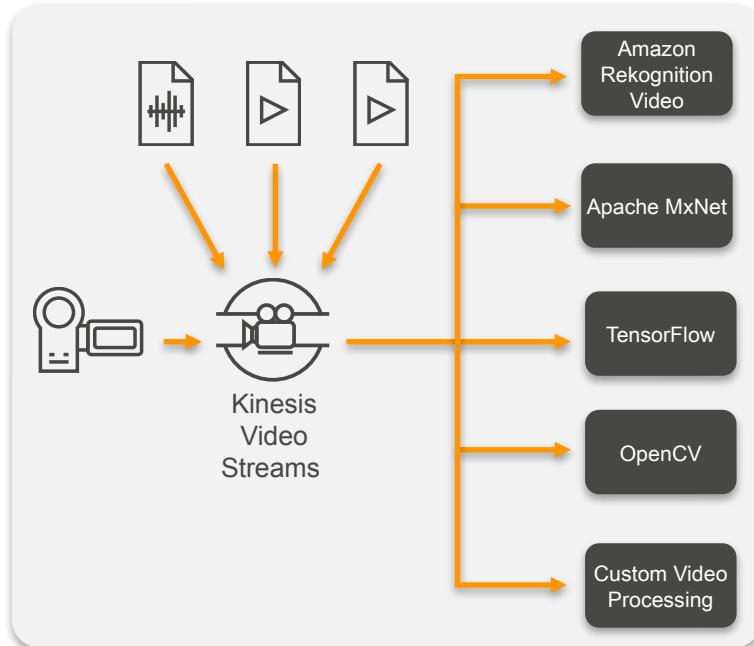
Artificial Intelligence

STREAMING DATA FROM CONNECTED DEVICES



Amazon Kinesis Video Streams (**GA**)

Stream video and time-encoded data for analytics



Stream video from millions of devices

Easily build vision-enabled apps

Secure

Durable, searchable storage

Serverless

Amazon Rekognition Video (GA)

Analyze activity, recognize and track
people in stored and live video stream

Four primary capabilities

1. Person tracking
2. Facial recognition
3. Object and activity detection
4. Video streaming support



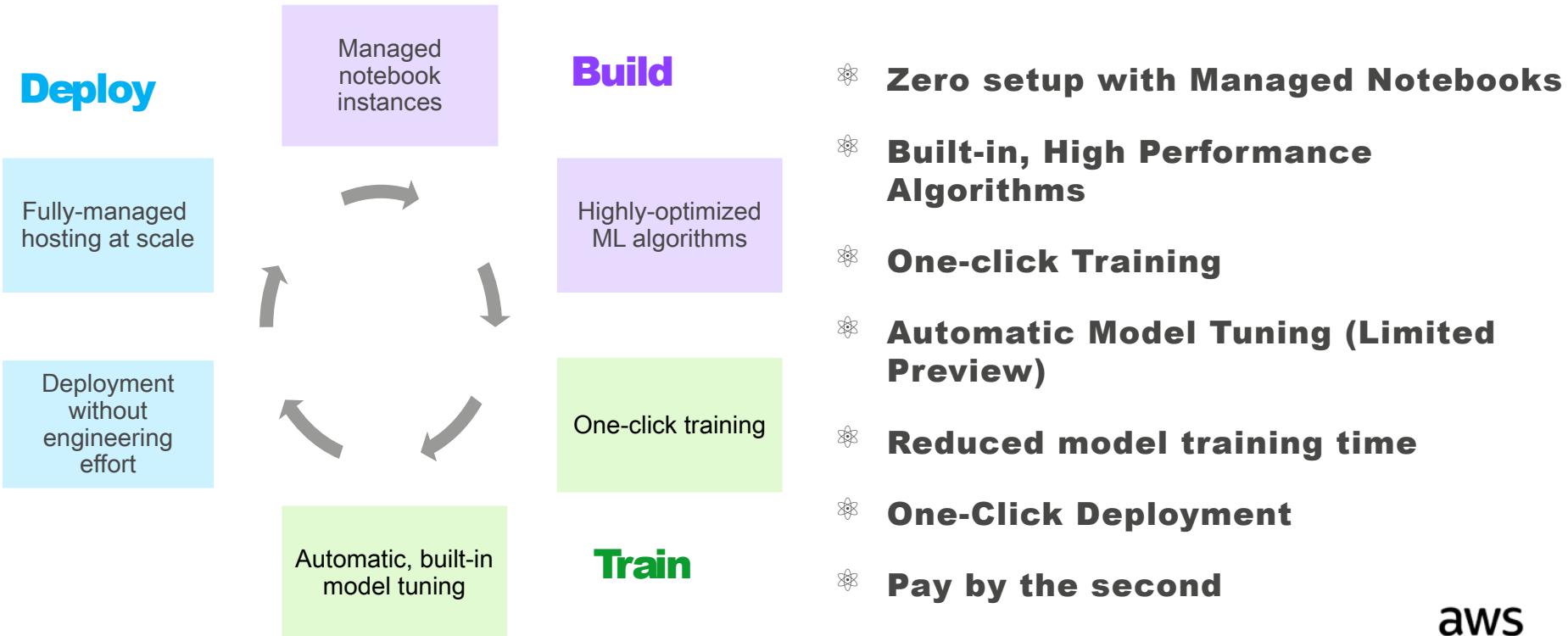
Real-time and batch
analysis



Motion and
time context

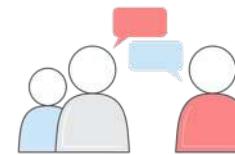
Amazon SageMaker (GA)

Build, Train, and Deploy Machine Learning Models at Scale



Amazon Transcribe (Preview)

Amazon Transcribe is an automatic speech recognition (ASR) service that makes it easy for developers to add speech to text capability to their applications



Support for both
regular &
telephony audio

S3 Integration

Time Stamps

Punctuation
&
formatting

Recognize
Multiple
Speakers

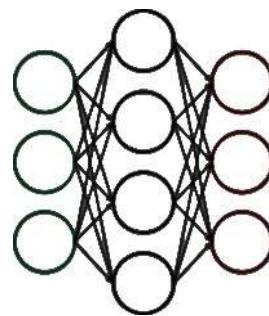
Custom
Vocabulary

Amazon Translate (Preview)

Amazon Translate is a neural machine translation service that delivers fast, high-quality, and affordable language translation.



REAL-TIME
TRANSLATION



POWERED BY DEEP
LEARNING



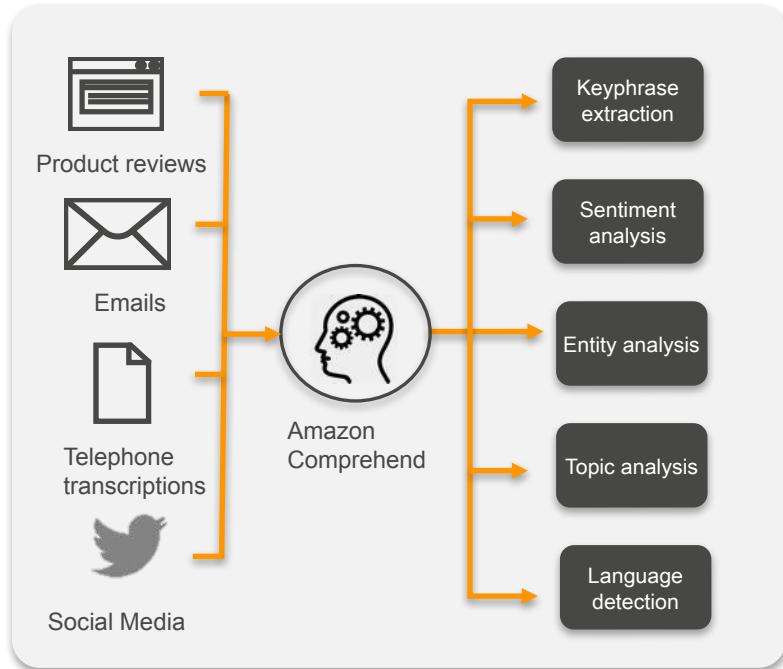
12 LANGUAGE PAIRS
(more to come)



LANGUAGE DETECTION

Amazon Comprehend (GA)

Natural Language Processing to discover insights from text



Classify language, extract key phrases, understand sentiment, identify / organize documents by topic

Continuously trained and constantly improving

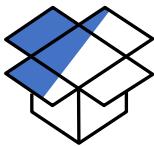
Integrated with Amazon S3 and AWS Glue

Internet of Things



AWS IoT 1-Click (Preview)

One click creation of an AWS Lambda function for any device



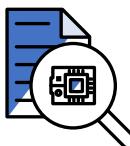
Ready out-of-the-box
HW without device
specific code



Securely connect
to the cloud



Create actions
using Lambda
functions with
one click



Fleet identification
and reporting

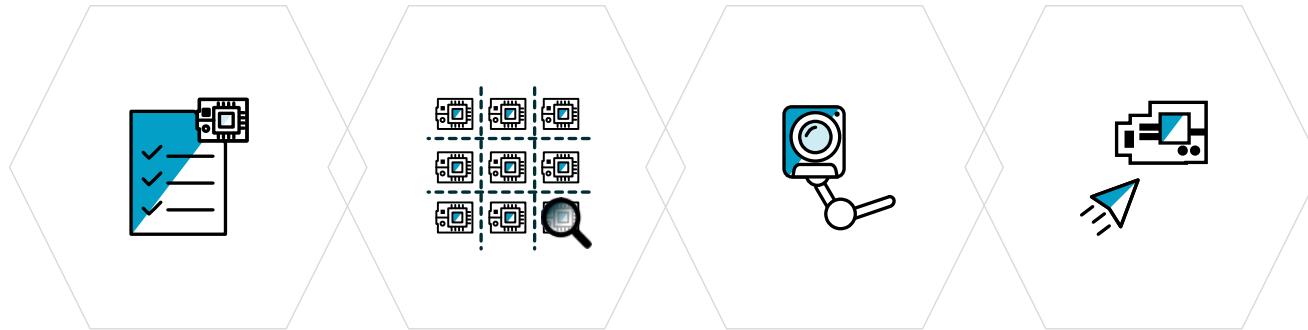


Mobile
application



AWS IoT Device Management (GA)

Fleet Management for Connected Devices



Batch fleet
provisioning

Real-time
fleet index & search

Fine grained
device logging
& monitoring

Over the
air updates

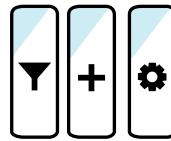


AWS IoT Analytics (Preview)

Easily analyze IoT data



IoT data collection from multiple sources



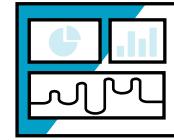
Data (pre-)processing and enrichment



Fully managed storage (raw & time-series)



Ad-hoc queries & QuickSight integration



In-depth analysis with Jupyter Notebooks



AWS Greengrass ML Inference (Preview)

Run machine learning models on the
edge



Access
models
trained in the
cloud



Built in
package for
Apache
MXNet



Local ML
inferencing



Access GPUs
and FPGAs

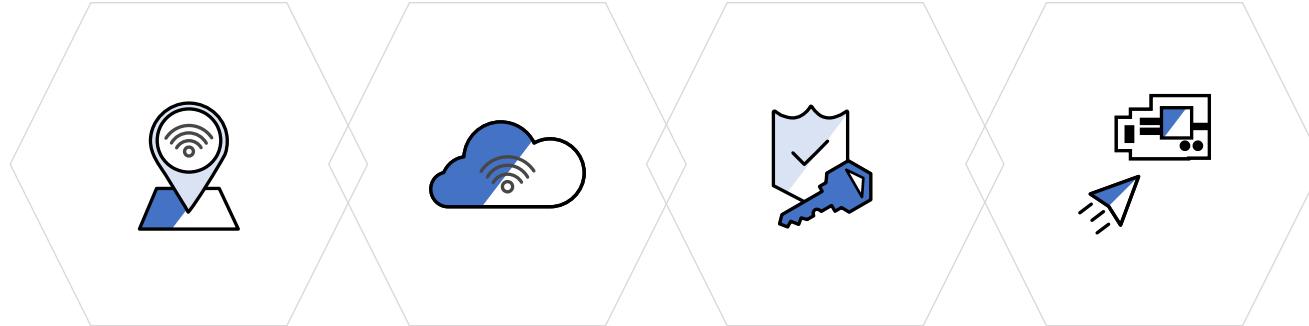


Amazon FreeRTOS (GA)

IoT operating system for microcontrollers



Based on FreeRTOS kernel



Local connectivity
libraries



AWS Greengrass

Cloud connectivity
libraries



AWS IoT Core

Security
libraries

OTA &
code signing

Coming soon

Augmented & Virtual Reality

Amazon Sumerian (Preview)

The fastest and easiest way to create VR, AR, and 3D experiences



Create and run VR, AR, and 3D apps

Browser editor based on WebGL/WebVR

Create digital characters called Hosts

Script the logic within scenes

Run on platforms including Oculus Rift, HTC Vive, and mobile devices





GO BUILD

<https://aws.amazon.com/new/reinvent/>



THANK YOU!

Julien Simon, Principal AI/ML Evangelist, EMEA
@julsimon