Amazon Web Services

Data at Scale

Why

- We should learn about AWS

  - Example for Cloud Computing

  - Reasonably cheap

    - Most covered in the free tier

How

- You need to use a credit / debit card

  - But most activities are covered by the free tier

  - Set up billing alerts (Use US North Virginia!)

  - You can work in groups

What

- AWS is a confusing conglomerates of different services

- Most prominent

  - EC2 — virtual machines

  - S3 — Storage capacity

- We will also use a Hadoop cluster to do exercises with Pig

  - **Those can cost a bit of money if you forget to take down your cluster when you are done**

  - If you make a mistake, the costs are less than 10$ per day

Use cases for AWS

- Machine learning

    - A specialized machine with large GPU costs several thousands dollar

    - If you don't do a lot of neural network computation, AWS is cheaper

- Setting up and hosting a web server

- Running a Hadoop cluster without the hazzle

- Running Apps in a private network

- Running highly available systems

- Lowering costs for compute infrastructures

- …

Alternatives to AWS

- Azure Virtual Machines

- Google Compute Engines & Google Cloud Storage

Billing

- AWS (and other cloud service providers) bill:

  - Per second

  - Per GB stores

  - Per GB moved

Interacting with AWS

- By hand:

  - Use the AWS management console

- Using a command-line interface

  - Allows to automize the renting of AWS resources

- SDK

  - Programmatic access

  - Available in various languages

- Blueprints

  - Description of your system

  - Compares with current state and figures out how to get additional resources

Signing up with AWS

- Provide login credentials (including strong password)

- Provide contact information

  - Your phone number is checked

- Provide payment details

- Verification of identity via phone

- Support plan

Activity

- Use your web-browser to access <u>aws.amazon.com</u>

- *Create a Free Account*

- Use a strong password (at least 20 characters)

Activity

- Create accurate contact information

Activity

- Create payment information

Activity

- Verify Identity

  - Message through smart-phone …

Activity

- Choose your support plan

  - **BASIC**

- unless you want to pay for something you'll probably do not need

Activity

- Now you can sign on to AWS

  - Go to the AWS management console
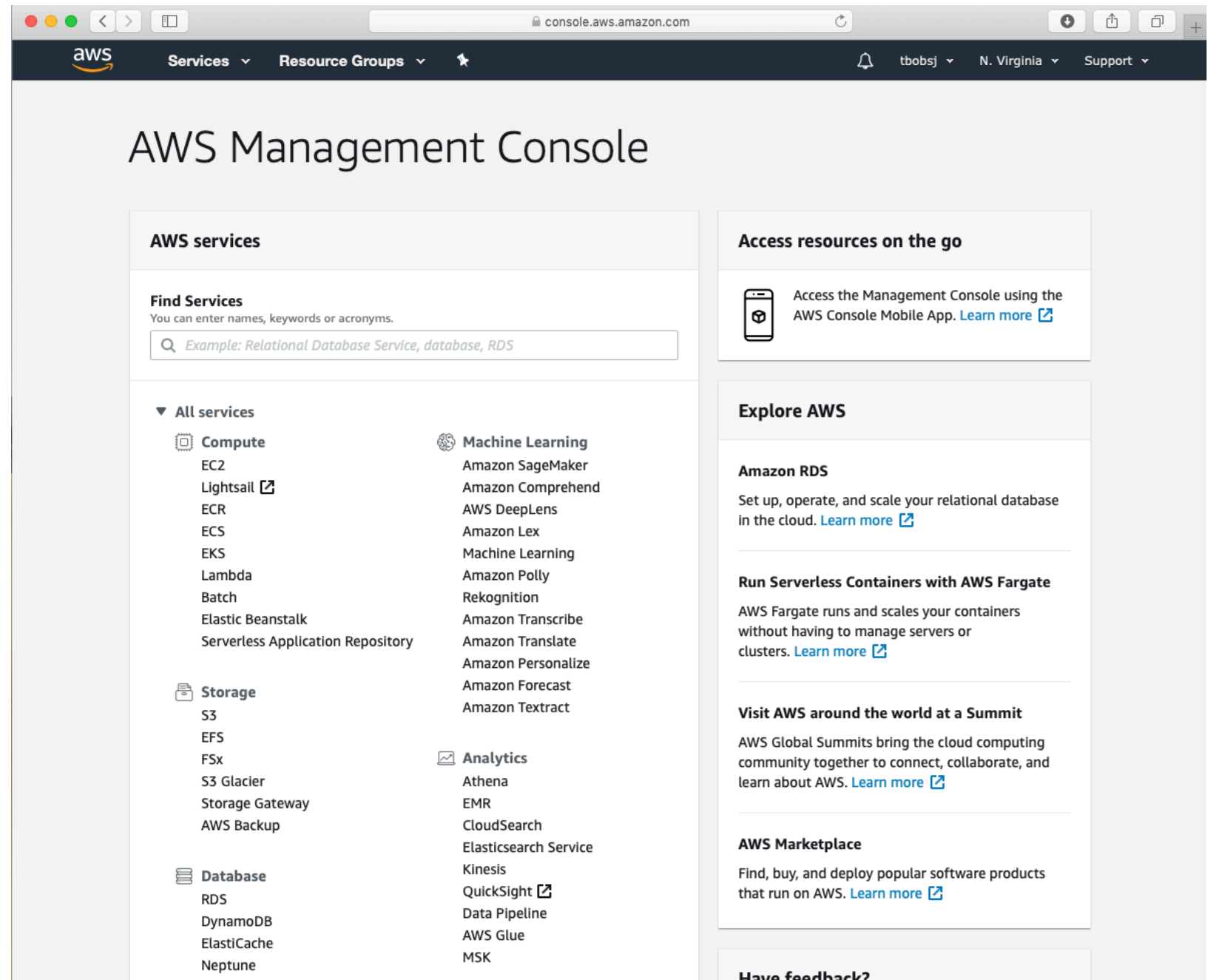
Activity

- Select the N.Virginia service center
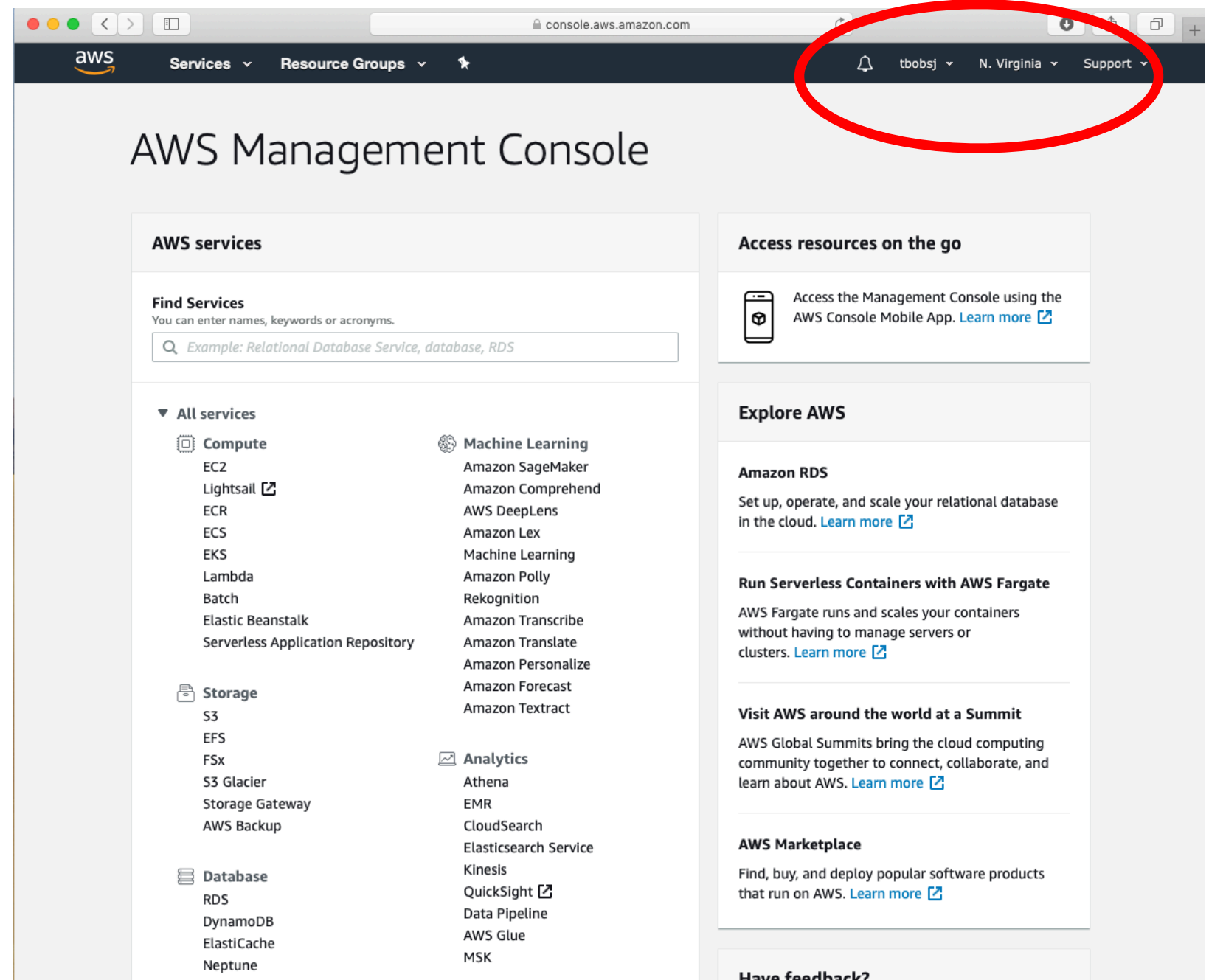
Activity

- Select Cost Explorer in Services
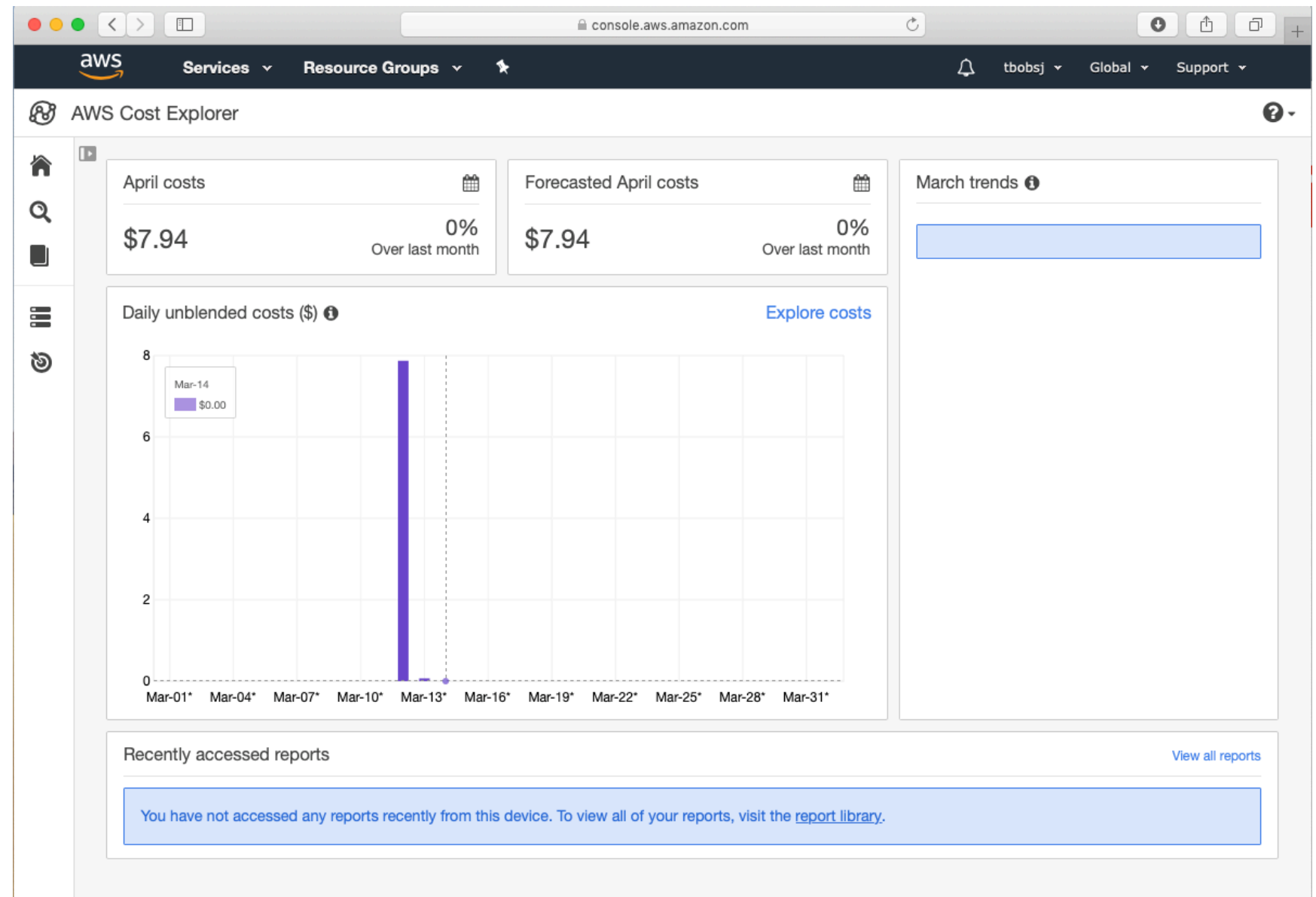
Activity

- Accessing AWS

  - MacOS and Linux: use SSH

- Access is via public—private key authentication

  - The public key is uploaded to your systems

  - The private key needs to be stored locally

    - Anywhere on your system, but you will need to provide the path to the key

  - The private key needs to have permission

    - AWS **does not** accepts keys that are publicly readable

Activity

- MacOS and Linux

  - Create the key

  - Put the private key in an easily remembered location

  - Use `chmod 400 mykey.pem` to change permissions

Activity

- Windows: Need to install and use PuTTY

- PuTTy has a tool PuTTYgen that changes key format

  - Open PuTTYGen

  - Select RSA

  - Use Load

  - Switch file type to all

  - Select mykey.pem file

  - Click save private key
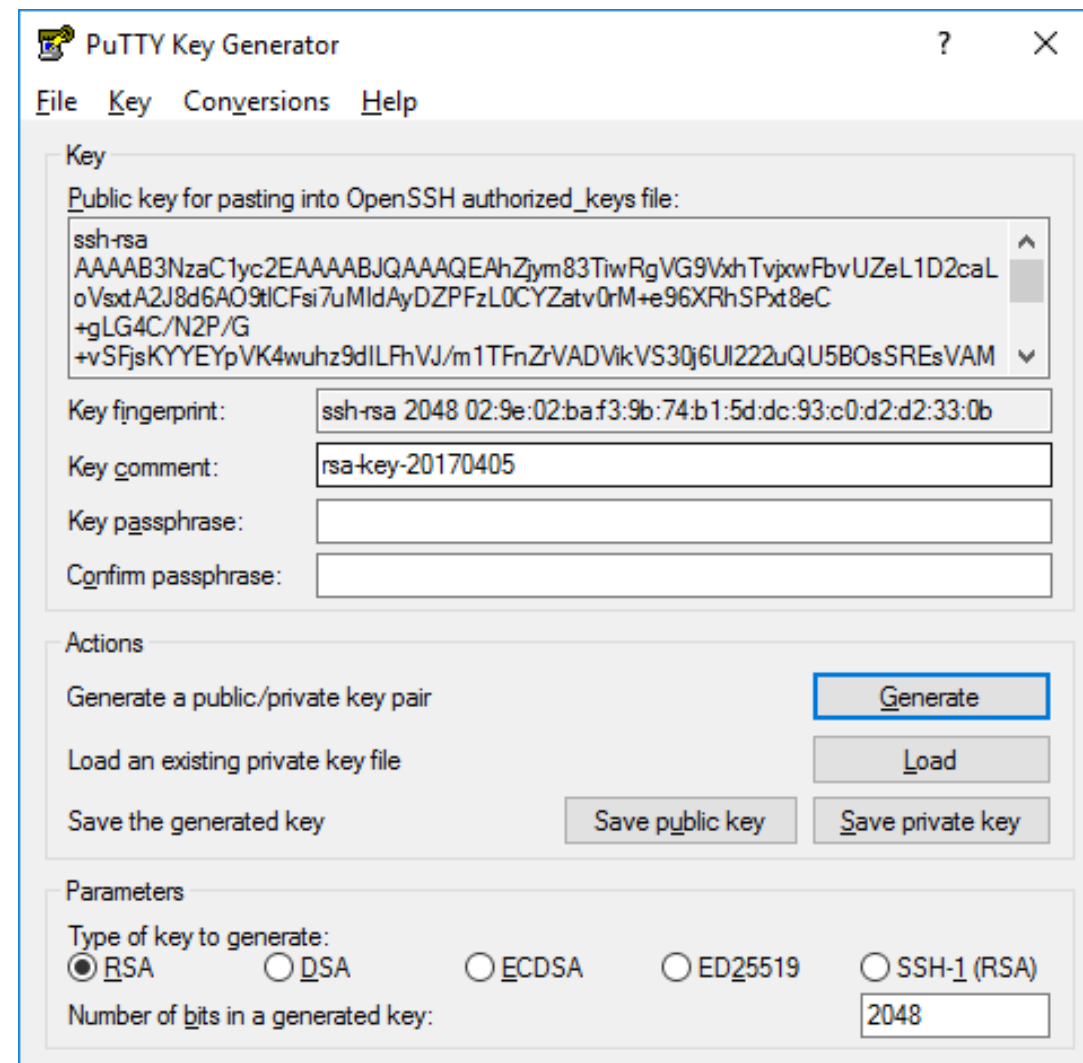
Activity

- Enable billing alarm

  - Select your name (in US Virginia)

  - Select "My Billing Dashboard"

  - Go to preferences

  - Select Receive Billing Alerts

  - Save preferences

Activity

- Now we are ready to use AWS

  - We create a simple webpage

  - This means installing PHP, MySQL, …

  - Use a number of services:

    - Elastic Load Balancing

    - Elastic Compute Cloud

    - Relational Database Service

    - Elastic File System

    - Security Groups

Activity

- Use AWS CloudFormation to do a number of things in the background

  - Create ELB (elastic load balancer)

  - Create RDS (relational database server)

  - Create and attach firewall rules

  - Create two virtual machines running web services

    - Create two VM

    - Mount file system

    - Install Apache and PHP

    - Install WordPress

    - Start the Apache Webserver

Activity

- In the AWS Management Console —>Services —> Cloudformation
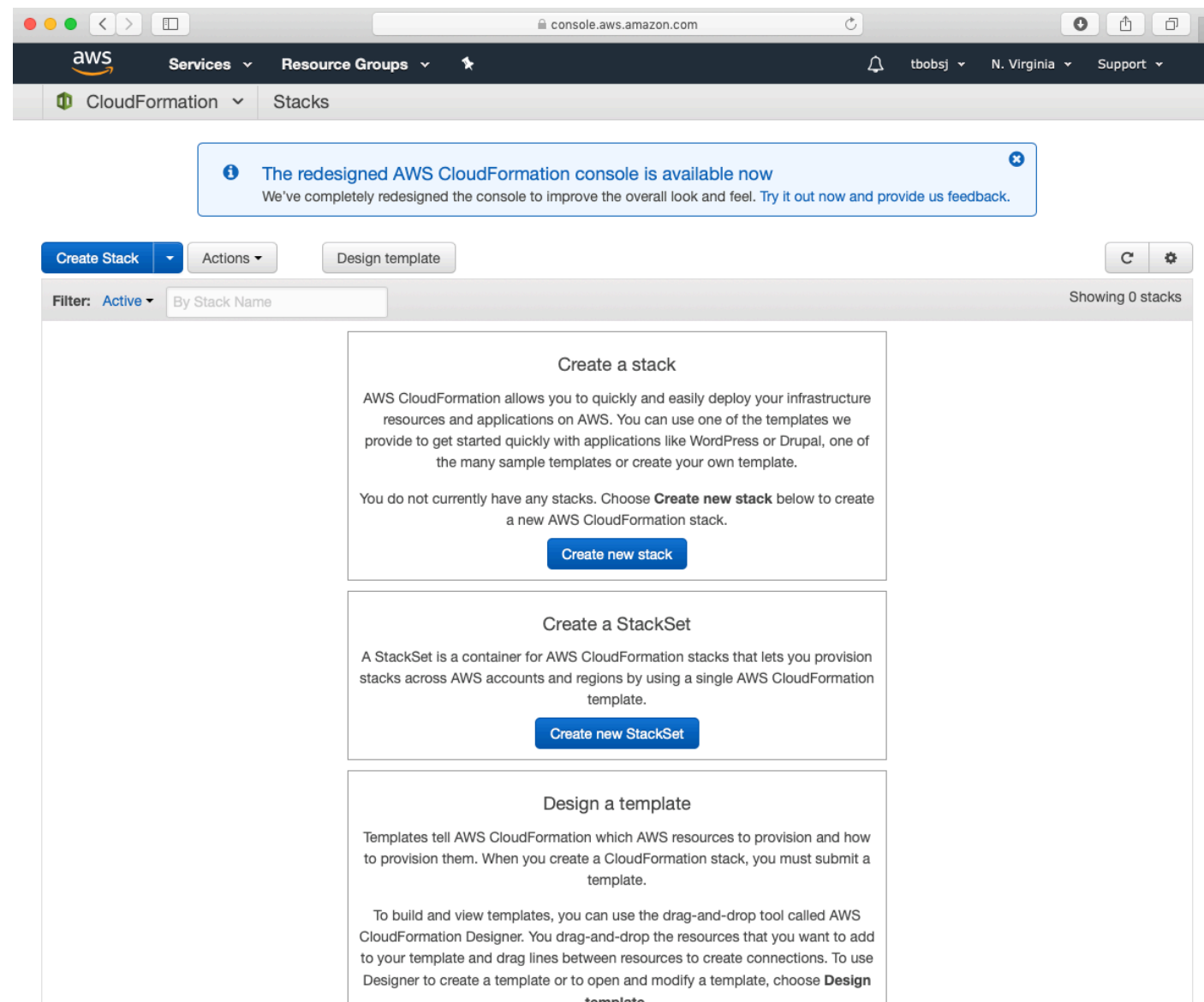
Activity

- Create the stack and select the template

    - Use template from a book

        - https://s3.amazonaws.com/awsinaction-code2/chapter02/template.yaml

        - Look at the specification: A simple document

    - Specify wordpress as stack name

    - Set the key name to the key that you created

    - Should use tags, e.g. to say that this is a wordpress system

        - `system: wordpress`

    - Specify URL in Template URL

Activity

- Cloudformation is now creating your resources

- The wordpress stack will be in status CREATE_IN_PROGRESS

- After a while, status changes to CREATE_COMPLETE

Activity

- Investigate your stack

  - Resource groups are collections of of AWS resources

- Create a resource group

  - Set group name to wordpress group

  - Tag `system: wordpress`

  - Select the region: North Virginia

  - Save

Activity

- Select instances in EC2 to see the virtual machines

- Get details on the virtual machines

Activity

- Select load balancer

  - Automatically created

Activity

- Go back to resource group wordpress

- Select DBinstances under RDS

- Get details of your SQL database

Activity

- Network file system EFS

- Cannot access through the resource group

- Need to go through EFS on the service menu

Activity

- Select the cost estimator

  - It will tell you that you have to pay about $35.00 per month for this instance

Activity

- Delete your blogging infrastructure

  - Go to CloudFormation Service in Management Console:

    - Select your wordpress stack

    - Open the action menu (Actions)

    - Click Delete Stack