Work with R on Amazon's Cloud

Alex Zolot (Zolotovitski)

StatVis Consulting

alex1@zolot.us alexzol@microsoft.com www.zolot.us www.statvis.com

Outline

- 1. Amazon's Web Services (AWS):
  - Elastic Compute Cloud (EC2)
  - Simple Storage Service (S3)
  - Elastic MapReduce
  - Amazon AWS console
  - Key Pairs
- 2. Basic instruments to work with remote server:
  - SSH
  - Putty
  - Xmings
- 3. Amazon Machine Images (AMIs)
  - Launching
  - Running
  - Saving
- 4. Transferring R code and data to and from AWS
  - Execution
  - Important R packages

Need:

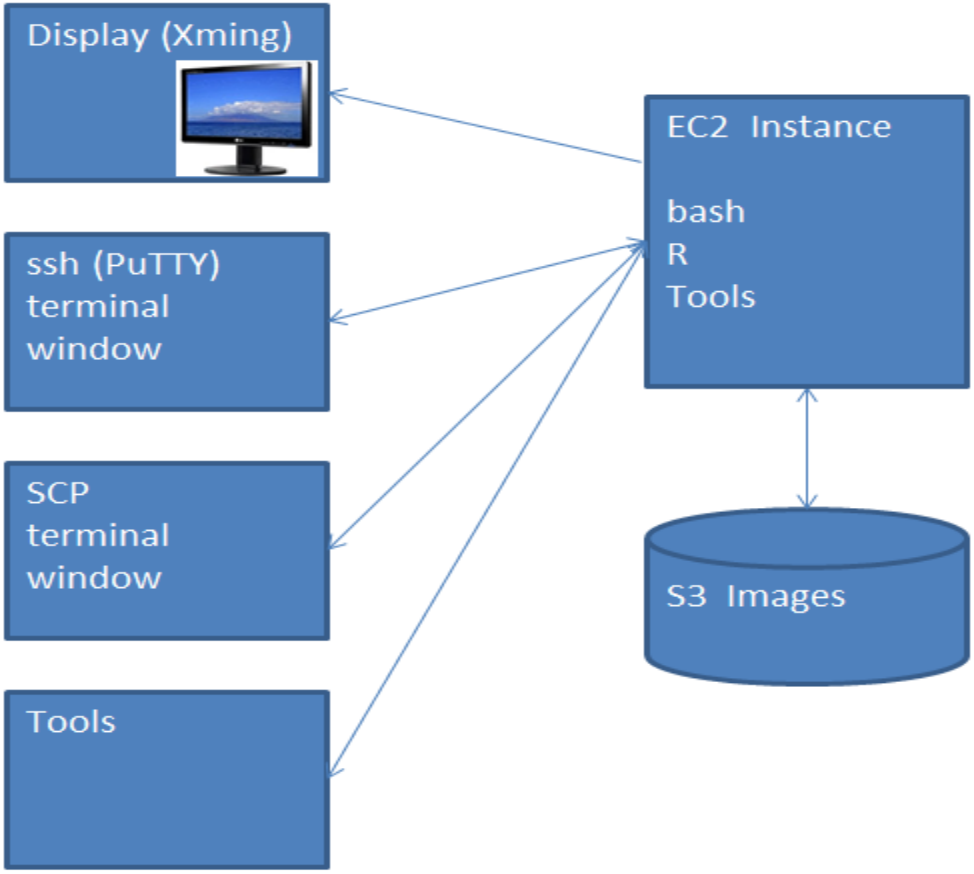
download and install free software:

1) R - any version
2) Mozilla Firefox with plugins: ElasticFox (Mozilla Firefox extension for managing your Amazon EC2 account): http://developer.amazonwebservices.com/connect/entry.jspa?externalID=609,
and Amazon S3 Firefox Organizer(S3Fox): https://addons.mozilla.org/en-US/firefox/addon/3247/.

Register for AWS EC2 and S3 account (http://aws.amazon.com/), get and keep handy your:
Account Number
Access Key ID
Secret Access Key
509 Certificate, and
Amazon EC2 API Tools

MS Windows users :

1. Any IDE for R, e.g. Tinn-R (www.sciviews.org/Tinn-R/), Download: http://sourceforge.net/projects/tinn-r/
2. Xming (www.straightrunning.com/XmingNotes/) , Download: http://sourceforge.net/projects/xming/
3. WinSCP, link and download: http://winscp.net/eng/index.php
4. PuTTY www.chiark.greenend.org.uk/~sgtatham/putty/
5. PuTTYgen: www.chiark.greenend.org.uk/~sgtatham/putty/download.html


Display (Xming)
ssh (PuTTY)
terminal
window
SCP
terminal
window
Tools
EC2 Instance
bash
R
Tools
S3 Images

Why AWS?

- **Simple and convenient to use.** An AMI contains your applications, libraries, data and all associated configuration settings. You simply access it. You don't need to configure it. This applies not only to applications like R, but also can include any third-party data that you require.
- **Available on-demand** over the Internet whenever you need them. You can configure the AMIs yourself without involving the service provider. You don't need to order any hardware and set it up.
- **Elastic access -** you can rapidly provision and access the additional resources you need. Again, no human intervention from the service provider is required. This type of elastic capacity can be used to handle surge requirements when you might need many machines for a short time in order to complete a computation.
- **Pay per use.** The cost of 1 AMI for 100 hours and 100 AMI for 1 hour is the same. With pay per use pricing, which is sometimes called utility pricing, you simply pay for the resources that you use.

Amazon's Web Services (AWS): www.aws.amazon.com
Amazon Elastic Compute Cloud (Amazon EC2)

- Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers.

  Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. AmazonEC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

Prices

Standard On-Demand Instances	**Linux/UNIX Usage**	**Windows Usage**
Small (Default)	$0.085 per hour	$0.12 per hour
Large	$0.34 per hour	$0.48 per hour
Extra Large	$0.68 per hour	$0.96 per hour
High-Memory On-Demand Instances		
Extra Large	$0.50 per hour	$0.62 per hour
Double Extra Large	$1.20 per hour	$1.44 per hour
Quadruple Extra Large	$2.40 per hour	$2.88 per hour
High-CPU On-Demand Instances		
Medium	$0.17 per hour	$0.29 per hour
Extra Large	$0.68 per hour	$1.16 per hour

Amazon Elastic MapReduce

Amazon Elastic MapReduce is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data. It utilizes a hosted Hadoop framework running on the web-scale infrastructure of Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Simple Storage Service (Amazon S3).

Using Amazon Elastic MapReduce, you can instantly provision as much or as little capacity as you like to perform data-intensive tasks for applications such as web indexing, data mining, log file analysis, machine learning, financial analysis, scientific simulation, and bioinformatics research. Amazon Elastic MapReduce lets you focus on crunching or analyzing your data without having to worry about time-consuming set-up, management or tuning of Hadoop clusters or the compute capacity upon which they sit.

Amazon Simple Storage Service (Amazon S3)

Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers.

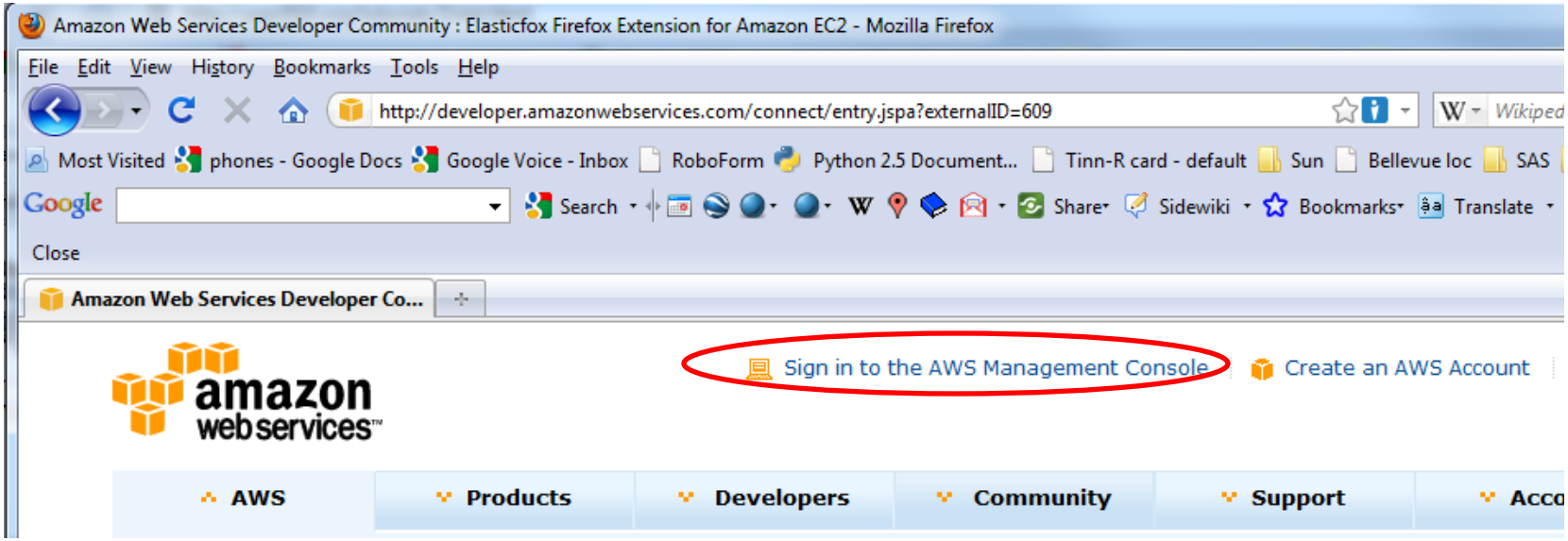
Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, secure, fast, inexpensive infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

Amazon S3, prices

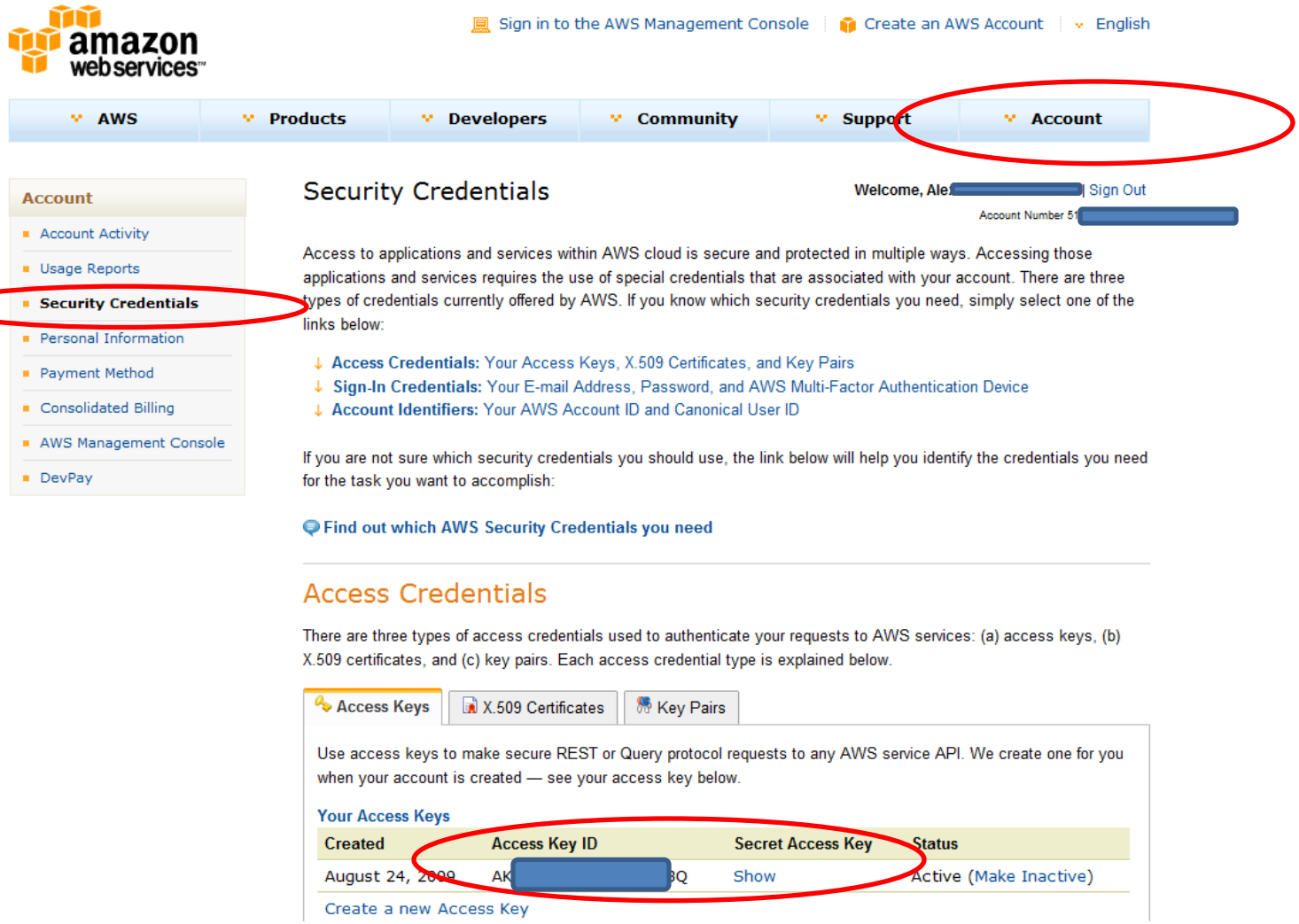
Storage (Designed for 99.999999999% Durability)		Reduced Redundancy Storage (Designed for 99.99% Durability)		Data Transfer*		Requests	
Tier	Pricing	Tier	Pricing	Tier	Pricing	Type	Pricing
First 50 TB / Month of Storage Used	$0.150 per GB	First 50 TB / Month of Storage Used	$0.100 per GB	All Data Transfer In	Free until June 30th, 2010**	PUT, COPY, POST, or LIST	$0.01 per 1,000 Requests
Next 50 TB / Month of Storage Used	$0.140 per GB	Next 50 TB / Month of Storage Used	$0.093 per GB	First 1 GB / month data transfer out	$0.000 per GB	GET and All Other Requests***	$0.01 per 10,000 Requests
Next 400 TB / Month of Storage Used	$0.130 per GB	Next 400 TB / Month of Storage Used	$0.087 per GB	Up to 10 TB / month data transfer out	$0.150 per GB		
Next 500 TB / Month of Storage Used	$0.105 per GB	Next 500 TB / Month of Storage Used	$0.070 per GB	Next 40 TB / month data transfer out	$0.110 per GB		
Next 4000 TB / Month of Storage Used	$0.080 per GB	Next 4000 TB / Month of Storage Used	$0.053 per GB	Next 100 TB / month data transfer out	$0.090 per GB		
Storage Used / Month Over 5000 TB	$0.055 per GB	Storage Used / Month Over 5000 TB	$0.037 per GB	Greater than 150 TB / month data transfer out	$0.080 per GB		

Setting up.

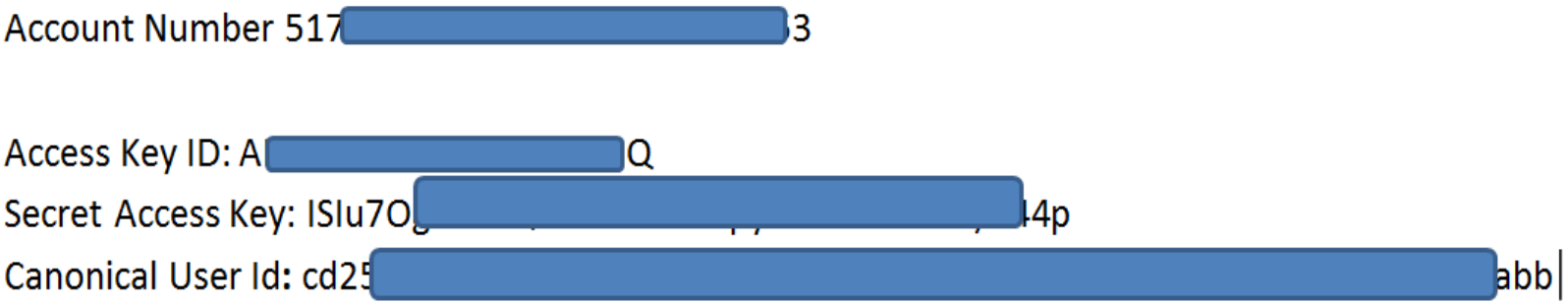
2) Firefox

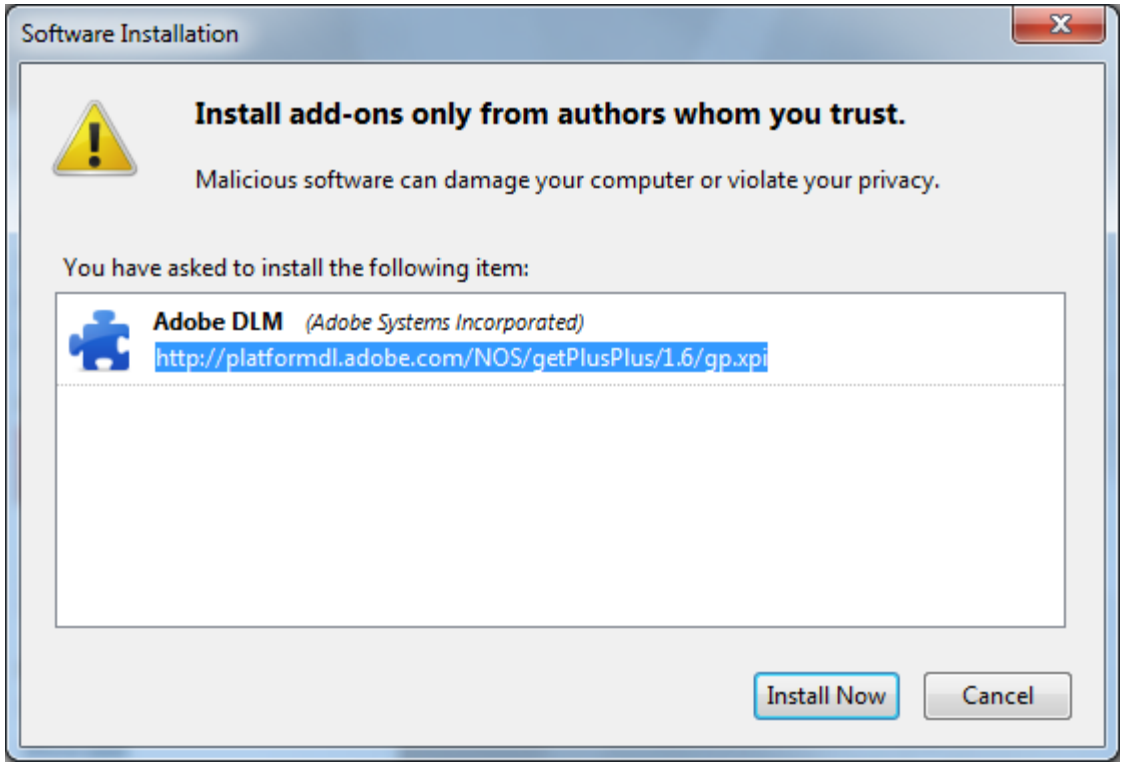


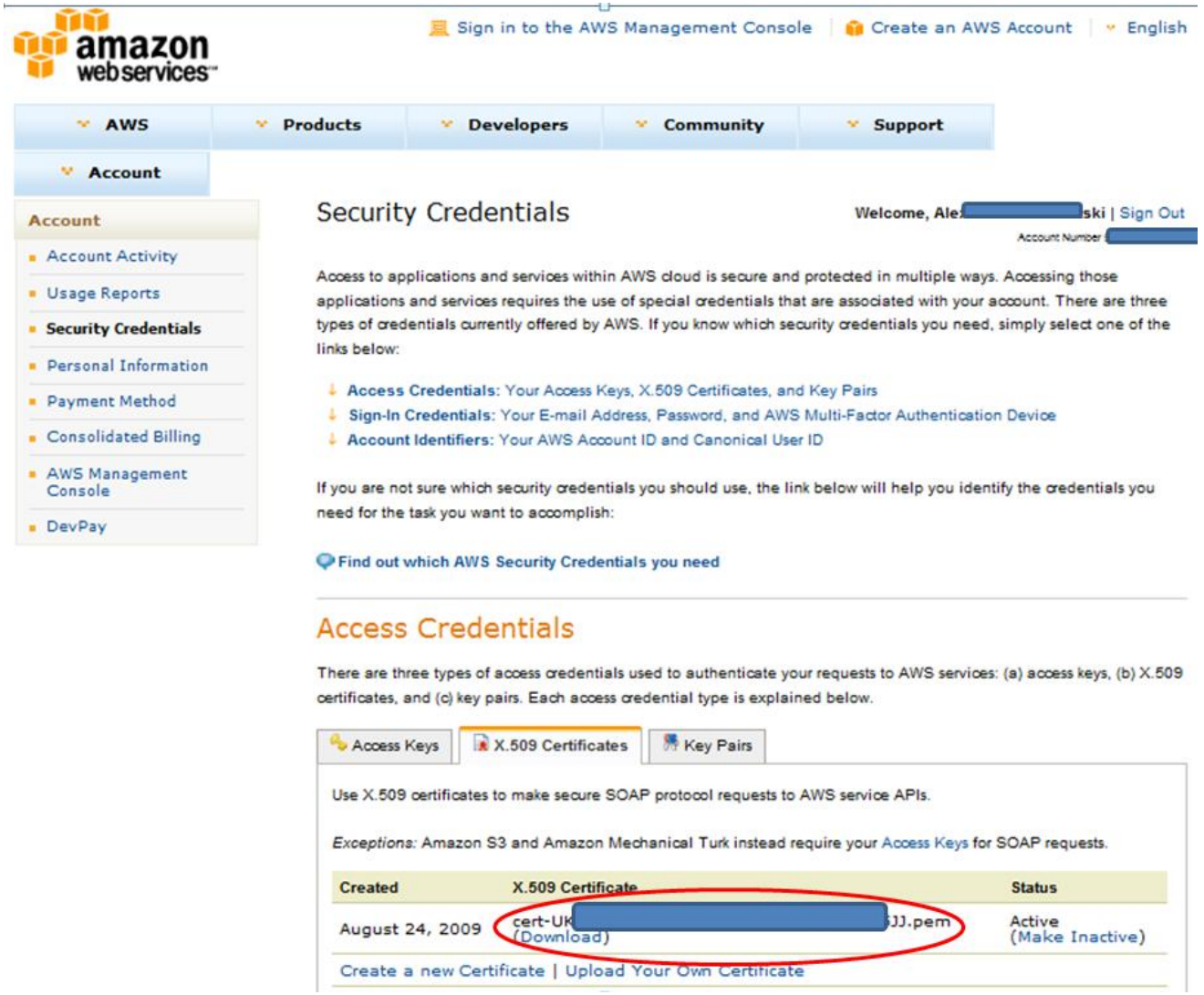
Setting up. Get Security Credentials



Keep credentials:

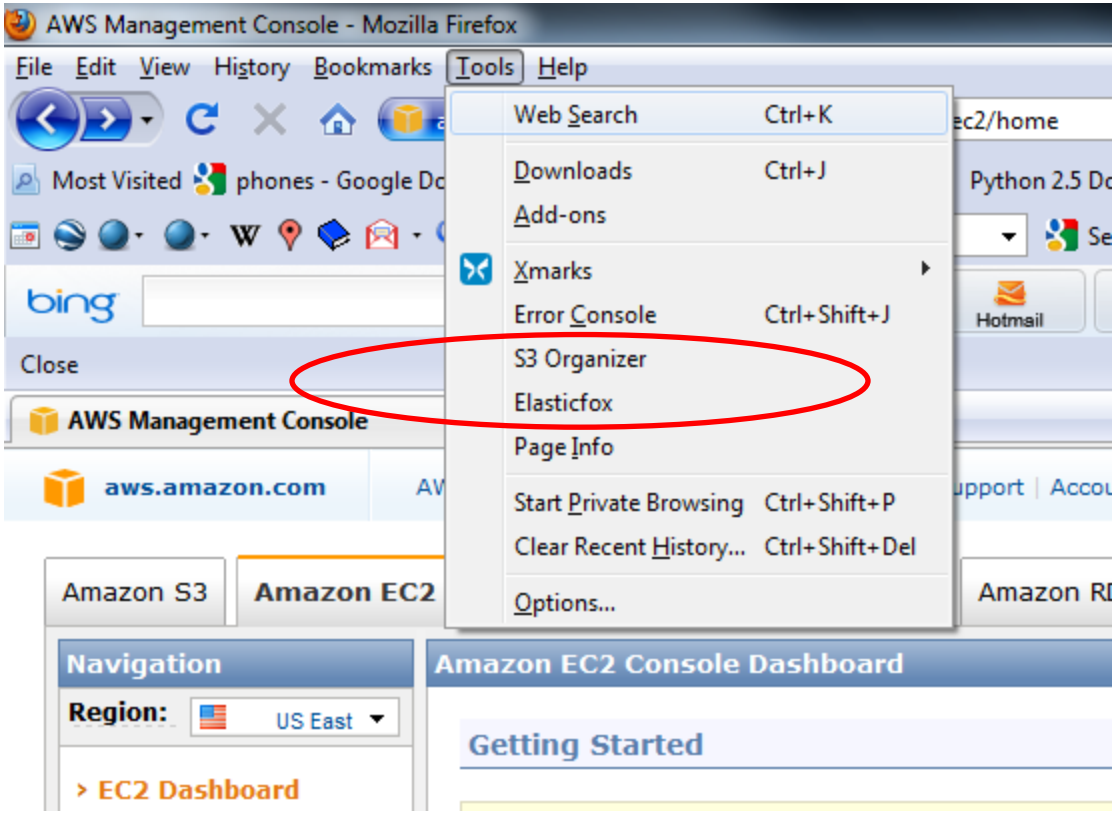



Software Installation
Install add-ons only from authors whom you trust.
Malicious software can damage your computer or violate your privacy.
You have asked to install the following item:
Adobe DLM (Adobe Systems Incorporated)
http://platformdl.adobe.com/NOS/getPlusPlus/1.6/gp.xpi
Install Now
Cancel

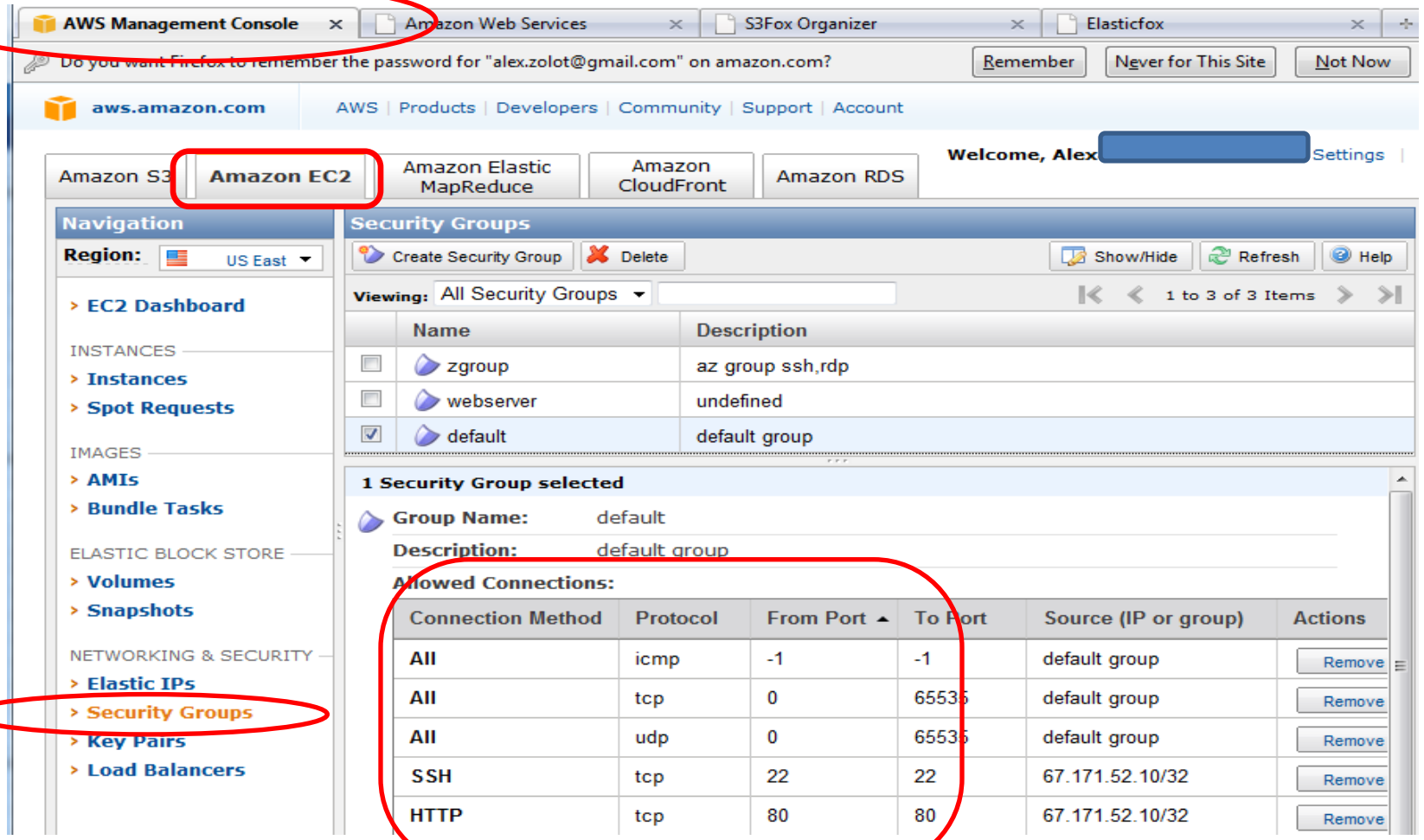

amazon web services
Sign in to the AWS Management Console
Create an AWS Account
English
AWS
Products
Developers
Community
Support
Account
Account
Account Activity
Usage Reports
Security Credentials
Personal Information
Payment Method
Consolidated Billing
AWS Management Console
DevPay
Security Credentials
Welcome, Ale[redacted]ski | Sign Out
Account Number [redacted]
Access to applications and services within AWS cloud is secure and protected in multiple ways. Accessing those applications and services requires the use of special credentials that are associated with your account. There are three types of credentials currently offered by AWS. If you know which security credentials you need, simply select one of the links below:
Access Credentials: Your Access Keys, X.509 Certificates, and Key Pairs
Sign-In Credentials: Your E-mail Address, Password, and AWS Multi-Factor Authentication Device
Account Identifiers: Your AWS Account ID and Canonical User ID
If you are not sure which security credentials you should use, the link below will help you identify the credentials you need for the task you want to accomplish:
Find out which AWS Security Credentials you need
Access Credentials
There are three types of access credentials used to authenticate your requests to AWS services: (a) access keys, (b) X.509 certificates, and (c) key pairs. Each access credential type is explained below.
Access Keys
X.509 Certificates
Key Pairs
Use X.509 certificates to make secure SOAP protocol requests to AWS service APIs.
Exceptions: Amazon S3 and Amazon Mechanical Turk instead require your Access Keys for SOAP requests.
Created
X.509 Certificate
Status
August 24, 2009
cert-UK[redacted]JJ.pem
(Download)
Active
(Make Inactive)
Create a new Certificate | Upload Your Own Certificate

Setting up.

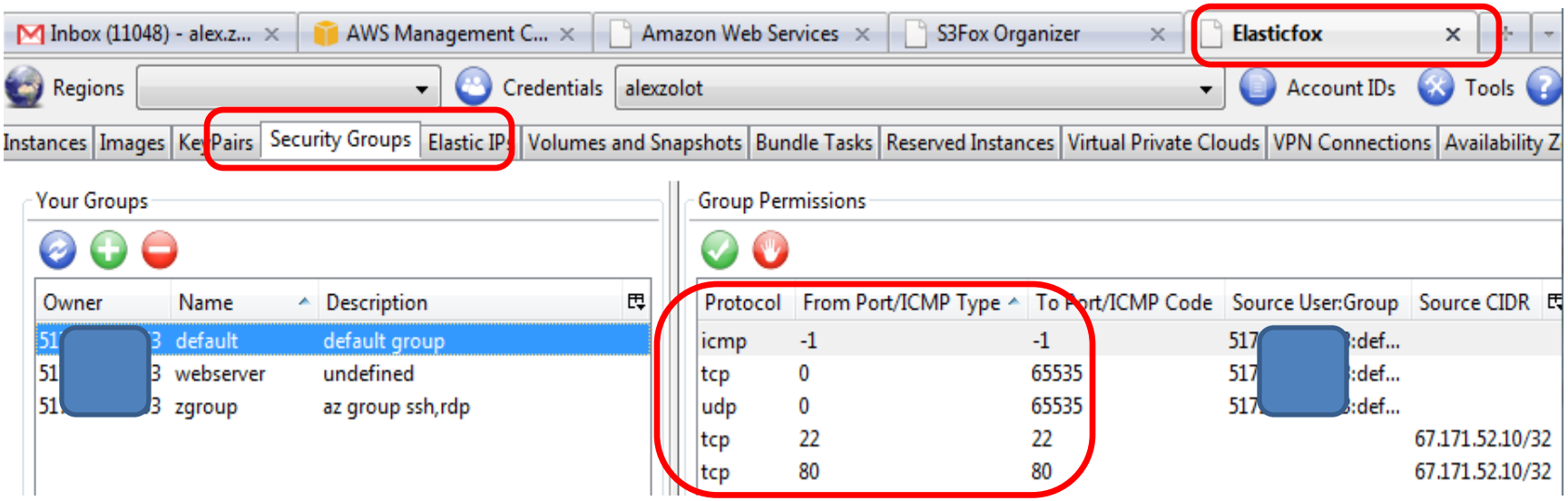
3) Firefox Plugins



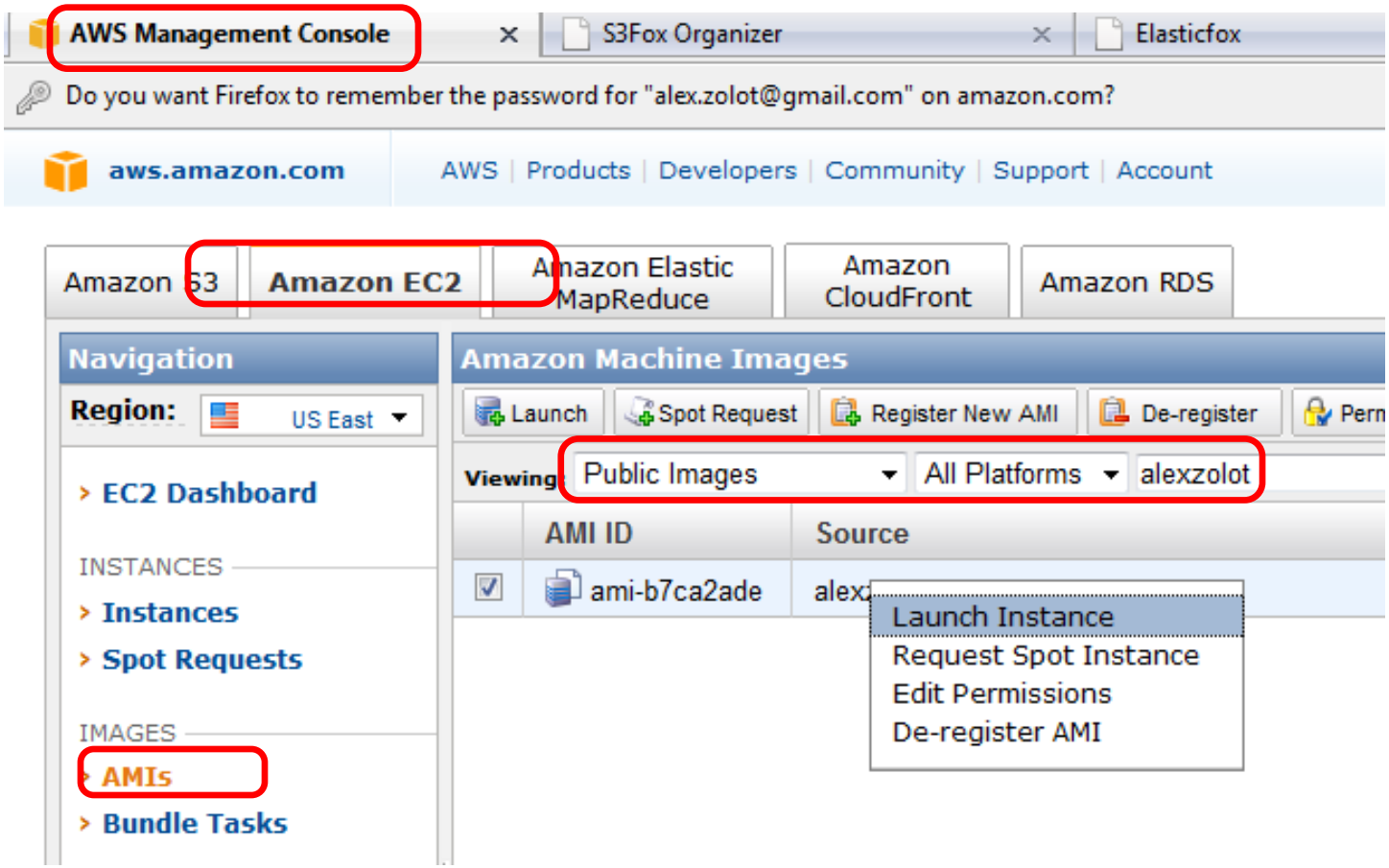
Set security group (in AWS Mgm.Cons)



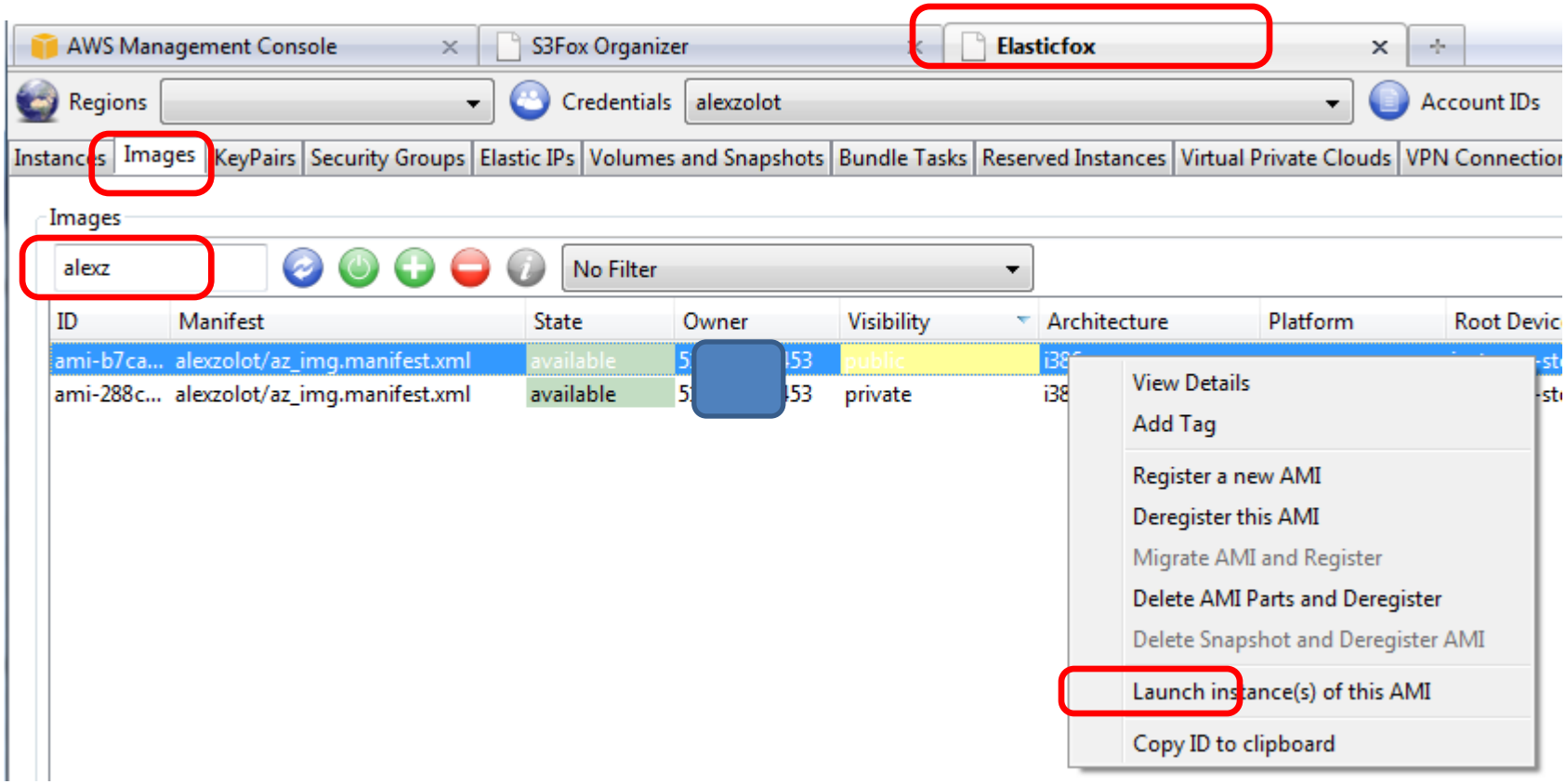
Set security group (In Elasticfox FF plugin)



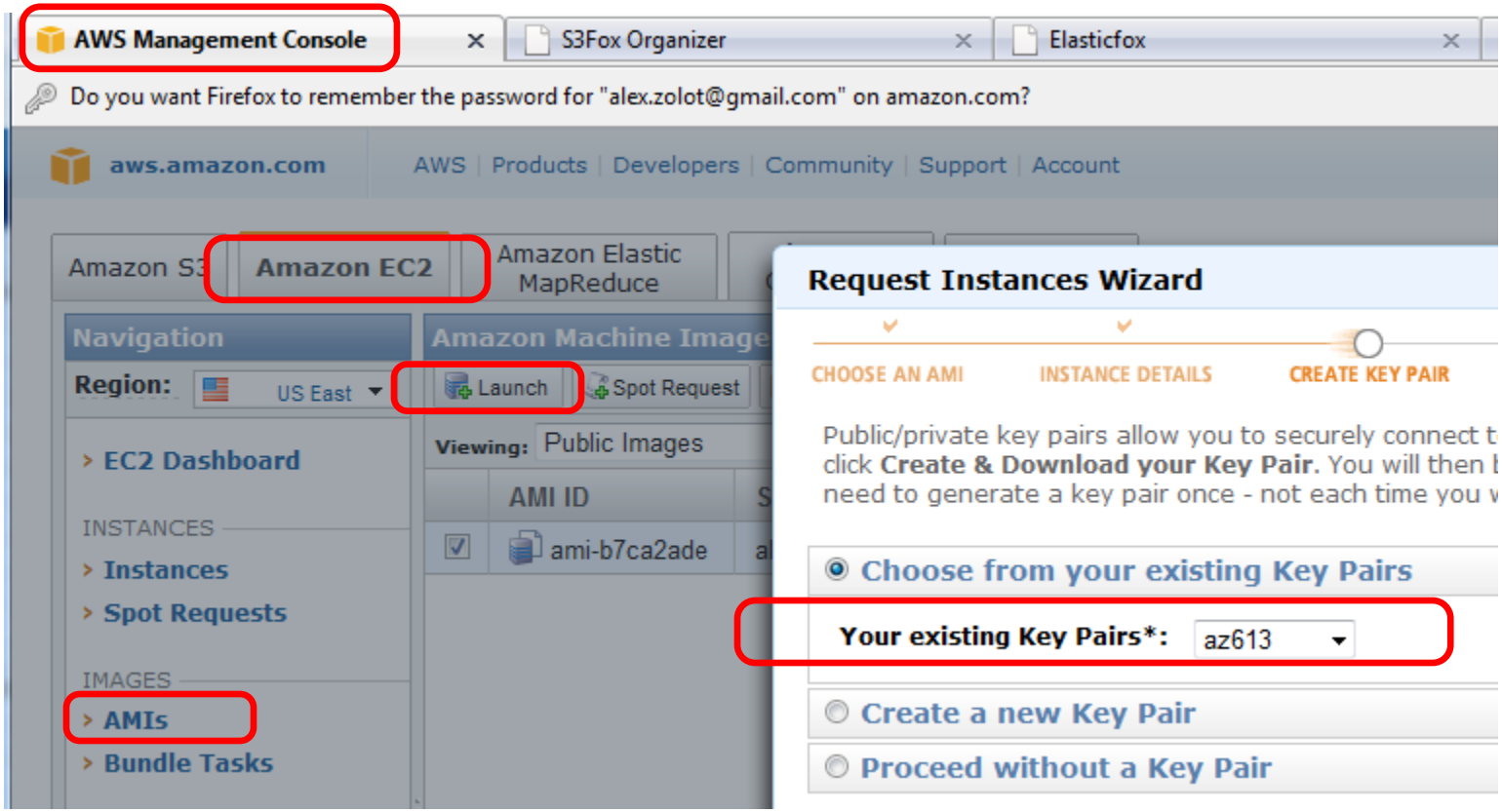
Choose AMI - Launch Instance (in AWS Mgm.Cons)



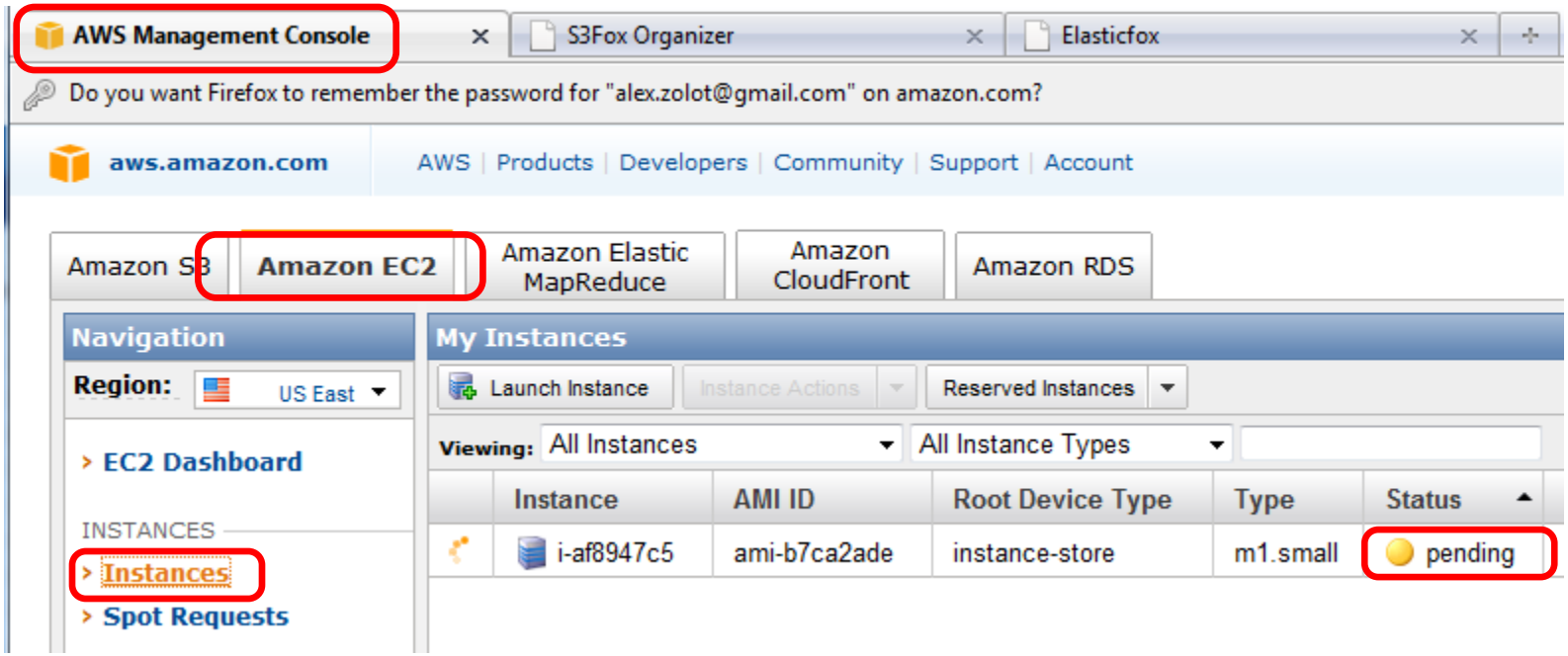
Choose AMI - Launch Instance (In Elasticfox FF plugin)



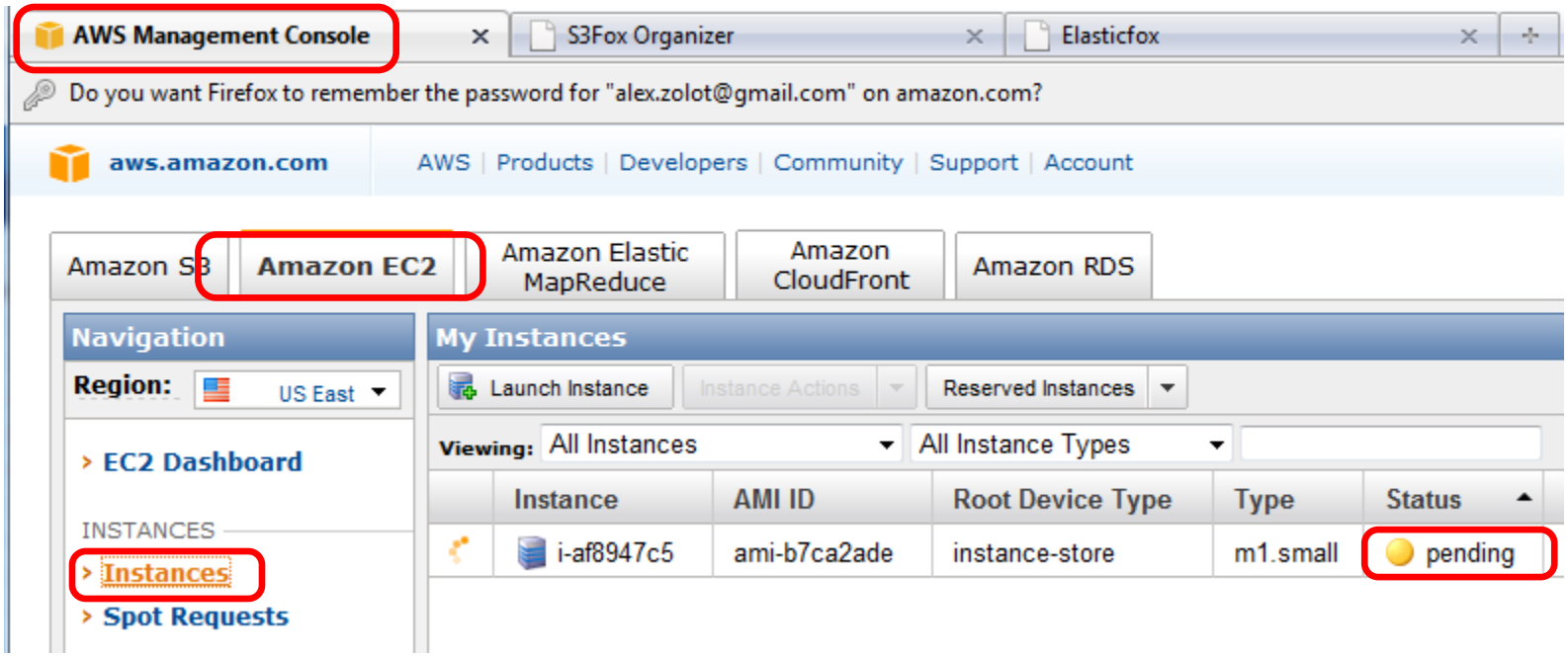
Choose AMI - Launch Instance (in AWS Mgm.Cons)



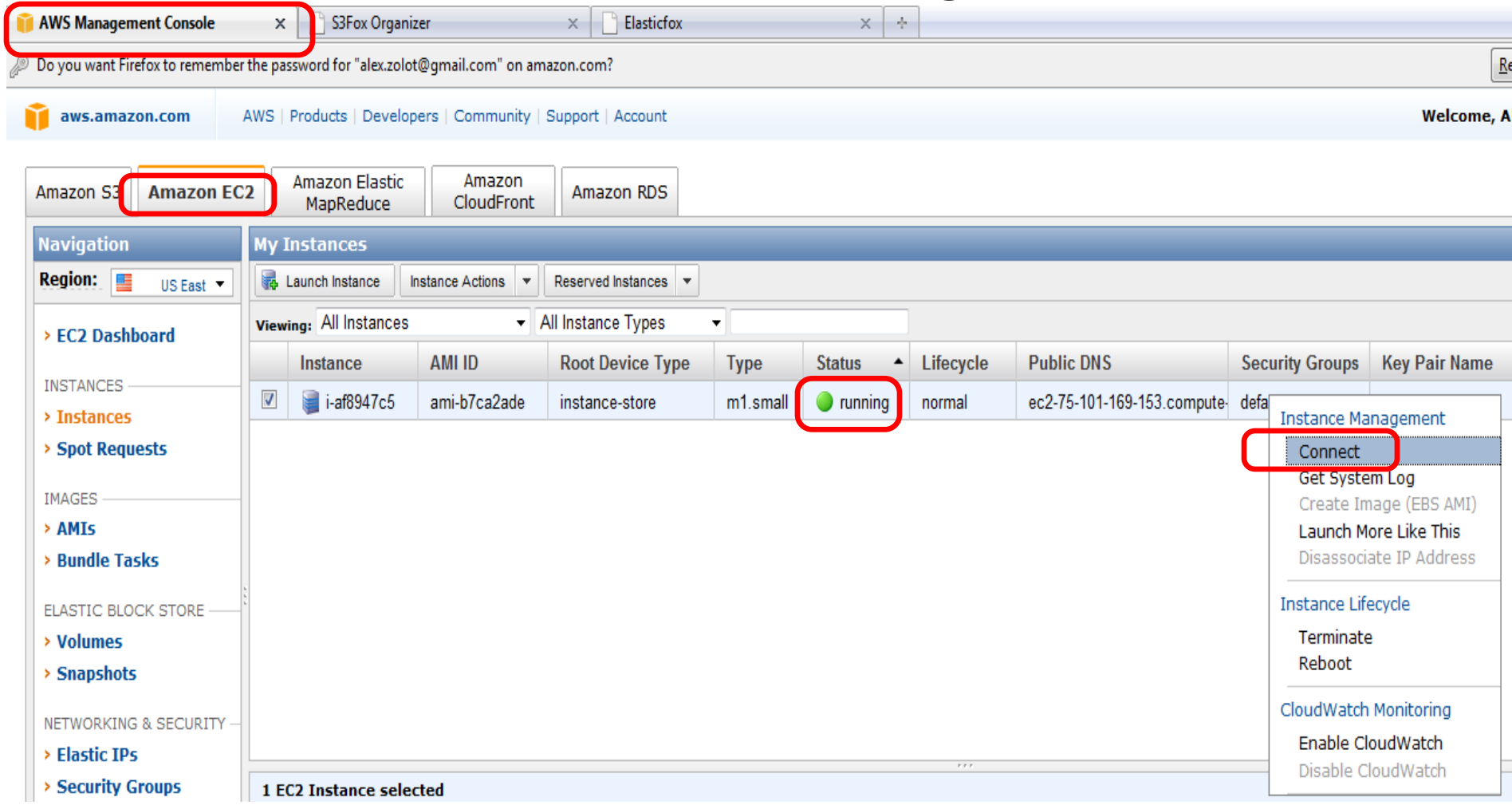
Choose AMI - Launch Instance (in AWS Mgm.Cons)



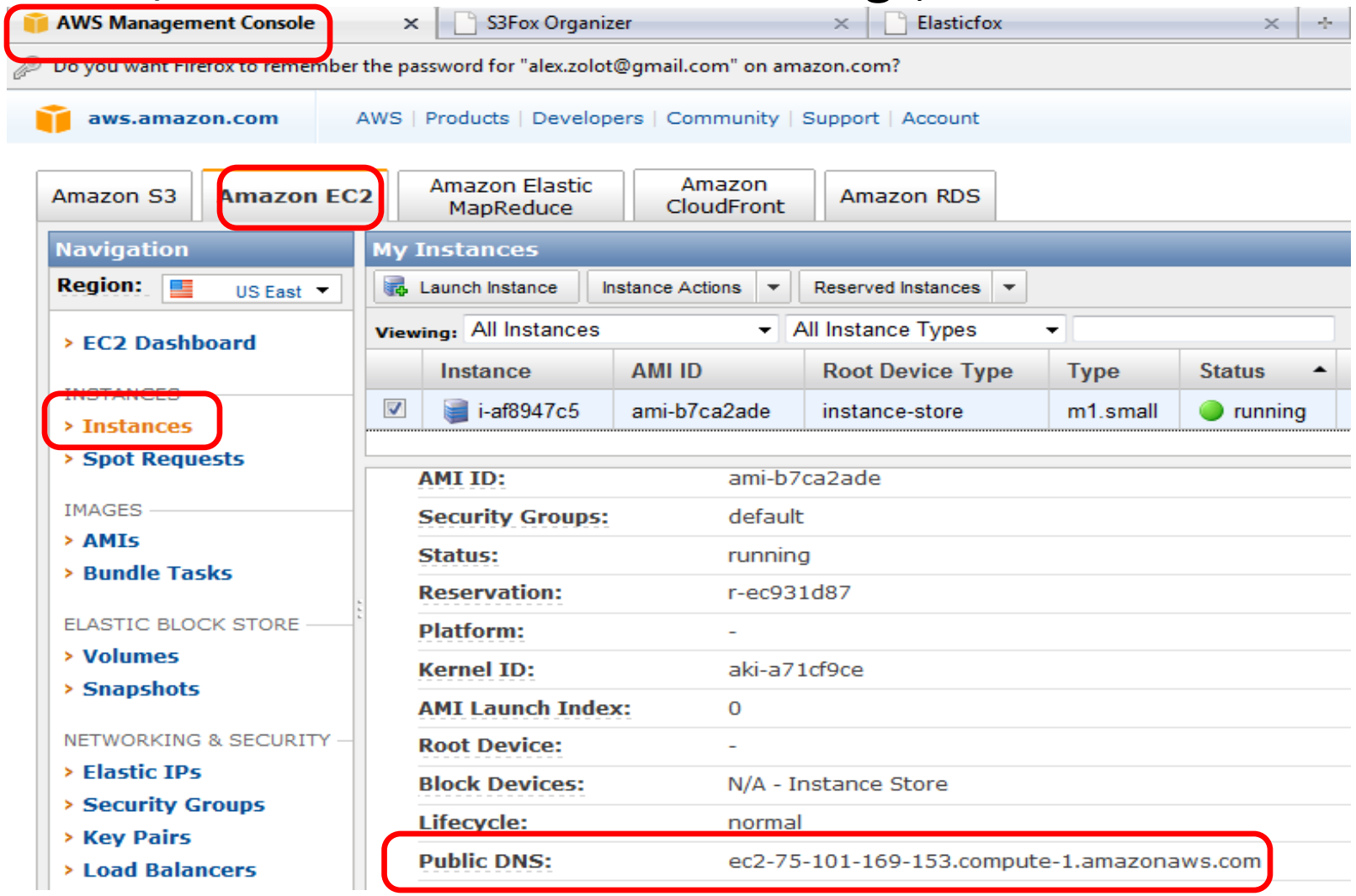
Choose AMI - Launch Instance (in AWS Mgm.Cons)



Choose AMI - Launch Instance (in AWS Mgm.Cons) wait/refresh till Status= “running”, then connect



Choose AMI - Launch Instance (in AWS Mgm.Cons) wait/refresh till Status= “running”, then connect

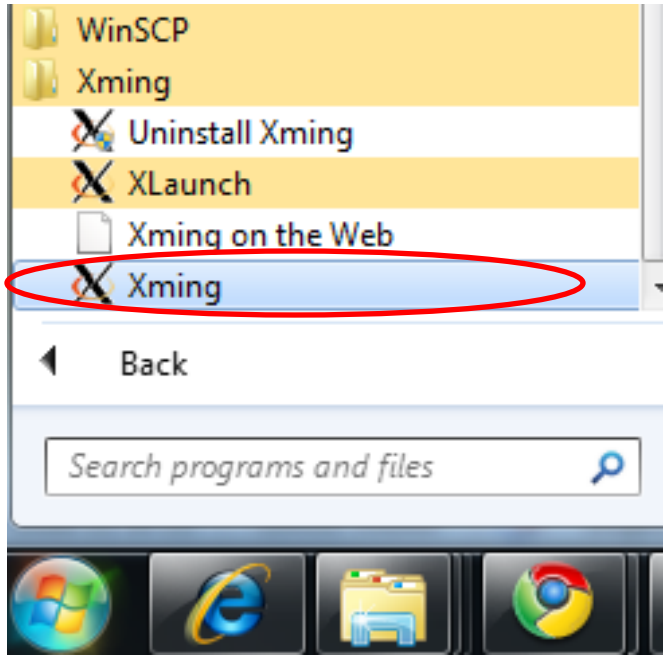


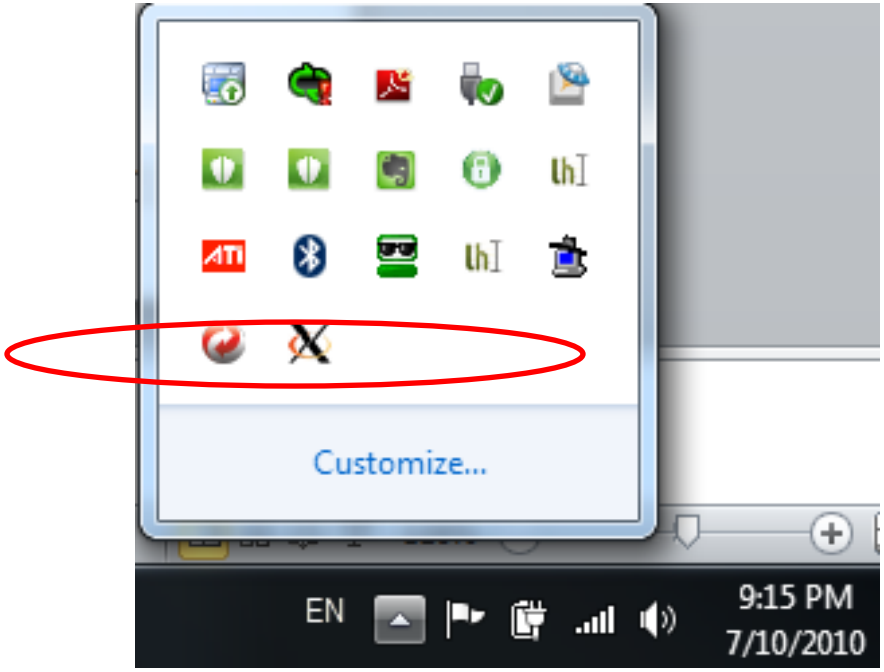
Choose AMI - Launch Instance (in AWS Mgm.Cons) wait/refresh till Status= “running”, then connect

Public DNS: ec2-75-101-169-153.compute-1.amazonaws.com

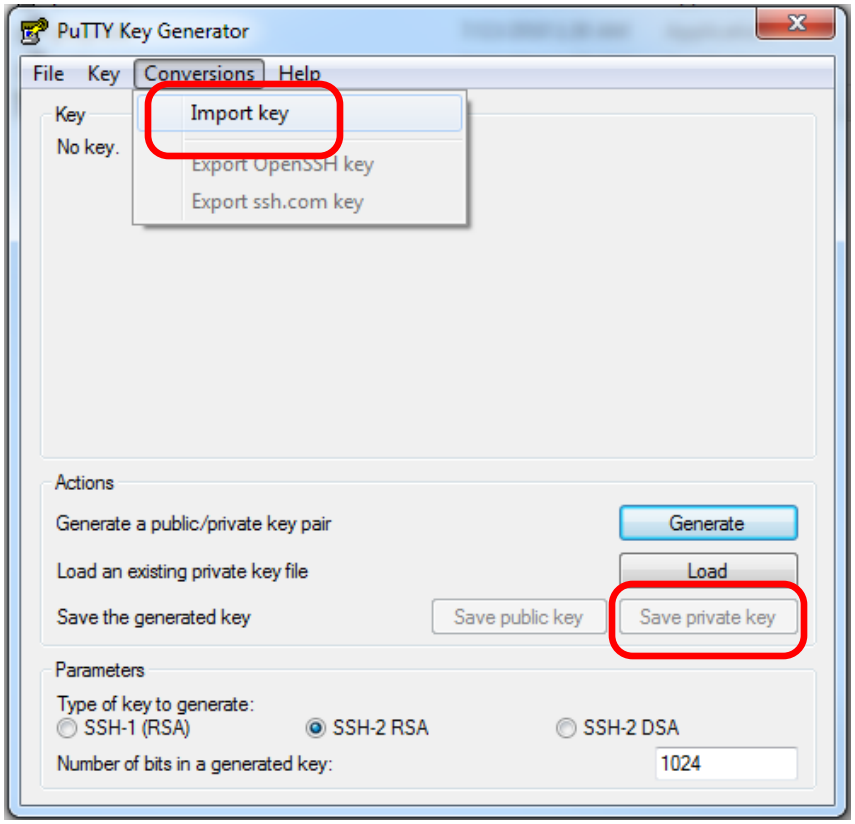
Setting up.

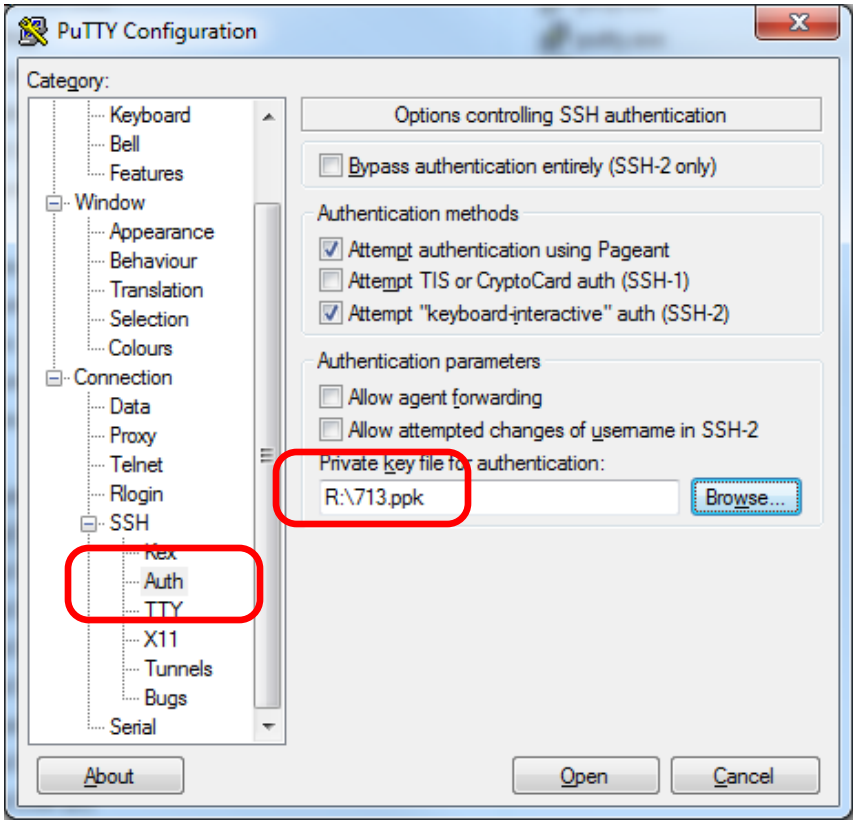
1) Xming



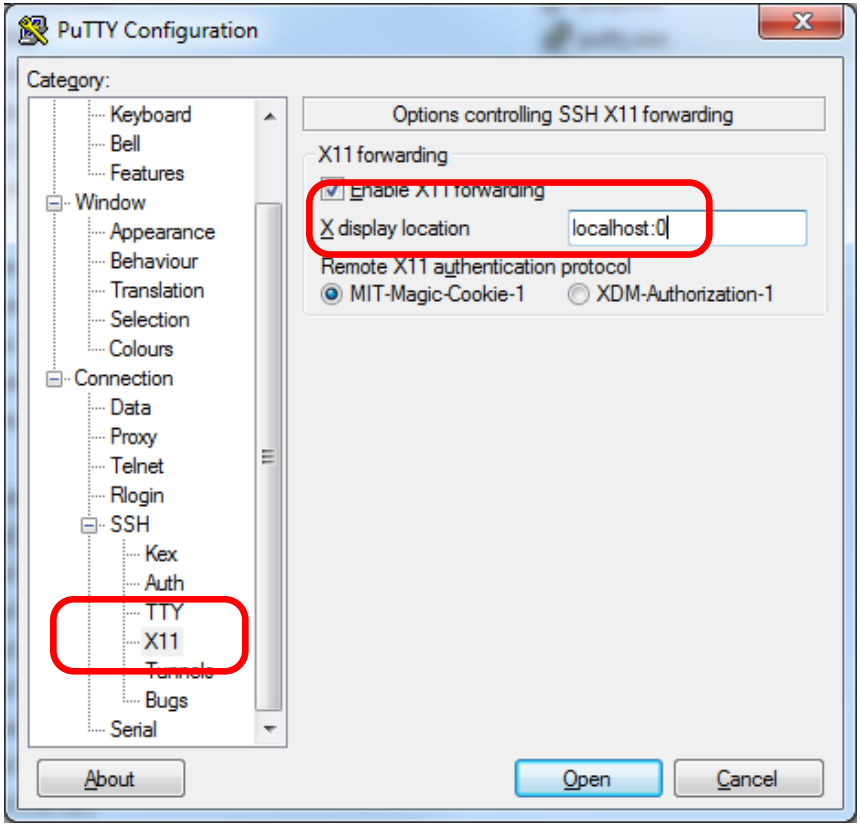


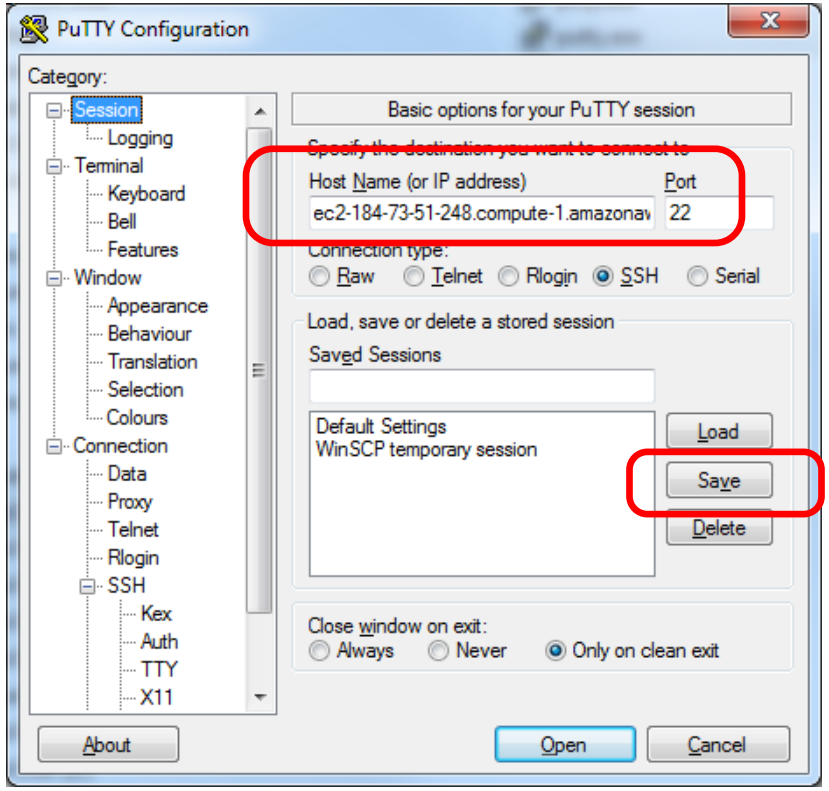
PuTTY 1.Convert .pem -> .ppk





PuTTY 2. Connection





Try R

```
# Example 1. (Xming)
plot(1:4)
plot(5:1)
x11()
plot(1:3)
```

Try R

WinSCP

Example 2. www – pdf()

```
setwd('/var/www/html')
pdf('z.pdf')
plot(1:4)
plot(5:1)
plot(1:3)
dev.off()
```

Try R

WinSCP

Example 3. www – R2HTML()

```
install.packages("R2HTML", repos = "http://cran.stat.ucla.edu/")
library(R2HTML)
fout= HTMLInitFile(outdir = "/var/www/",
        filename="z",
        CSSFile="http://www.stat.ucl.ac.be/R2HTML/Pastel.css")
data(iris)
HTML(as.title("Fisher Iris dataset / Correlation 1"),file=fout)
HTML(cor(iris[,1:4]), file=fout)
HTML(as.title("Fisher Iris dataset / Correlation 2"),file=fout)
HTML.cormat(cor(iris[,1:4]), file=fout)
# File is generated, you can call the browser:
## Not run: browseURL(tmpfic)
```

R. Send data files from local PC to EC2 Instance and back.

WinSCP (pscp.exe)

Create your AMI

1. Download tools to the running Instance. Set Envir. Variables.
2. Upload private key and certificate to /mnt/c
3. Create bundle at S3.
4. Register AMI.

- **1. Getting the Command Line Tools**

- The command line tools: http://s3.amazonaws.com/ec2-downloads/ec2-ami-tools.zip These tools are written in Java and include shell scripts for both Windows 2000/XP and Linux/Unix/Mac OSX. The ZIP file is self-contained; no installation is required.
- 1b. Set Envir. Variables.

- C:\> set EC2_HOME=<*path-to-tools*>
- S:\51_AWS_cloud\ec2-api-tools-1.3-51254
- C:\> set PATH=%PATH%;%EC2_HOME%\bin
-
- set JAVA_HOME=C:\Windows\System32
- set EC2_HOME=S:\51_AWS_cloud\ec2-api-tools-1.3-51254
- set PATH=%PATH%;%EC2_HOME%\bin
-
- C:\> set **EC2_PRIVATE_KEY=***c:\ec2\pk-HKZYKTAIG2ECMXYIBH3HXV4ZBZQ55CLO.pem*
- C:\>

To take a snap shot of the running server. While still on the server run the following command:
ec2-bundle-vol -d /mnt -k /mnt/pk-STRINGOFTHIRTYTWOLETTERSORDIGIT

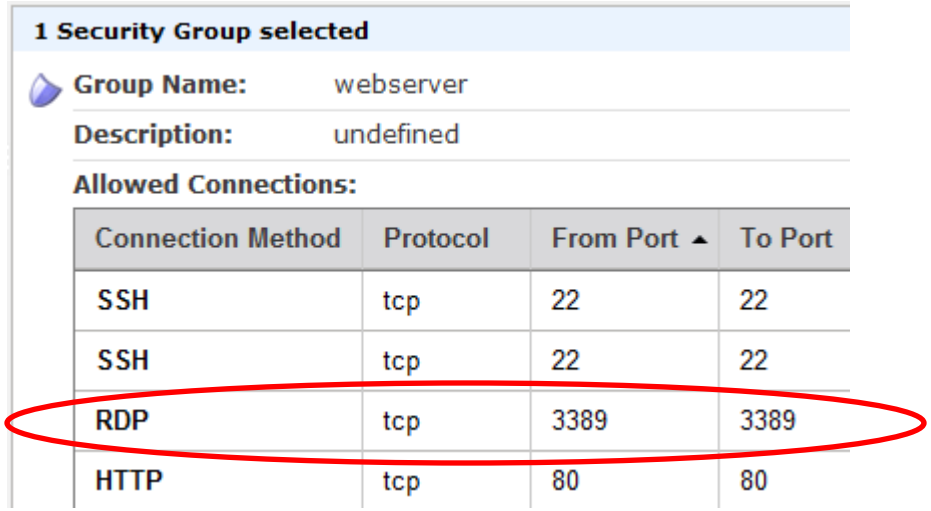
Installation R using AMI w/o R.

1. Run Instance from AMI w/o R.
2. Install R from CRAN following instructions. – binary or from source

E.g. take debian xxx - see R-file

From source:

Win AMI



1 Security Group selected

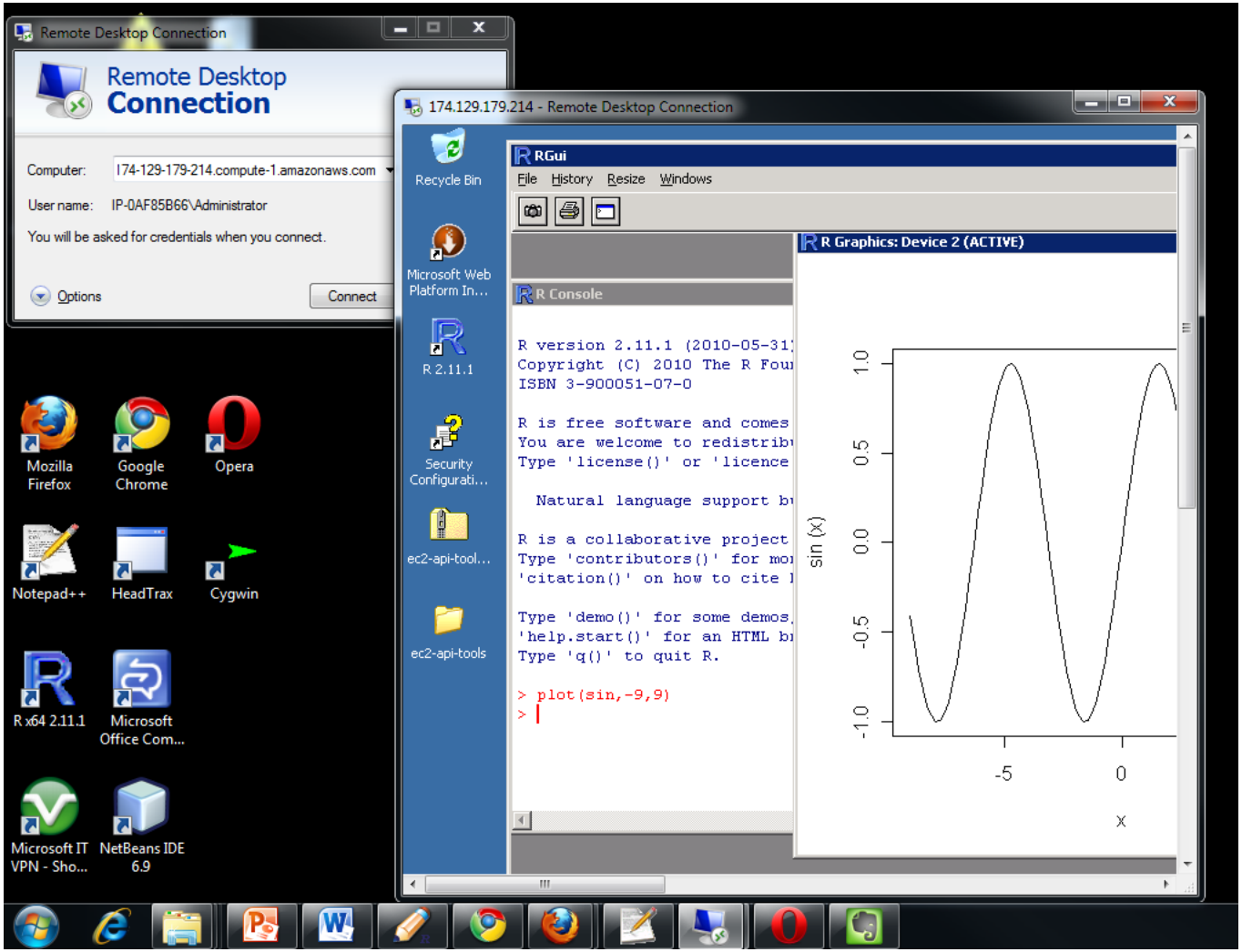
Group Name: webserver

Description: undefined

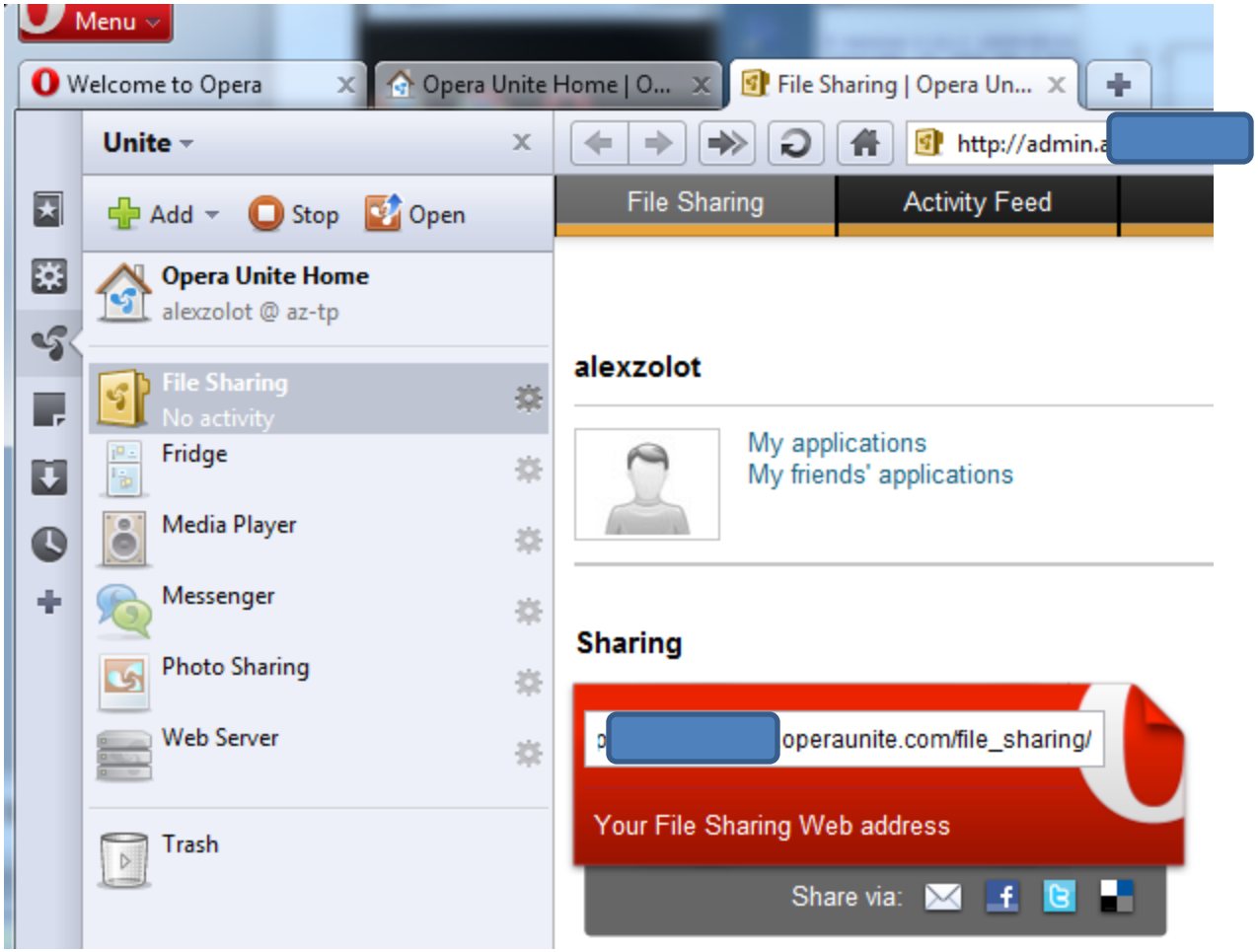
Allowed Connections:

Connection Method	Protocol	From Port ▲	To Port
SSH	tcp	22	22
SSH	tcp	22	22
RDP	tcp	3389	3389
HTTP	tcp	80	80

Win AMI


Remote Desktop Connection
Remote Desktop Connection
Computer: 174-129-179-214.compute-1.amazonaws.com
User name: IP-0AF85B66\Administrator
You will be asked for credentials when you connect.
Options
Connect
174.129.179.214 - Remote Desktop Connection
Recycle Bin
Microsoft Web Platform In...
R 2.11.1
Security Configurati...
ec2-api-tool...
ec2-api-tools
RGui
File History Resize Windows
R Console
R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Fou
ISBN 3-900051-07-0
R is free software and comes
You are welcome to redistrib
Type 'license()' or 'licence
Natural language support b
R is a collaborative project
Type 'contributors()' for mo
'citation()' on how to cite
Type 'demo()' for some demos,
'help.start()' for an HTML b
Type 'q()' to quit R.
> plot(sin,-9,9)
>
R Graphics: Device 2 (ACTIVE)
sin (x)
1.0
0.5
0.0
-0.5
-1.0
-5
0
x
Mozilla Firefox
Google Chrome
Opera
Notepad++
HeadTrax
Cygwin
R x64 2.11.1
Microsoft Office Com...
Microsoft IT VPN - Sho...
NetBeans IDE 6.9

Win AMI. WinSCP -> Opera Unite



References

Work with R on Amazon's Cloud. http://user2010.org/tutorials/Zolot.html

1. Robert Grossman, Running R on Amazon's EC2 http://rgrossman.com/2009/05/17/running-r-on-amazons-ec2
2. Amazon Web Services (AWS) www.aws.amazon.com.
3. A quick overview of PuTTY and SSH for AWS Newbies http://clouddb.info/2009/05/17/a-quick-overview-of-putty-and-ssh-for-aws-newbies
4. Using Xming for X11 applications on an EC2 machine http://www.full360.com/content/using-xming-x11-applications-ec2-machine
5. Command Line Tools Amazon EC2 Resource Center
6. Deploying a Web Server on Amazon EC2 www.ondemandbeat.com/2009/04/30/deploying-a-web-server-on-amazon-ec2/
7. SAVING A CUSTOMISED LINUX AMAZON INSTANCE (EC2 AND S3) http://robrohan.com/2009/01/30/saving-a-customised-linux-amazon-instance-ec2-and-s3/
8. How to Run Windows & IIS in the Cloud on Amazon EC2 (in 15 mins) http://blogs.iis.net/bills/archive/2009/01/13/how-to-run-windows-amp-iis-in-the-cloud-on-amazon-ec2-in-15-mins.aspx
9. www.robertsosinski.com/2008/01/26/starting-amazon-ec2-with-mac-os-x