

Keeping your sanity with the Deep Learning AMI

Julien Simon

Principal Technical Evangelist, AI and Machine Learning, AWS

@julsimon

Getting into Deep Learning, day 1

Deep Learning, yeah!

Let's create an EC2 instance and get to work!

Getting into Deep Learning, day 5

distributed training

virtualenv

CUDA 8.x/9.x

GPU drivers

Jupyter

Microsoft CNTK

ONNX

Horovod

Tensorboard

Chainer

Caffe

pip

Deep Learning, WTF!

Theano

Torch

Numpy

Pandas

Keras

Security Group

Python 2

Python 3

Tensorflow

Intel MKL

Apache MXNet

PyTorch

Model Serving

Gluon

Anaconda

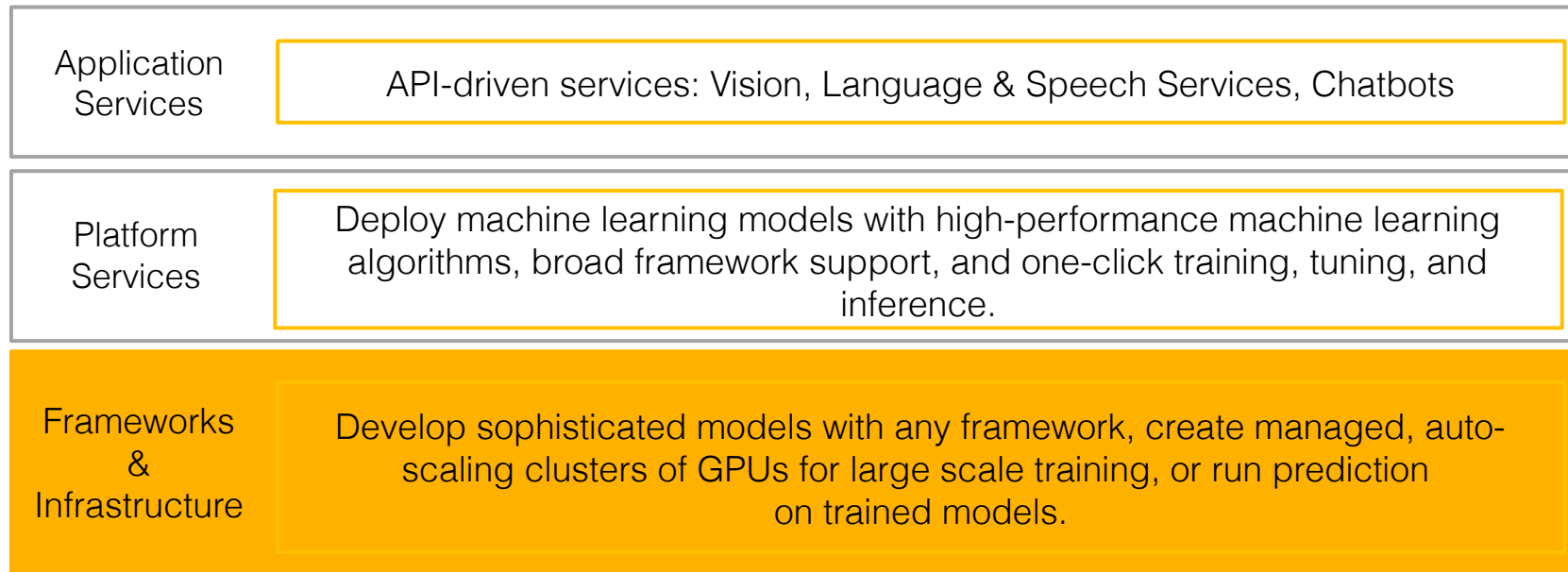
Our mission

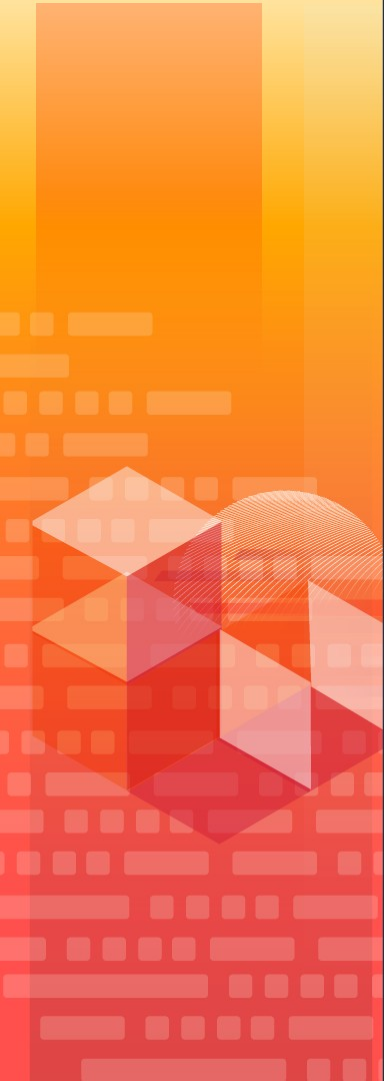
Put Machine Learning in the hands
of every developer and data scientist

AWS ML Stack

Application Services	API-driven services: Vision, Language & Speech Services, Chatbots
Platform Services	Deploy machine learning models with high-performance machine learning algorithms, broad framework support, and one-click training, tuning, and inference.
Frameworks & Infrastructure	Develop sophisticated models with any framework, create managed, auto-scaling clusters of GPUs for large scale training, or run prediction on trained models.

AWS ML Stack





The Deep Learning AMI

The Deep Learning AMI, minute 1

distributed training

virtualenv

CUDA 8.x/9.x

GPU drivers

Jupyter

Microsoft CNTK

ONNX

Horovod

Tensorboard

Chainer

Caffe

pip

Theano

Focus on ML/DL, not on setup!

Torch

Numpy

Pandas

Keras

Security Group

Python 2

Python 3

Tensorflow

Intel MKL

Apache MXNet

PyTorch

Model Serving

Gluon

Anaconda

AWS Deep Learning AMIs

Preconfigured environments to quickly build Deep Learning applications

Conda AMI

For developers who want pre-installed pip packages of DL frameworks in separate virtual environments.

Base AMI

For developers who want a clean slate to set up private DL engine repositories or custom builds of DL engines.

AMI with source code

For developers who want preinstalled DL frameworks and their source code in a shared Python environment.



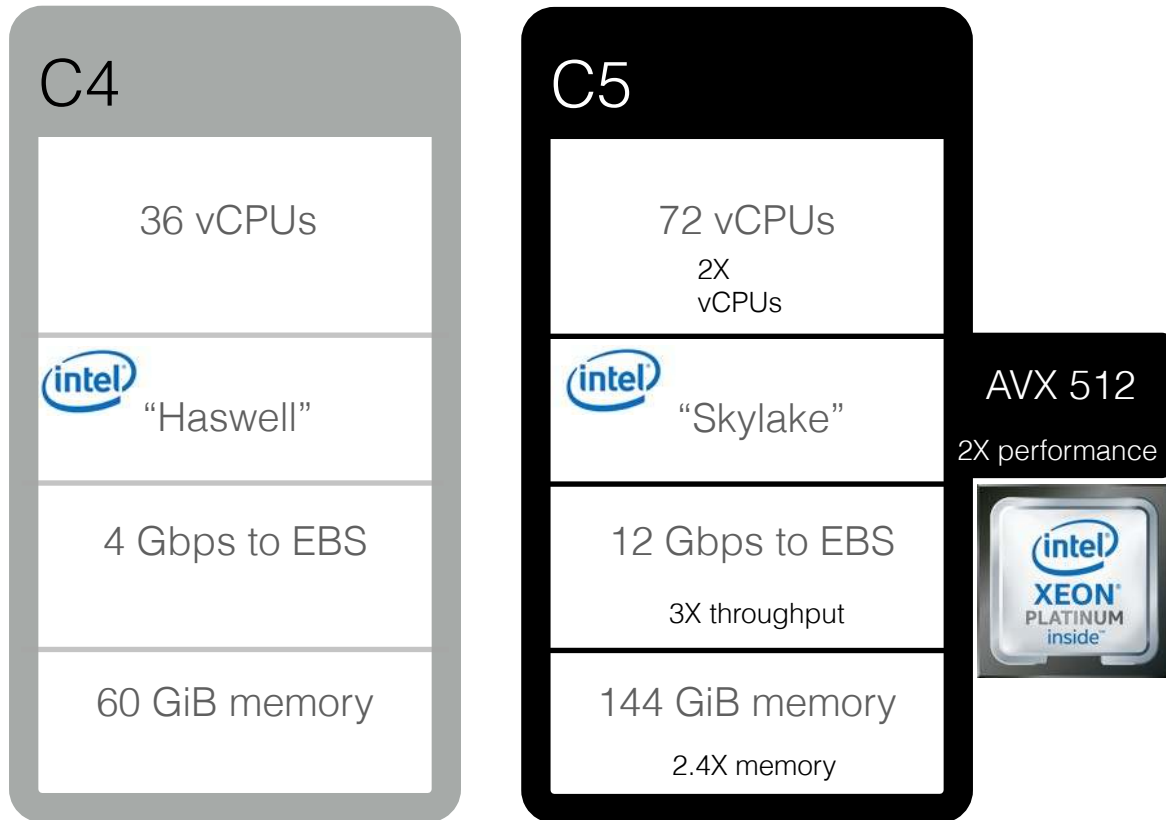
<https://aws.amazon.com/machine-learning/amis/>

Amazon EC2 C5 / C5d instances

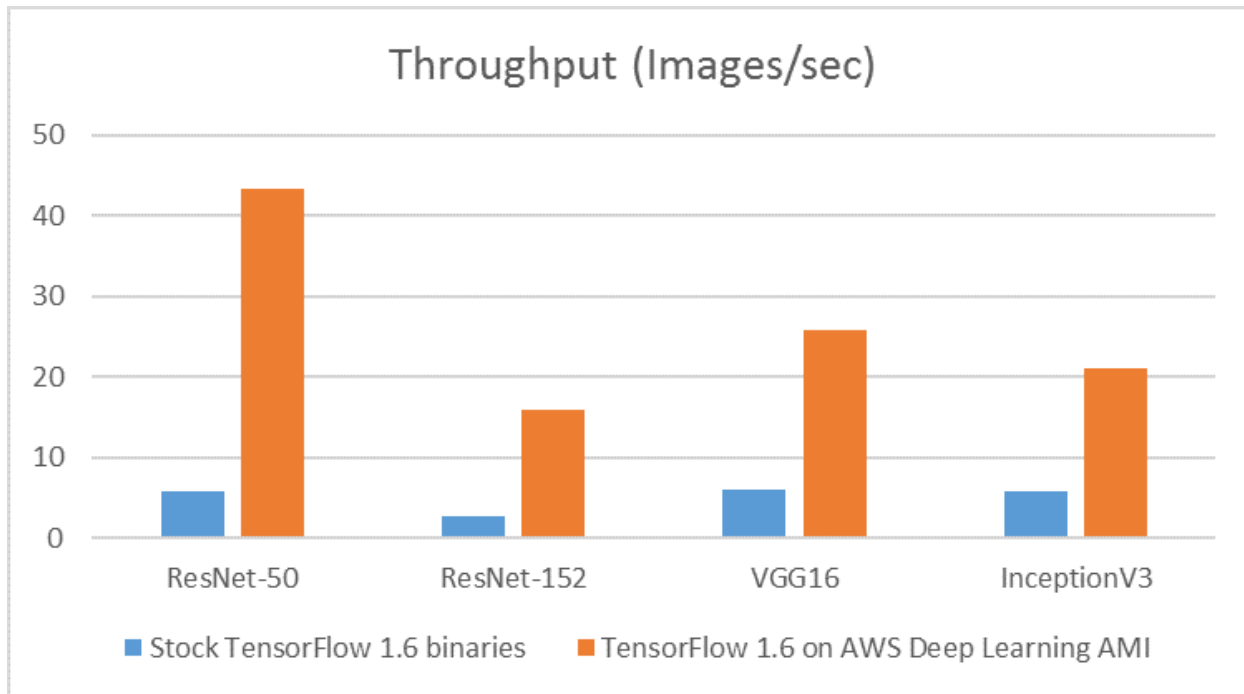
C5: Next Generation
Compute-Optimized
Instances with
Intel® Xeon® Scalable
Processor

AWS Compute optimized
instances support the new
Intel® AVX-512 advanced
instruction set, enabling you to
more efficiently run vector
processing workloads with
single and double floating point
precision, such as AI/machine
learning or video processing.

25% improvement in
price/performance over C4



Faster TensorFlow training on C5



<https://aws.amazon.com/blogs/machine-learning/faster-training-with-optimized-tensorflow-1-6-on-amazon-ec2-c5-and-p3-instances/>



Amazon EC2 P3 Instances

The fastest, most powerful GPU instances in the cloud

- P3.2xlarge, P3.8xlarge, P3.16xlarge
- Up to eight NVIDIA Tesla V100 GPUs in a single instance
 - 40,960 CUDA cores, 5120 Tensor cores
 - 128GB of GPU memory
- 1 PetaFLOPs of computational performance – 14x better than P2
- 300 GB/s GPU-to-GPU communication (NVLink) – 9x better than P2



Demo:

start a p3 instance running the
Deep Learning AMI



Demo:

train a Tensorflow model and
visualize training with TensorBoard



Demo:

train a Tensorflow model and
serve it with Tensorflow Model Server



Demo:

work with Jupyter notebooks

... or you could use notebook
instances in Amazon SageMaker ;)

Resources

<https://ml.aws>

<https://aws.amazon.com/machine-learning/amis/>

<https://medium.com/@julsimon>

Floor28 Agenda

Big Data Day
14 Oct

Technical Sessions

Serverless Data Workshop

Big Data UG Meetup

ML & DL Day
15 Oct

Technical Sessions

SageMaker Workshop

ML&DL Meetup

DevOps Day
16 Oct

Technical Sessions

K8s Workshop

DevOps Meetup

DevOps Day
17 Oct

Technical Sessions

Spot Workshop

Databases Day
18 Oct

Technical Sessions

PyTorch Meetup

Builders' Day
Serverless backend
21 Oct

Technical Sessions

Serverless Workshop

Virtual assistants UG Meetup

Builders' Day
AppSync, Alexa & IoT
22 Oct

Technical Sessions

CDK Workshop

AWS IL UG Meetup

Enterprise IT Day
23 Oct

Technical Sessions

GameDay
24 Oct





GAME DAY

PUT YOUR SKILLS TO THE TEST

OCT 24

Register now: bit.ly/Floor28GameDay



SSID: Guest

Password: Cube@11999

INNOVATE2018

ONLINE CONFERENCE

DEVELOPER EDITION

<https://amzn.to/2DAEUW8>



Thank you!

Julien Simon

Principal Technical Evangelist, AI and Machine Learning

@julsimon