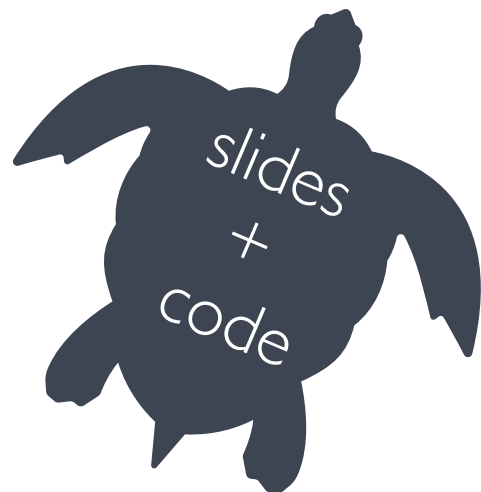


# Machine learning in the cloud with Amazon Web Services

---



Miami Machine Learning Meetup - 2/8/18



<https://github.com/rikturr/aws-ml-experimenter>

# About Me

---

- Aaron Richter
- Data Scientist, PhD Student
-  @rikturr  [rikturr@gmail.com](mailto:rikturr@gmail.com)
- \*\* We're hiring!
- [modmed.com/careers](http://modmed.com/careers)



# Agenda

---

- The cloud?
- AWS concepts (acronyms galore!)
- AWS account setup (walkthrough)
- Run models! (Jupyter)
- aws-ml-experimenter introduction (time permitting)

# The Cloud

---



# The Cloud

---

- Ease of use
  - No hardware to deal with
- Scalability
  - Instances with dozens of CPUs and hundreds of GBs of memory
- Cost
  - Pay for usage

# The Cloud

◀ Back to search results for "tesla k80"



Roll over image to zoom in

HP

HP J0G95A NVIDIA Tesla K80 - GPU computing processor - 2 GPUs - Tesla K80 - 24 GB GDDR5 - PCI Express 3.0 x16 - fanless



11 customer reviews | 5 answered questions

Price: **\$2,699.99 & FREE Shipping**

Item is eligible: No interest if paid in full within 12 months with the Amazon.com Store Card.

**Note:** Not eligible for Amazon Prime.

**Only 3 left in stock - order soon.**

**Get it as soon as Feb. 6 - 8** when you choose **Expedited Shipping** at checkout.

Ships from and sold by **SabrePC** in **easy-to-open packaging**.

Service: **Get professional installation**[Details](#)

**Without expert installation**

Include installation  
+\$76.39

⌵ [See more](#)

[https://www.amazon.com/HP-J0G95A-NVIDIA-Tesla-K80/dp/B00TWFEIWA/ref=sr\\_1\\_3?ie=UTF8&qid=1517774449&sr=8-3&keywords=tesla+k80&dpID=51C7-cCsGGL&preST=\\_SY300\\_QL70\\_&dpSrc=srch](https://www.amazon.com/HP-J0G95A-NVIDIA-Tesla-K80/dp/B00TWFEIWA/ref=sr_1_3?ie=UTF8&qid=1517774449&sr=8-3&keywords=tesla+k80&dpID=51C7-cCsGGL&preST=_SY300_QL70_&dpSrc=srch)

# The Cloud

## P2

P2 instances are intended for general-purpose GPU compute applications.

**Features:**

- High frequency Intel Xeon E5-2686 v4 (Broadwell) processors
- High-performance NVIDIA K80 GPUs, each with 2,496 parallel processing cores and 12GiB of GPU memory
- Supports GPUDirect™ for peer-to-peer GPU communications
- Provides Enhanced Networking using Elastic Network Adapter (ENA) with up to 25 Gbps of aggregate network bandwidth within a Placement Group
- EBS-optimized by default at no additional cost

Model	GPUs	vCPU	Mem (GiB)	GPU Memory (GiB)
p2.xlarge	1	4	61	12
p2.8xlarge	8	32	488	96
p2.16xlarge	16	64	732	192

<https://aws.amazon.com/ec2/instance-types/>

**GPU Instances - Current Generation**

p2.xlarge	4	12	61	EBS Only	\$0.9 per Hour
p2.8xlarge	32	94	488	EBS Only	\$7.2 per Hour
p2.16xlarge	64	188	732	EBS Only	\$14.4 per Hour

<https://aws.amazon.com/ec2/pricing/>



- Elastic Compute Cloud (EC2)
  - Instance types/pricing: on-demand vs. spot
  - Amazon Machine Images (AMI)
  - Security: VPC, subnets, security groups
- Simple Storage Service (S3)
- Identity and Access Management (IAM)
  - Users, groups, roles, policies



# AWS account setup

---

- Create AWS account
- Create user with S3 and EC2 access
- Configure credentials locally
- Create S3 bucket (+CyberDuck)
- Setup IAM role (for EC2 to access S3)
- Open ports in security group (SSH - 22, Tensorboard - 6006)
- Launch EC2 instance from console and SSH into it

Code time!



```
ssh -i <PEM> -L 8000:localhost:8888 ec2-user@<HOST>
```

```
jupyter notebook --no-browser --port=8888
```

<https://towardsdatascience.com/setting-up-and-using-jupyter-notebooks-on-aws-61a9648db6c5>