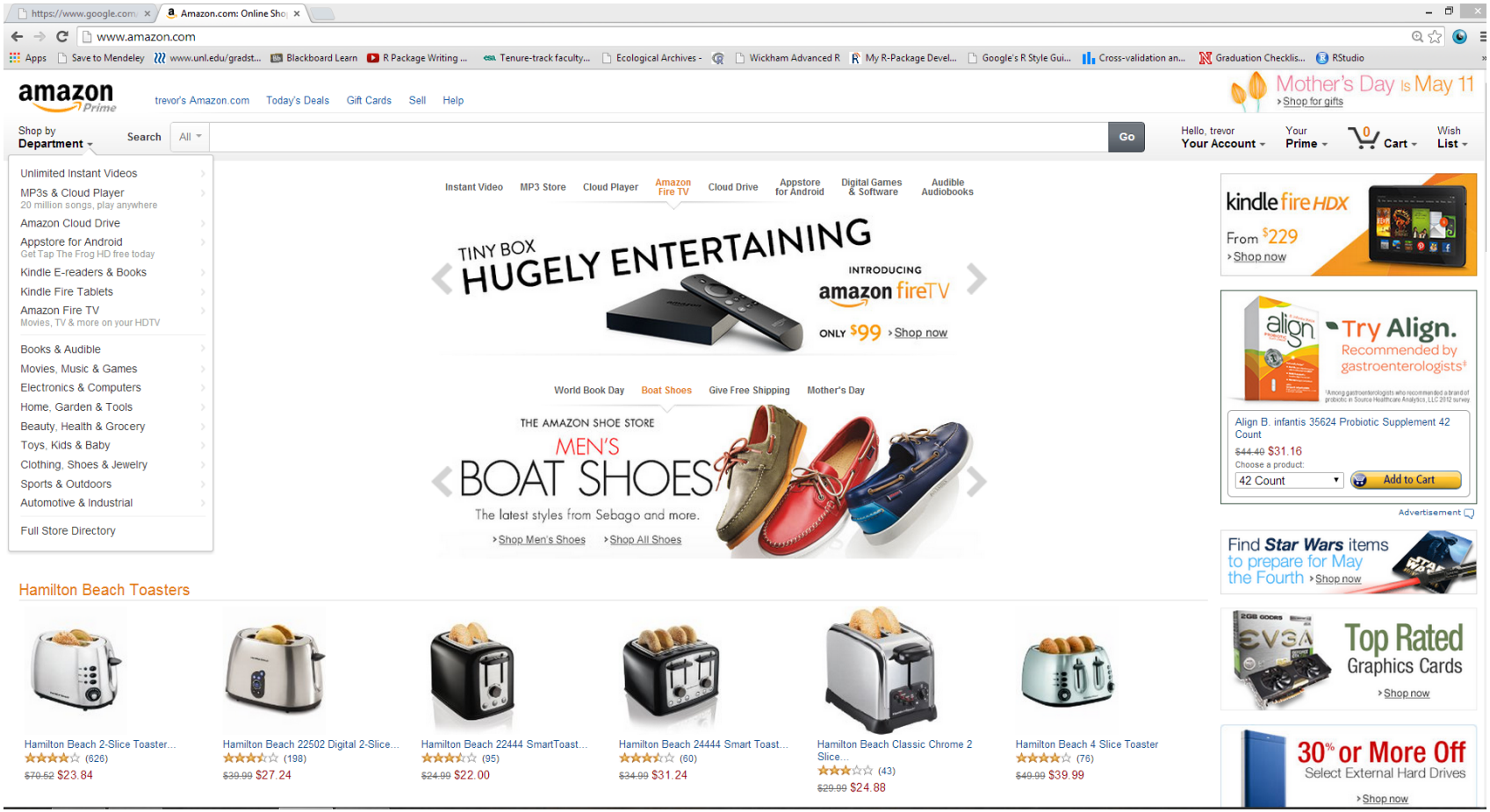
Running R from Amazon's Elastic Compute Cloud

Trevor Hefley

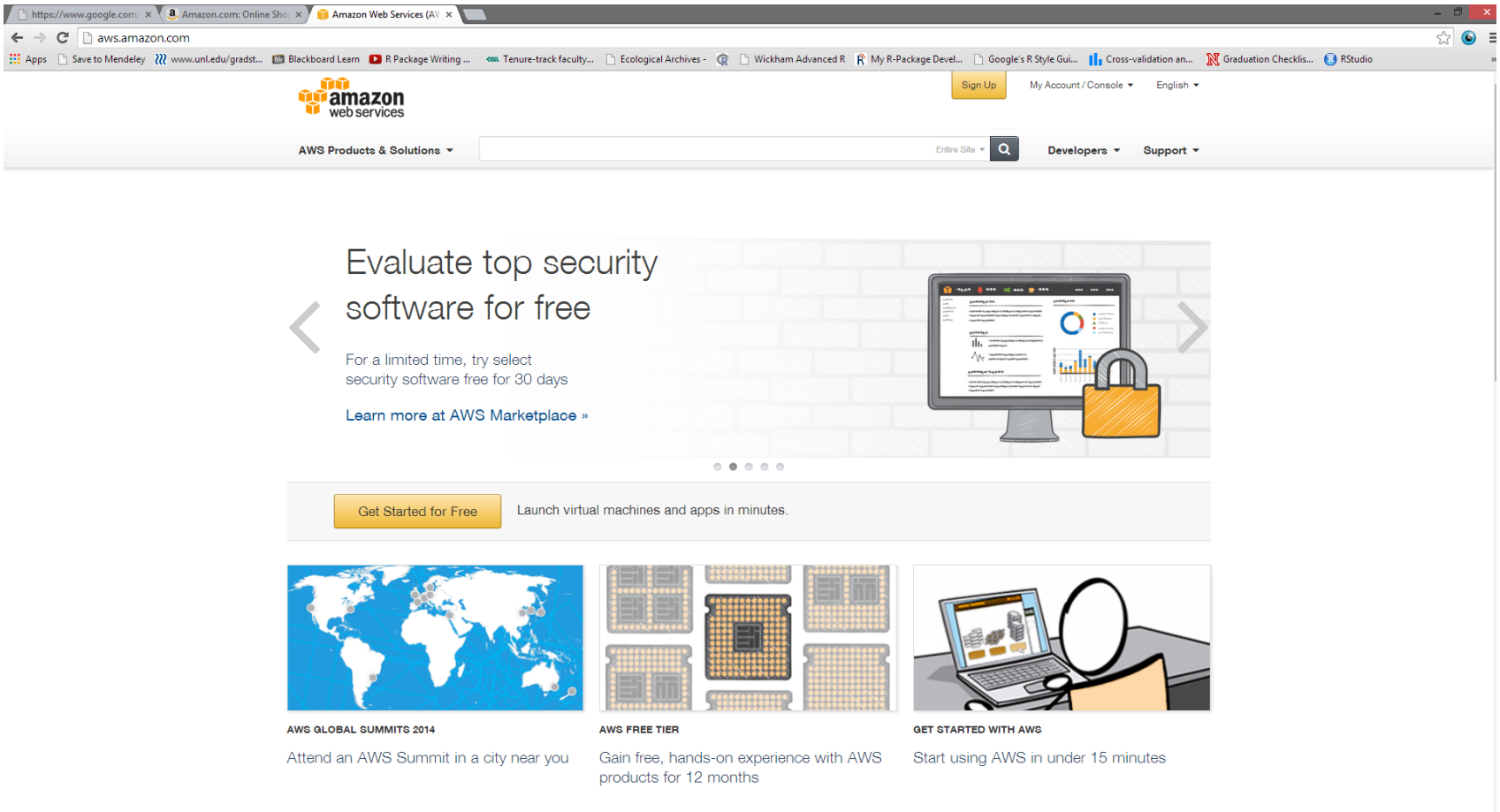
Department of Statistics
University of Nebraska–Lincoln

April 30, 2014

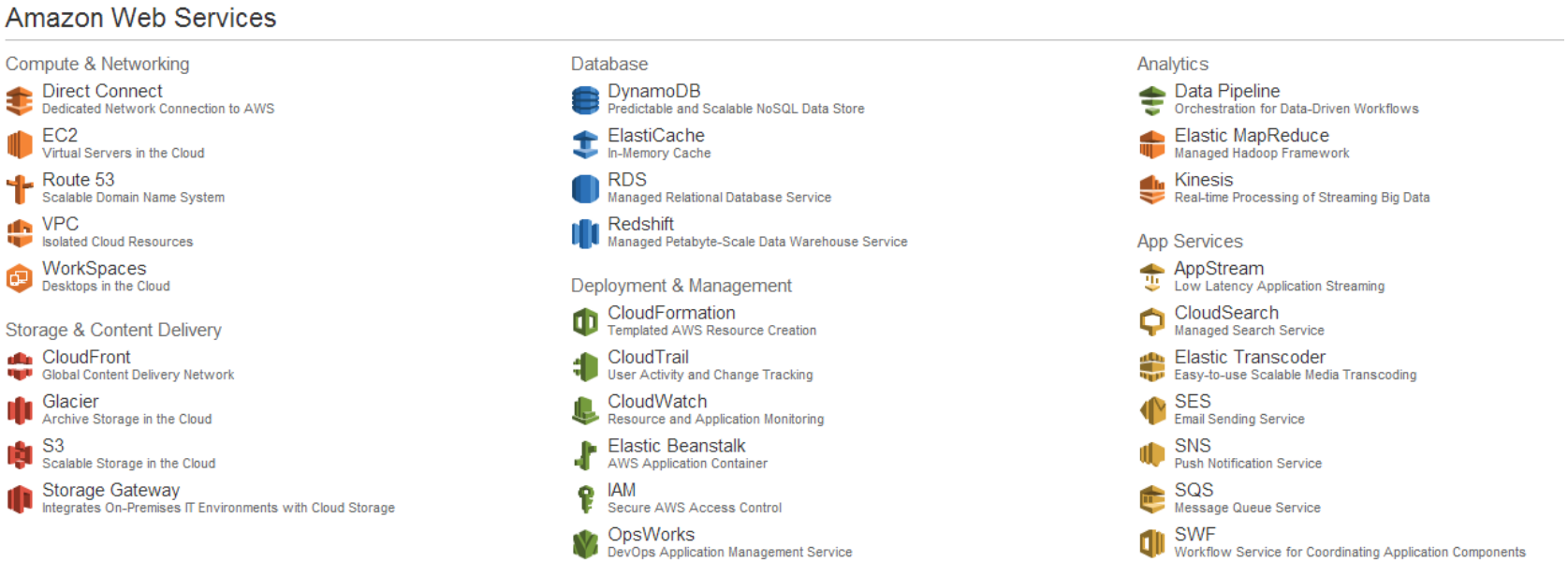
Introduction



Introduction



Introduction



Elastic Compute Cloud (EC2)

- On demand, scalable cloud computing service provided by Amazon.
- You pay for the service when you need it.
- You pay only for the amount you use.
- There are other ways to purchase (e.g., bidding on unused services, renting a server).

Terminology

- Instance = Node
- Amazon Machine Image (AMI) = operating system + programs
- Instance + AMI = "virtual computer"

Instances

- Instance = Node
- Free tier (750 hours EC2 running Linux/Unix Micro Instance and 750 hours running Microsoft Windows Servers Micro Instance)
- `http://aws.amazon.com/ec2/pricing/`

Instances

- vCPU = threads
- ECU = One EC2 Compute Unit provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor.
- Memory (GiB) = RAM
- Instance Storage (GB) = "hard drive" storage
- SSD = solid state drive
- `http://aws.amazon.com/ec2/pricing/`

Amazon Machine Image

- Amazon Machine Image (AMI) = operating system + programs
- Build your own AMI
- You can save, sell, or share your AMI
- Pre-made AMI (e.g., Revolution Analytics `https://aws.amazon.com/marketplace/seller-profile?id=3c6536d3-8115-4bc0-a713-be58e257a7be`

The easy way

- Louis Aslett (postdoc at Oxford)
- See link below and follow instructions or my video tutorial on Black Board
- Demonstration
- `http://www.louisaslett.com/RStudio_AMI/`

The easy way

- Easy to install packages
- Easy to upload files (e.g., R code)
- Run Rstudio from web browser
- Easy to change instance size/type
- I have not found an easy way to install other software with pre-made AMIs (e.g., JAGS, LYX/knitr).

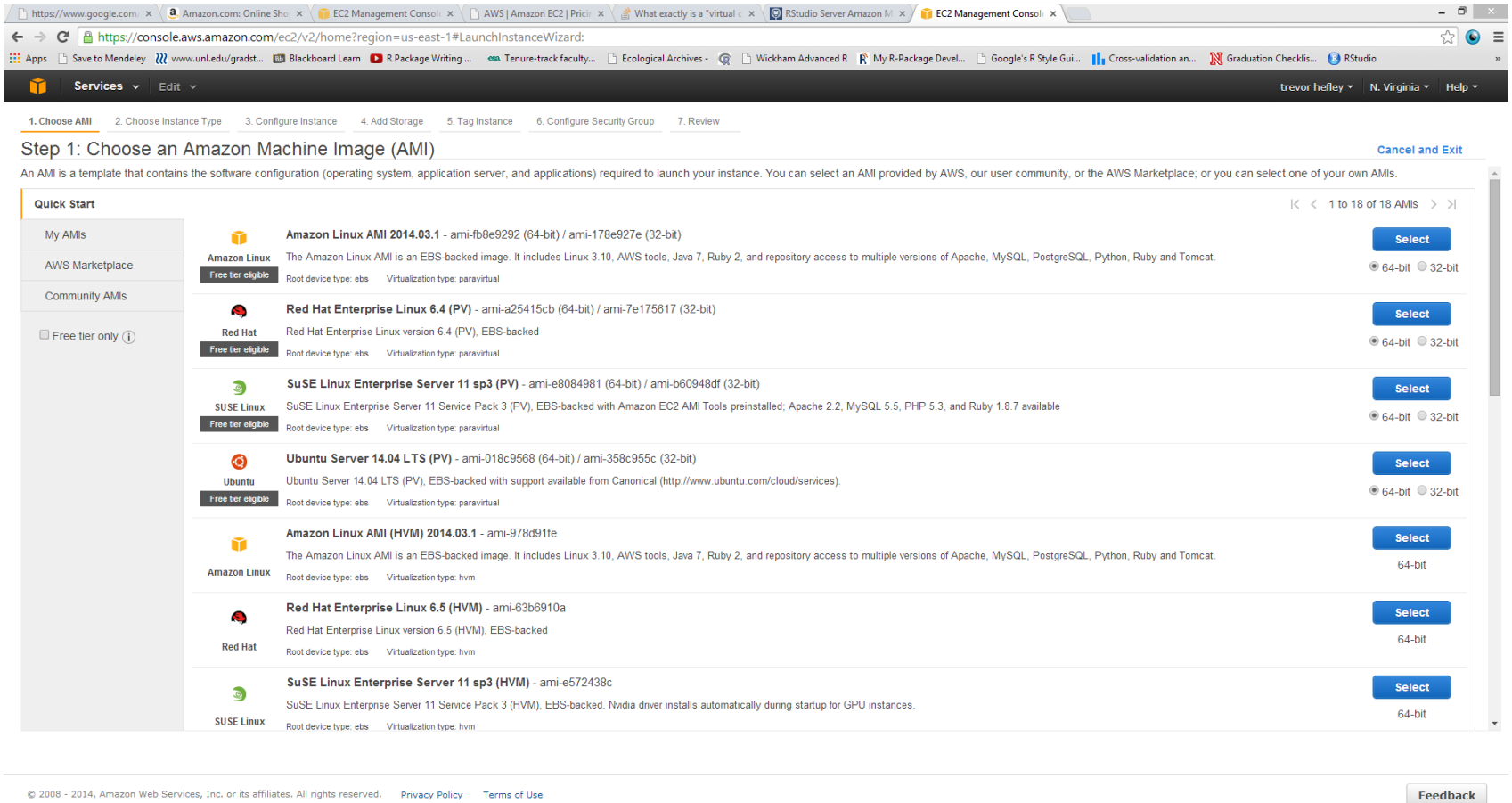
Building your own AMI

- Equivalent to customizing your own computer
- You can put whatever programs you want on your own AMI
- You can choose the operating system

Building your own AMI

- "Launching an instance" = first step to building and running your virtual machine
- Choose a AMI

Building your own AMI



Building your own AMI

- Many different option (18 as of 4/24/2014)
- Linux dominates the list and is cheaper
- Windows Server

Building your own AMI

- Not enough time in this presentation to build a Windows Server AMI
- See tutorial video on Black Board
- Demonstration of Window Server and R

Parallel computing assignment revisited

- All times reported were using the foreach package
- My computer is a Lenovo Y510p with an Intel i7-4700MQ processor (3.4 GHz) with four cores and 16 GB of RAM.
- I tested the t1.micro, m3.xlarge and c3.2xlarge instances

Parallel computing assignment revisited

	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Windows Usage
General Purpose - Current Generation					
m3.medium	1	3	3.75	1 x 4 SSD	$0.133 per Hour
m3.large	2	6.5	7.5	1 x 32 SSD	$0.266 per Hour
m3.xlarge	4	13	15	2 x 40 SSD	$0.532 per Hour
m3.2xlarge	8	26	30	2 x 80 SSD	$1.064 per Hour
Compute Optimized - Current Generation					
c3.large	2	7	3.75	2 x 16 SSD	$0.188 per Hour
c3.xlarge	4	14	7.5	2 x 40 SSD	$0.376 per Hour
c3.2xlarge	8	28	15	2 x 80 SSD	$0.752 per Hour
c3.4xlarge	16	55	30	2 x 160 SSD	$1.504 per Hour
c3.8xlarge	32	108	60	2 x 320 SSD	$3.008 per Hour

Parallel computing assignment revisited

Computer/instance	Number of threads	Time
Lenovo	1	4.45
Lenovo	2	2.41
Lenovo	8	1.24
t1.micro (free)	1	9.55
m3.xlarge	4	2.76
c3.2xlarge	4	1.48
c3.2xlarge	8	1.17

Stopping vs. terminating instances

- You will be charged if you leave an instance running (or you will use up all of your free tier hours)
- You will not be charged for computing resources if you stop the instance.
- You can re-start any stopped instance.
- You will be charged for storage on any stopped instance (I think ~$0.05 per GB per month)

Stopping vs. terminating instances

- You can terminate an instance, but your AMI will be lost.
- After terminating a node, no charges will occur.

Thoughts about cloud computing

- A real life problem faced by grad students: which laptop to purchase?
- I recently purchased a Lenovo Y510p with an Intel i7-4700MQ processor (3.4 GHz) and 16 GB of RAM
- Cost was ~$1100
- I use my laptops full computing capability <5% of the time.
- The equivalent spec system on the EC2 is about $0.75 an hour.
- My laptop is not future proof
- EC2 is future proof

How I am going to use it

- Amazon's EC2 is portable
- Window Server 2012 is convenient for Windows users
- RStudio on the server facilitates collaboration

Running R from Amazon's Elastic Compute Cloud

Trevor Hefley

Department of Statistics
University of Nebraska–Lincoln

April 30, 2014