RUNNING AND ANALYZING MACHINE LEARNING EXPERIMENTS IN THE CLOUD WITH AWS, PYTHON, AND R

Aaron Richter, PhD

IEEE ICMLA 2019

HI, I'M AARON

Data Scientist, Modernizing Medicine

PhD, Florida Atlantic University

@rikturr

rikturr@gmail.com

TUTORIAL GOALS

- ***Machine learning experimentation***

- Training and evaluating a large number of machine learning models with differing hardware and software requirements
- Enhance efficiency using the cloud (bigger machines, more machines, parallelization)

- Out of scope:
  - Data processing
  - Machine learning theory
  - Model deployment / serving

TECHNOLOGIES

- **Python**
  - Jupyter
  - *pandas*
  - *scikit-learn*
  - *keras*
- **R**
  - RStudio
  - *tidyverse (dplyr/ggplot2)*
  - *shiny (shinyapps.io)*
- **Amazon web services**
  - S3
  - EC2
  - SageMaker

RUNNING AND ANALYZING MACHINE LEARNING EXPERIMENTS IN THE CLOUD WITH AWS, PYTHON, AND R

SCHEDULE

Part 1	1:00pm - 1:30pm	Introduce the cloud and AWS	Slides
	1:30pm - 2:15pm	Setup AWS environment	Web / Terminal
	2:15pm - 2:30pm	*Break*	Coffee
Part 2	2:30pm - 3:15pm	ML with Medicare data	Jupyter
	3:15pm - 4:00pm	Shiny dashboard with R	R Studio / Web
	4:00pm - 4:15pm	*Break*	Coffee
Part 3	4:15pm - 5:00pm	Heavy experimentation	All of the above

WHAT IS THE CLOUD?


There is no cloud
it's just someone else's computer

WHY 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 <$10 preparing for this tutorial)

Nvidia Tesla v100 16GB

by PNY

8 ratings | 8 answered questions

Price: $5,834.99 & FREE Shipping

Thank you for being a Prime member. Get $100 off instantly: Pay $5,734.99 upon approval for the Amazon Prime Rewards Visa Card. No annual fee.

Note: Not eligible for Amazon Prime.

Arrives before Christmas. Choose delivery option in checkout.

Service: **Get professional installation** Details

Without expert installation | Include installation +$70.76 per unit

See more

- TCSV100MPCIE-PB
- 3536403359041

Compare with similar items

New (3) from $5,834.99 + FREE Shipping

Report incorrect product information.

https://www.amazon.com/PNY-TCSV100MPCIE-PB-Nvidia-Tesla-v100/dp/B076P84525/ref=sr_1_3?keywords=Tesla+v100&qid=1576356059&s=pc&sr=1-3

P3 | P2 | Inf1 | G4 | G3 | F1

P3 instances are the latest generation of general purpose GPU instances.

Features:

- Up to 8 NVIDIA Tesla V100 GPUs, each pairing 5,120 CUDA Cores and 640 Tensor Cores
- High frequency Intel Xeon E5-2686 v4 (Broadwell) processors for p3.2xlarge, p3.8xlarge, and p3.16xlarge.
- High frequency 2.5 GHz (base) Intel Xeon P-8175M processors for p3dn.24xlarge.
- Supports NVLink for peer-to-peer GPU communication
- Provides up to 100 Gbps of aggregate network bandwidth.
- EFA support on p3dn.24xlarge instances

Instance	GPUs	vCPU	Mem (GiB)	GPU Mem (GiB)	GPU P2P	Storage (GB)	Dedicated EBS Bandwidth	Networking Performance
p3.2xlarge	1	8	61	16	-	EBS-Only	1.5 Gbps	Up to 10 Gigabit
p3.8xlarge	4	32	244	64	NVLink	EBS-Only	7 Gbps	10 Gigabit

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

P3 | P2 | Inf1 | G4 | G3 | F1

P3 instances are the latest generation of general purpose GPU instances.

Features:

- Up to 8 NVIDIA Tesla V100 GPUs, each pairing 5,120 CUDA Cores and 640 Tensor Cores
- High frequency Intel Xeon E5-2686 v4 (Broadwell) processors for p3.2xl
- High frequency 2.5 GHz (base) Intel Xeon P-8175M processors for p3dn.
- Supports NVLink for peer-to-peer GPU communication
- Provides up to 100 Gbps of aggregate network bandwidth.
- EFA support on p3dn.24xlarge instances

GPU Instances - Current Generation

p3.2xlarge	8	26	61 GiB	EBS Only	$3.06 per Hour
p3.8xlarge	32	94	244 GiB	EBS Only	$12.24 per Hour
p3.16xlarge	64	188	488 GiB	EBS Only	$24.48 per Hour

Instance	GPUs	vCPU	Mem (GiB)	GPU Mem (GiB)	GPU P2P	Storage (GB)	Dedicated EBS Bandwidth	Networking Performance
p3.2xlarge	1	8	61	16	-	EBS-Only	1.5 Gbps	Up to 10 Gigabit
p3.8xlarge	4	32	244	64	NVLink	EBS-Only	7 Gbps	10 Gigabit

https://aws.amazon.com/ec2/pricing/on-demand/

P3 | P2 | Inf1 | G4 | G3 | F1

P3 instances are the latest generation of general purpose GPU instances.

Features:

- Up to 8 NVIDIA Tesla V100 GPUs, each pairing 5,120 CUDA Cores and 640 Tensor Cores
- High frequency Intel Xeon E5-2686 v4 (Broadwell) processors for p3.2xl
- High frequency 2.5 GHz (base) Intel Xeon P-8175M processors for p3dn.
- Supports NVLink for peer-to-peer GPU communication
- Provides up to 100 Gbps of aggregate network bandwidth.
- EFA support on p3dn.24xlarge instances

GPU Instances - Current Generation

p3.2xlarge	8	26	61 GiB	EBS Only	$3.06 per Hour
p3.8xlarge	32	94	244 GiB	EBS Only	$12.24 per Hour
p3.16xlarge	64	188	488 GiB	EBS Only	$24.48 per Hour

Instance	GPUs	vCPU	Mem (GiB)	GPU Mem (GiB)	GPU P2P	Storage (GB)	Dedicated EBS Bandwidth	Networking Performance
p3.2xlarge	1	8	61	16	-	EBS-Only	1.5 Gbps	Up to 10 Gigabit
p3.8xlarge	4	32	244	64	NVLink			

p3.2xlarge	$0.9225 per Hour
p3.8xlarge	$3.672 per Hour
p3.16xlarge	$7.344 per Hour

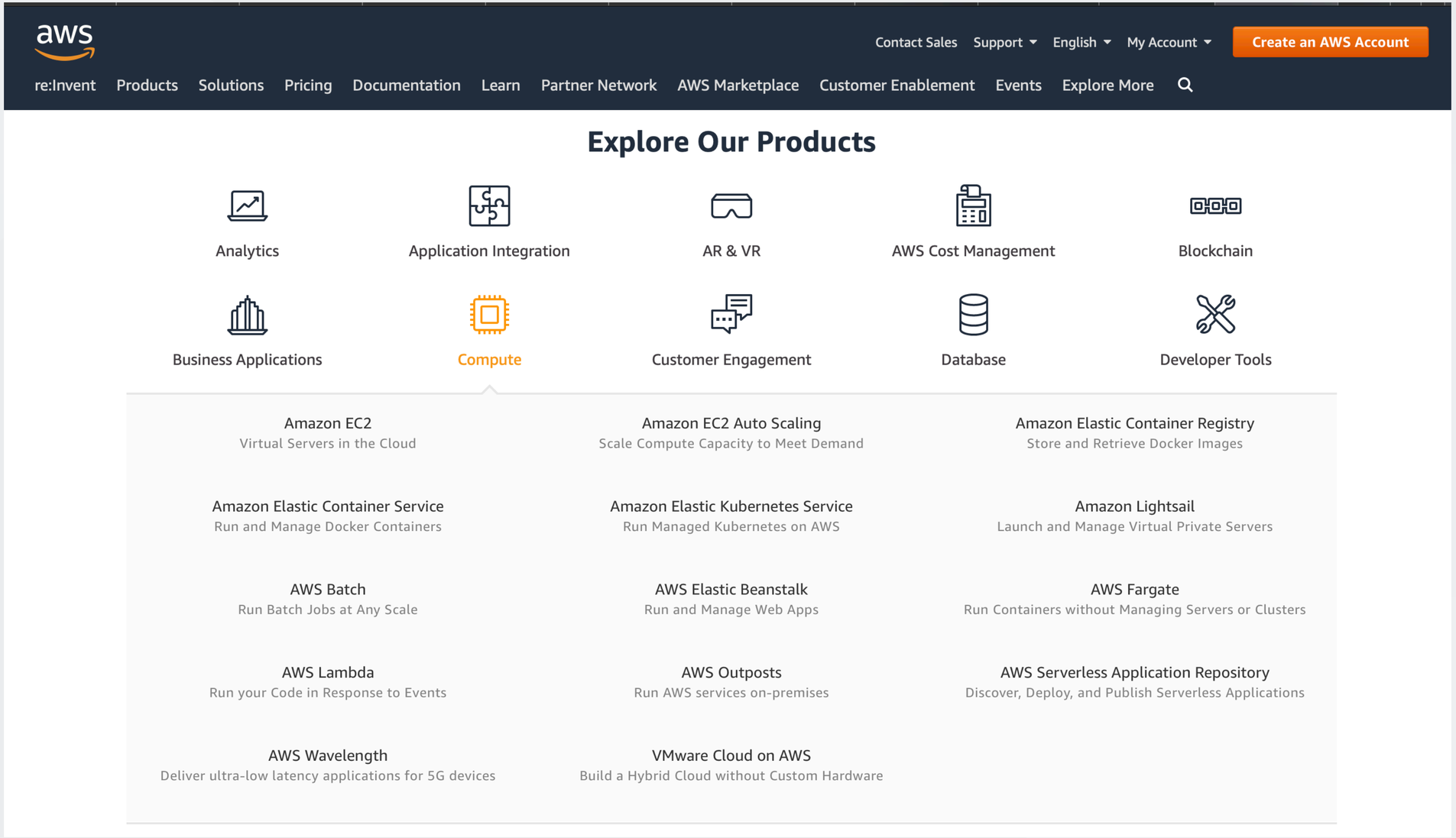
https://aws.amazon.com/ec2/spot/pricing/

AWS SERVICES

S3

EC2

SageMaker



https://aws.amazon.com/products/

S3

- *Simple Storage Service*
- “infinite”, “always up” data storage

- Cheap to store data
- Free to transfer data in
- Cheap to transfer between AWS services
- Less cheap to transfer data out (download to laptop, access from web)
  - $0.09/GB

- Object storage
  - There are no directories
  - s3://<BUCKET>/PREFIX
    - bucket: think domain name
    - prefix: think file path

EC2

- Elastic Compute Cloud
- Computing machinery on-demand

- Price depends on type of instance
  - Compute optimized, memory optimized, GPU, etc.
  - Spot pricing helps for temporary workloads

- Can launch from pre-configured images
  - Amazon Machine Images (AMI)
  - Linux and Windows (but lets be honest)

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

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

EC2

- Other things to know…

- Identity and Access Management (IAM)
  - Users, groups (people)
  - Roles (services)
  - Policies (permissions)

- Storage: EBS volumes
- Networking/Security: Availability zones, VPC, subnets, security groups
- Key pairs

SAGEMAKER

- Machine learning lifecycle services
  - Labelling, experimenting, training, inference, hosting

- Easier (less configuration/management) than EC2
- Instances more expensive than EC2

https://aws.amazon.com/sagemaker/

https://aws.amazon.com/sagemaker/pricing/

https://docs.aws.amazon.com/sagemaker/latest/dg/model-managed-spot-training.html

LET'S USE THE CLOUD!

- Zero to Jupyter notebook running in the cloud

LET'S USE THE CLOUD!

(EC2)

1. Create AWS account
2. Create S3 bucket
3. Setup IAM role (for EC2 to access S3)
4. Launch EC2 instance from AWS Console
5. Download key file!
6. Open ports in security group: SSH - 22
7. SSH onto instance and launch Jupyter

(SageMaker)

1. Create AWS account
2. Create S3 bucket
3. SageMaker -> Notebook instance
4. Click "Open Jupyter"

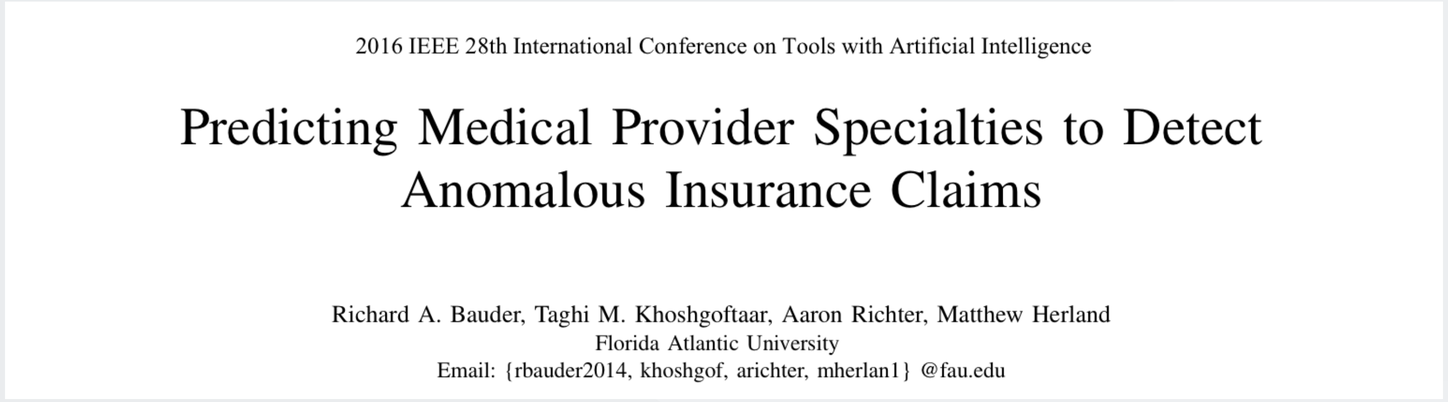
```
ssh -i <PEM> -L 8000:localhost:8000 ec2-user@<HOST>

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

BREAK

Come back by 2:45pm!

MEDICARE ML EXAMPLES



2016 IEEE 28th International Conference on Tools with Artificial Intelligence

Predicting Medical Provider Specialties to Detect Anomalous Insurance Claims

Richard A. Bauder, Taghi M. Khoshgoftaar, Aaron Richter, Matthew Herland
Florida Atlantic University
Email: {rbauder2014, khoshgof, arichter, mherlan1} @fau.edu

https://github.com/rikturr/aws-ml-experimenter/blob/master/examples/medicare-data.ipynb

https://github.com/rikturr/aws-ml-experimenter/blob/master/examples/scikit-learn.ipynb

https://github.com/rikturr/aws-ml-experimenter/blob/master/examples/keras.ipynb

BREAK

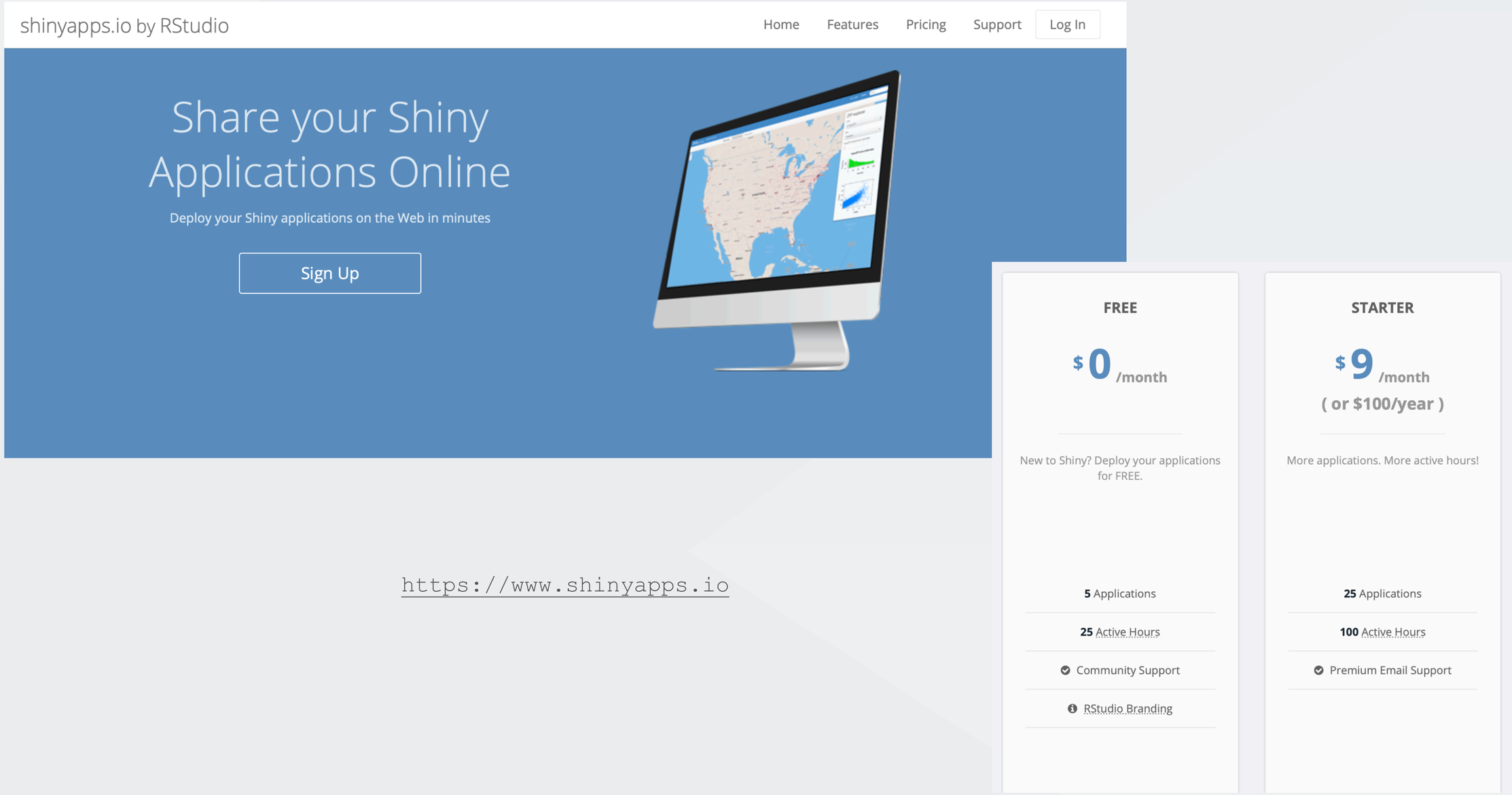
Come back by 4:15pm!

SHINY DASHBOARD WITH R

SET UP R/RSTUDIO

- Install R https://cran.r-project.org
- Install RStudio https://rstudio.com/products/rstudio/download/

- Packages
  - tidyverse https://www.tidyverse.org
  - aws.s3 https://github.com/cloudyr/aws.s3



https://www.shinyapps.io

HEAVY EXPERIMENTATION

(How I made all my papers)

HOW TO ACCESS AWS

- Console (Web UI) https://console.aws.amazon.com
- Terminal (AWS CLI) https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-install.html
- boto3 (Python API) https://boto3.amazonaws.com/v1/documentation/api/latest/index.html

RANDOM THINGS TO DO

- Request limit increases (support ticket)
- Install aws cli `pip install awscli`
- Create user with S3 and EC2 access
  - IAM pass role
- Configure credentials locally (region us-east-1)
- Forklift for nice S3

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

https://apps.apple.com/us/app/forklift-file-manager-and-ftp-sftp-webdav-amazon-s3-client/id412448059?mt=12

WORKING EXAMPLE

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

https://faubigdata.shinyapps.io/dashboard/

THANK YOU!

Data Scientist, Modernizing Medicine

PhD, Florida Atlantic University

@rikturr

rikturr@gmail.com