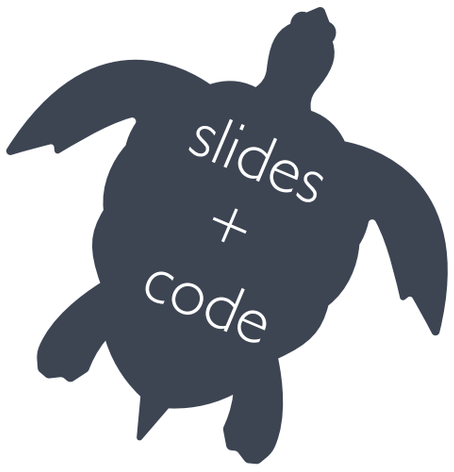
Machine learning in the cloud with Amazon Web Services

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

MODERNIZING MEDICINE

FAU
FLORIDA ATLANTIC UNIVERSITY

- ** We're hiring!
  - modmed.com/careers

Agenda

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

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
  - (I spent $1.55 preparing for this talk)

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

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/

AWS

- Elastic Compute Cloud (EC2)
  - Instance types/pricing: on-demand vs. spot
  - Amazon Machine Images (AMI)
  - Storage: EBS volumes
  - Networking/Security: Availability zones, VPC, subnets, security groups
  - Key pairs

https://aws.amazon.com/ec2/

AWS

- Simple Storage Service (S3)
  - Object storage
- Identity and Access Management (IAM)
  - Users, groups (people)
  - Roles (services)
  - Policies (permissions)

https://aws.amazon.com/s3/ https://aws.amazon.com/iam/

AWS account setup - notebook

- Create AWS account
- Create S3 bucket
- Setup IAM role (for EC2 to access S3)
- Launch EC2 instance from AWS Console
  - Download key file!
- Open ports in security group: SSH - 22, Tensorboard - 6006
- SSH onto instance and launch Jupyter

Code time!

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

jupyter notebook --no-browser —port=8000
```

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

AWS account setup - production

- Install AWS CLI (`pip install awscli`)
- Create user with S3 and EC2 access
  - IAM pass role
- Configure credentials locally (region us-east-1)
- CyberDuck for nice S3

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": “*"
     }]
}
```

https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-getting-started.html

https://aws.amazon.com/blogs/security/granting-permission-to-launch-ec2-instances-with-iam-roles-passrole-permission/

More code!

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

Resources + Extra Stuff

- ***Good intro article***: https://www.dataquest.io/blog/introduction-to-aws-for-data-scientists/

- ***Boto3 Python package***: https://github.com/boto/boto3

- ***Amazon SageMaker***: https://aws.amazon.com/sagemaker/

- ***Pywren***: http://pywren.io/

    - ***AWS Lambda***: https://aws.amazon.com/lambda/

Thank you!

- Aaron Richter
- Data Scientist, PhD Student
- @rikturr rikturr@gmail.com

MODERNIZING MEDICINE

FAU
FLORIDA ATLANTIC UNIVERSITY

- ** We're hiring!
  - modmed.com/careers