

# Running R from Amazon's Elastic Compute Cloud

Trevor Hefley

Department of Statistics  
University of Nebraska–Lincoln

April 30, 2014

- 1 Introduction
- 2 EC2
- 3 Running R on the EC2
  - Pre-made AMI
  - Building a AMI
- 4 Performance
- 5 Important information
- 6 Conclusions

Introduction  
EC2  
Running R on the EC2  
Performance  
Important information  
Conclusions

# Introduction

The screenshot shows the Amazon.com homepage. At the top, there's a navigation bar with the Amazon logo, search bar, and various account options. Below the navigation bar, there's a featured advertisement for Amazon Fire TV, titled "TINY BOX HUGELY ENTERTAINING". The ad shows the Fire TV device and its remote control, with a price of \$99. Below the ad, there's another advertisement for men's boat shoes, titled "THE AMAZON SHOE STORE MEN'S BOAT SHOES".

The main content area is titled "Hamilton Beach Toasters" and displays a grid of six different toaster models. Each model is shown with an image, a brief description, and its price and rating. The ratings are represented by star icons and a number in parentheses.

Hamilton Beach 2-Slice Toaster	Hamilton Beach 2592 Digital 2-Slice	Hamilton Beach 2644 SmartFast	Hamilton Beach 2644 Smart Toast	Hamilton Beach Classic Chrome 2-Slice	Hamilton Beach 4-Slice Toaster
★★★★☆ (102)	★★★★☆ (138)	★★★★☆ (95)	★★★★☆ (85)	★★★★☆ (43)	★★★★☆ (75)
\$69.99 \$23.84	\$69.99 \$27.24	\$69.99 \$22.00	\$69.99 \$31.24	\$69.99 \$24.98	\$69.99 \$39.99






# Introduction

The screenshot shows the AWS website homepage. At the top, there is a navigation bar with the AWS logo, a search bar, and links for "Sign Up", "My Account / Console", and "English". Below the navigation bar, there is a main promotional banner for security software. The banner text reads: "Evaluate top security software for free", "For a limited time, try select security software free for 30 days", and "Learn more at AWS Marketplace". To the right of the text is an illustration of a computer monitor displaying a dashboard and a padlock. Below the banner, there are three offer cards: "Get Started for Free" (Launch virtual machines and apps in minutes), "AWS GLOBAL SUMMITS 2014" (Attend an AWS Summit in a city near you), "AWS FREE TIER" (Gain free, hands-on experience with AWS products for 12 months), and "GET STARTED WITH AWS" (Start using AWS in under 15 minutes). The bottom of the page features a set of navigation icons.

# Introduction

## Amazon Web Services





### Compute & Networking

-  **Direct Connect**  
Dedicated Network Connection to AWS
-  **EC2**  
Virtual Servers in the Cloud
-  **Route 53**  
Scalable Domain Name System
-  **VPC**  
Isolated Cloud Resources
-  **WorkSpaces**  
Desktops in the Cloud







### Storage & Content Delivery

-  **CloudFront**  
Global Content Delivery Network
-  **Glacier**  
Archive Storage in the Cloud
-  **S3**  
Scalable Storage in the Cloud
-  **Storage Gateway**  
Integrates On-Premises IT Environments with Cloud Storage

### Database

-  **DynamoDB**  
Predictable and Scalable NoSQL Data Store
-  **ElastiCache**  
In-Memory Cache
-  **RDS**  
Managed Relational Database Service
-  **Redshift**  
Managed Petabyte-Scale Data Warehouse Service








### Deployment & Management

-  **CloudFormation**  
Templated AWS Resource Creation
-  **CloudTrail**  
User Activity and Change Tracking
-  **CloudWatch**  
Resource and Application Monitoring
-  **Elastic Beanstalk**  
AWS Application Container
-  **IAM**  
Secure AWS Access Control
-  **OpsWorks**  
DevOps Application Management Service

### Analytics

-  **Data Pipeline**  
Orchestration for Data-Driven Workflows
-  **Elastic MapReduce**  
Managed Hadoop Framework
-  **Kinesis**  
Real-time Processing of Streaming Big Data

### App Services

-  **AppStream**  
Low Latency Application Streaming
-  **CloudSearch**  
Managed Search Service
-  **Elastic Transcoder**  
Easy-to-use Scalable Media Transcoding
-  **SES**  
Email Sending Service
-  **SNS**  
Push Notification Service
-  **SQS**  
Message Queue Service
-  **SWF**  
Workflow Service for Coordinating Application Components

## 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/](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

The screenshot shows the AWS Management Console interface for the 'Launch Instance Wizard'. The current step is 'Step 1: Choose an Amazon Machine Image (AMI)'. The page provides instructions on selecting an AMI and lists several options in a table:

AMI Name	AMI ID	Architecture	Root Device Type	Virtualization Type	Size
Amazon Linux AMI 2014.03.1	ami-fb8e292 (64-bit) / ami-178e927e (32-bit)	x86_64	ebs	paravirtual	64-bit @ 32-bit
Red Hat Enterprise Linux 6.4 (PV)	ami-a25415cb (64-bit) / ami-7e175617 (32-bit)	x86_64	ebs	paravirtual	64-bit @ 32-bit
SUSE Linux Enterprise Server 11 sp3 (PV)	ami-e8084881 (64-bit) / ami-b60948df (32-bit)	x86_64	ebs	paravirtual	64-bit @ 32-bit
Ubuntu Server 14.04 LTS (PV)	ami-018c956b (64-bit) / ami-358c956c (32-bit)	x86_64	ebs	paravirtual	64-bit @ 32-bit
Amazon Linux AMI (HVM) 2014.03.1	ami-97d91f1e	x86_64	ebs	hvm	64-bit
Red Hat Enterprise Linux 6.5 (HVM)	ami-63b6910a	x86_64	ebs	hvm	64-bit
SUSE Linux Enterprise Server 11 sp3 (HVM)	ami-e572438c	x86_64	ebs	hvm	64-bit

At the bottom of the page, there is a copyright notice: © 2008 - 2014, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use. A Feedback button is also visible in the bottom right corner.

# Building your own AMI

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



Windows

Free tier eligible

**Microsoft Windows Server 2012 Base** - ami-5f938e36

Microsoft Windows 2012 Standard edition with 64-bit architecture. [English]

Root device type: ebs

Virtualization type: hvm



## 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
<b>General Purpose - Current Generation</b>					
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
<b>Compute Optimized - Current Generation</b>					
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