

# INTRO TO DATA SCIENCE

## AMAZON WEB SERVICES

- I. SETTING UP AMAZON WEB SERVICES (AWS)**
- II. KEEPING AN EYE ON COSTS**
- III. GENERATING KEY PAIRS**
- IV. LAUNCHING AN EC2 INSTANCE**
- V. LAUNCHING A SPARK CLUSTER**

---

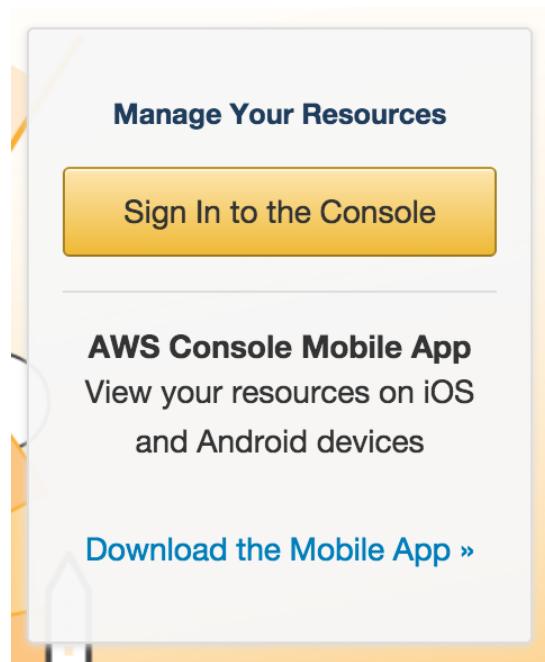
## INTRO TO DATA SCIENCE

---

# AMAZON WEB SERVICES (AWS)

*Some of the most important services are:*

- *Elastic Compute Cloud (EC2)*
- *Simple Storage Service (S3)*
- *Relational Database Service*
- *Amazon Elastic MapReduce (EMR)*
- *Tools for Deployment & Management*




AWS Services Edit

[Naeff](#) [N. Virginia](#) [Support](#)

### Amazon Web Services

#### Compute

-  **EC2**  
Virtual Servers in the Cloud
-  **Lambda**  
Run Code in Response to Events
-  **EC2 Container Service**  
Run and Manage Docker Containers

#### Storage & Content Delivery

-  **S3**  
Scalable Storage in the Cloud
-  **Elastic File System** [PREVIEW](#)  
Fully Managed File System for EC2
-  **Storage Gateway**  
Integrates On-Premises IT Environments with Cloud Storage
-  **Glacier**  
Archive Storage in the Cloud
-  **CloudFront**  
Global Content Delivery Network

#### Database

-  **RDS**  
MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
-  **DynamoDB**  
Predictable and Scalable NoSQL Data Store
-  **ElastiCache**  
In-Memory Cache
-  **Redshift**  
Managed Petabyte-Scale Data Warehouse Service

#### Networking

-  **VPC**  
Isolated Cloud Resources
-  **Direct Connect**  
Dedicated Network Connection to AWS
-  **Route 53**  
Scalable DNS and Domain Name Registration

#### Administration & Security

-  **Directory Service**  
Managed Directories in the Cloud
-  **Identity & Access Management**  
Access Control and Key Management
-  **Trusted Advisor**  
AWS Cloud Optimization Expert
-  **CloudTrail**  
User Activity and Change Tracking
-  **Config**  
Resource Configurations and Inventory
-  **CloudWatch**  
Resource and Application Monitoring
-  **Service Catalog**  
Personalized Catalog of AWS Resources

#### Deployment & Management

-  **Elastic Beanstalk**  
AWS Application Container
-  **OpsWorks**  
DevOps Application Management Service
-  **CloudFormation**  
Templated AWS Resource Creation
-  **CodeDeploy**  
Automated Deployments
-  **CodeCommit**  
Managed Git Repositories
-  **CodePipeline**  
Continuous Delivery

#### Analytics

-  **EMR**  
Managed Hadoop Framework
-  **Kinesis**  
Real-time Processing of Streaming Big Data
-  **Data Pipeline**  
Orchestration for Data-Driven Workflows
-  **Machine Learning**  
Build Smart Applications Quickly and Easily

#### Application Services

-  **SQS**  
Message Queue Service
-  **SWF**  
Workflow Service for Coordinating Application Components
-  **AppStream**  
Low Latency Application Streaming
-  **Elastic Transcoder**  
Easy-to-use Scalable Media Transcoding
-  **SES**  
Email Sending Service
-  **CloudSearch**  
Managed Search Service
-  **API Gateway**  
Build, Deploy and Manage APIs

#### Mobile Services

-  **Cognito**  
User Identity and App Data Synchronization
-  **Device Farm**  
Test Android, Fire OS, and iOS apps on real devices in the Cloud
-  **Mobile Analytics**  
Collect, View and Export App Analytics
-  **SNS**  
Push Notification Service

#### Enterprise Applications

-  **WorkSpaces**  
Desktops in the Cloud
-  **WorkDocs**  
Secure Enterprise Storage and Sharing Service
-  **WorkMail** [PREVIEW](#)  
Secure Email and Calendaring Service

### Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#)

[Tag Editor](#)

### Additional Resources

#### Getting Started

Read our [documentation](#) or view our [training](#) to learn more about AWS.

#### AWS Console Mobile App

View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

#### AWS Marketplace

Find and buy software, launch with 1-Click and pay by the hour.

#### AWS Lambda

Run your code without managing servers. Try AWS Lambda for free today.

#### Service Health

 All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

[Service Health Dashboard](#)

---

INTRO TO DATA SCIENCE

---

# KEEPING AN EYE ON COSTS

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

8

AWS Services Edit ▾

N. Virginia ▾ Support ▾

### Amazon Web Services

Compute

- EC2** Virtual Servers in the Cloud
- Lambda** Run Code in Response to Events
- EC2 Container Service** Run and Manage Docker Containers

Storage & Content Delivery

- S3** Scalable Storage in the Cloud
- Elastic File System** PREVIEW Fully Managed File System for EC2
- Storage Gateway** Integrates On-Premises IT Environments with Cloud Storage
- Glacier** Archive Storage in the Cloud
- CloudFront** Global Content Delivery Network

Database

- RDS** MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DynamoDB** Predictable and Scalable NoSQL Data Store
- ElastiCache** In-Memory Cache
- Redshift** Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC** Isolated Cloud Resources
- Direct Connect** Dedicated Network Connection to AWS
- Route 53** Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service** Managed Directories in the Cloud
- Identity & Access Management** Access Control and Key Management
- Trusted Advisor** AWS Cloud Optimization Expert
- CloudTrail** User Activity and Change Tracking
- Config** Resource Configurations and Inventory
- CloudWatch** Resource and Application Monitoring
- Service Catalog** Personalized Catalog of AWS Resources

Deployment & Management

- Elastic Beanstalk** AWS Application Container
- OpsWorks** DevOps Application Management Service
- CloudFormation** Templated AWS Resource Creation
- CodeDeploy** Automated Deployments
- CodeCommit** Managed Git Repositories
- CodePipeline** Continuous Delivery

Analytics

- EMR** Managed Hadoop Framework
- Kinesis** Real-time Processing of Streaming Big Data
- Data Pipeline** Orchestration for Data-Driven Workflows
- Machine Learning** Build Smart Applications Quickly and Easily

Application Services

- SQS** Message Queue Service
- SWF** Workflow Service for Coordinating Applications
- AppStream** Low Latency Application Streaming
- Elastic Transcoder** Easy-to-use Scalable Media Transcoding
- SES** Email Sending Service
- CloudSearch** Managed Search Service
- API Gateway** Build, Deploy and Manage APIs

Mobile Services

- Cognito** User Identity and App Data Synchronization
- Device Farm** Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics** Collect, View and Export App Analytics
- SNS** Push Notification Service

Enterprise Applications

- WorkSpaces** Desktops in the Cloud
- WorkDocs** Secure Enterprise Storage and Sharing Service
- WorkMail** PREVIEW Secure Email and Calendaring Service

My Account

Billing & Cost Management

Security Credentials

Sign Out

AWS Console Mobile App

AWS Marketplace

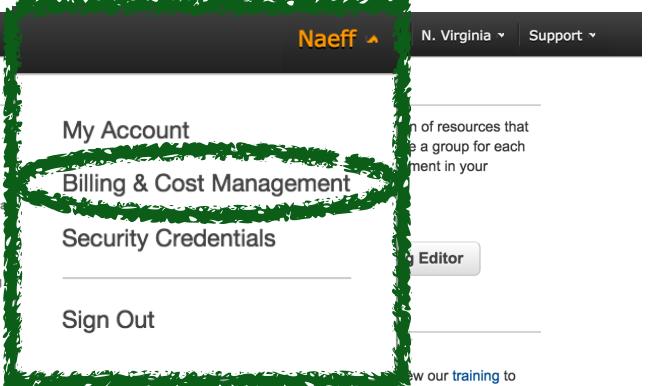
AWS Lambda

Service Health

All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

Service Health Dashboard



The screenshot shows the AWS Management Console with the user 'Naeff' logged in. The main navigation bar includes 'AWS', 'Services', 'Edit', 'N. Virginia', and 'Support'. Below the navigation is a grid of service icons. A large green circle highlights the 'Billing & Cost Management' section under 'My Account'. To the right of the grid, there's a sidebar with links like 'Editor', 'View our training to learn more about AWS.', 'AWS Console Mobile App', 'AWS Marketplace', 'AWS Lambda', 'Service Health', and status information ('All services operating normally.' and 'Updated: Aug 25 2015 15:57:00 GMT-0400').

# AMAZON WEB SERVICES

AWS Services Edit Naeff Global Support

**Dashboard** (circled)

Cost Explorer  
Budgets  
Payment Methods  
Payment History  
Consolidated Billing  
Reports  
Preferences  
Credits  
Tax Settings  
DevPay

## Billing & Cost Management Dashboard

**Spend Summary**

Welcome to the AWS Account Billing console. Your last month, month-to-date, and month-end forecasted costs appear below.

*Current month-to-date balance for September 2015*

**\$0.00**

► Important Information about these Costs  Include Subscription Charges

**Month-to-Date Spend by Service** Bill Details

The chart below shows the proportion of costs spent for each service you use.

**Month-to-Date Top Services by Spend**

	Amount
SimpleDB	\$0.00
S3	\$0.00
Tax	\$0.00
Total	\$0.00

**Alerts & Notifications**

Your account is enabled for monitoring estimated charges. Set your first billing alarm to receive an e-mail when charges reach a threshold you define.

IAM access to your account's billing information is not enabled. You can enable it on the [Account Information](#) page.



Dashboard

Bills

Cost Explorer

Budgets

Payment Methods

Payment History

Consolidated Billing

Reports

**Preferences**

Tax Settings

DevPay

## Preferences

**Receive PDF Invoice By Email**  
Turn on this feature to receive a PDF version of your invoice by email. Invoices are generally available within the first three days of the month.

**Receive Billing Alerts**  
Turn on this feature to monitor your AWS usage charges and recurring fees automatically, making it easier to track and manage your spending on AWS. You can set billing alerts to receive email notifications when your charges reach a specified threshold. Once enabled, this preference cannot be disabled. [Manage Billing Alerts](#)

**Receive Billing Reports**  
Turn on this feature to receive ongoing reports of your AWS charges once or more daily. AWS delivers these reports to the Amazon S3 bucket that you specify where indicated below. For consolidated billing customers, AWS generates reports only for paying accounts. Linked accounts cannot sign up for billing reports.

Save to S3 Bucket:  [Verify](#)

[Save preferences](#)

The screenshot shows the AWS CloudWatch Billing Alarms interface. On the left, a sidebar menu includes 'Dashboard', 'Alarms' (which is selected and highlighted in orange), 'ALARM' (0 notifications), 'INSUFFICIENT' (0 notifications), 'Billing' (which is circled in green), and 'Metrics'. Under 'Metrics', there are links for 'Selected Metrics', 'Billing', 'EBS', 'EC2', 'EMR', and 'S3'. The main content area is titled 'Billing Alarms' and contains the following text:

Amazon CloudWatch can help you monitor the charges on your [AWS bill](#) by sending you email alerts when charges exceed a threshold you define.

Once you update your preferences in the Account Billing console, you will begin receiving Amazon CloudWatch metrics that reflect your month-to-date AWS charges. Then, you can create a billing alarm by specifying a spending threshold and an e-mail address to notify. [Learn more about billing alerts](#)

You get 10 free alarms and 1,000 free e-mail notifications each month as part of the [AWS Free Tier](#).

A prominent blue button labeled 'Create Alarm' is located at the bottom of this section. The entire 'Create Alarm' button is also circled in green.

On the right side, there is an 'Additional Info' sidebar with links to 'Getting Started Guide', 'Monitoring Scripts Guide', 'Overview and Features', 'Documentation', 'Forums', and 'Report an Issue'.

At the top right of the main content area, there are user account links for 'Naeff', 'N. Virginia', and 'Support'.

AWS Services Edit

Dashboard Alarms ALARM INSUFFICIENT OK Billing Logs Metrics Selected Metrics Billing EBS EC2 EMR S3

Billing Alarm Amazon CloudWatch Once you update Then, you can cr You get 10 free a Create Alarm

**Create Alarm**

## Billing Alarm

You can create a billing alarm to receive e-mail alerts when your AWS charges exceed a threshold you choose. Simply:

1. Enter a spending threshold
2. Provide an email address
3. Check your inbox for a confirmation email and click the link provided

When my total estimated charges for the month exceed: \$  USD send a notification to:

Reminder: for each address you specify, we'll send an email from AWS with the subject "AWS Notification - Subscription Confirmation". Click the link provided in the message to confirm that AWS may deliver alerts to that address.

showing simple options | show advanced

**Alarm Preview**

This alarm will trigger when the blue line goes above the red line

EstimatedCharges > 0

0 0.25 0.5 0.75 1

8/27 00:00 8/29 00:00 8/31 00:00

**More resources**

[AWS Billing console](#)  
[Getting started with billing alarms](#)  
[More help with billing alarms](#)  
[AWS Billing FAQs](#)

**Create Alarm**

Naeff N. Virginia Support Additional Info Getting Started Guide Monitoring Scripts Guide Overview and Features Documentation Forums Report an Issue

AWS Services
Naeff ▾ N. Virginia ▾ Support ▾

---

**Compute**

- EC2 Virtual Servers in the Cloud
- Lambda Run Code in Response to Events
- EC2 Container Service Run and Manage Docker Containers

**Storage & Content Delivery**

- S3 Scalable Storage in the Cloud
- Elastic File System PREVIEW Fully Managed File System for EC2
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
- Glacier Archive Storage in the Cloud
- CloudFront Global Content Delivery Network

**Database**

- RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DynamoDB Predictable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift Managed Petabyte-Scale Data Warehouse Service

**Networking**

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Scalable DNS and Domain Name Registration

**Administration & Security**

- Directory Service Managed Directories in the Cloud
- Identity & Access Management Access Control and Key Management
- Trusted Advisor AWS Cloud Optimization Expert
- CloudTrail User Activity and Change Tracking
- Config Resource Configurations and Inventory
- CloudWatch Resource and Application Monitoring
- Service Catalog Personalized Catalog of AWS Resources

**Deployment & Management**

- Elastic Beanstalk AWS Application Container
- OpsWorks DevOps Application Management Service
- CloudFormation Templated AWS Resource Creation
- CodeDeploy Automated Deployments
- CodeCommit Managed Git Repositories
- CodePipeline Continuous Delivery

**Analytics**

- EMR Managed Hadoop Framework
- Kinesis Real-time Processing of Streaming Big Data
- Data Pipeline Orchestration for Data-Driven Workflows
- Machine Learning Build Smart Applications Quickly and Easily

**Application Services**

- SQS Message Queue Service
- SWF Workflow Service for Coordinating Application Components
- AppStream Low Latency Application Streaming
- Elastic Transcoder Easy-to-use Scalable Media Transcoding
- SES Email Sending Service
- CloudSearch Managed Search Service
- API Gateway Build, Deploy and Manage APIs

**Mobile Services**

- Cognito User Identity and App Data Synchronization
- Device Farm Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics Collect, View and Export App Analytics
- SNS Push Notification Service

**Enterprise Applications**

- WorkSpaces Desktops in the Cloud
- WorkD Secure
- WorkL Secure

**Resource Groups**

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#)
[Tag Editor](#)

**Additional Resources**

**Getting Started**  
Read our [documentation](#) or view our [training](#) to learn more about AWS.

**AWS Console Mobile App**  
View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

**AWS Marketplace**  
Find and buy software, launch with 1-Click and pay by the hour.

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

14

AWS Services

Compute

**EC2** Virtual Servers in the Cloud

Run and Manage Docker Containers

Storage & Content Delivery

- S3 Scalable Storage in the Cloud
- Elastic File System PREVIEW Fully Managed File System for EC2
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
- Glacier Archive Storage in the Cloud
- CloudFront Global Content Delivery Network

Database

- RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DYNAMODB Predictable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service Managed Directories in the Cloud
- Identity & Access Management Access Control and Key Management
- Trusted Advisor AWS Cloud Optimization Expert
- CloudTrail User Activity and Change Tracking
- Config Resource Configurations and Inventory
- CloudWatch Resource and Application Monitoring
- Service Catalog Personalized Catalog of AWS Resources

Deployment & Management

- Elastic Beanstalk AWS Application Container
- OpsWorks DevOps Application Management Service
- CloudFormation Templated AWS Resource Creation
- CodeDeploy Automated Deployments
- CodeCommit Managed Git Repositories
- CodePipeline Continuous Delivery

Analytics

- EMR Managed Hadoop Framework
- Kinesis Real-time Processing of Streaming Big Data
- Data Pipeline Orchestration for Data-Driven Workflows
- Machine Learning Build Smart Applications Quickly and Easily

Application Services

- SQS Message Queue Service
- SWF Workflow Service for Coordinating Application Components
- AppStream Low Latency Application Streaming
- Elastic Transcoder Easy-to-use Scalable Media Transcoding
- SES Email Sending Service
- CloudSearch Managed Search Service
- API Gateway Build, Deploy and Manage APIs

Mobile Services

- Cognito User Identity and App Data Synchronization
- Device Farm Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics Collect, View and Export App Analytics
- SNS Push Notification Service

Enterprise Applications

- WorkSpaces Desktops in the Cloud
- WorkD Secure
- Work Secure L

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

Create a Group Tag Editor

Additional Resources

Getting Started

Read our [documentation](#) or view our [training](#) to learn more about AWS.

AWS Console Mobile App

View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

AWS Marketplace

Find and buy software, launch with 1-Click and pay by the hour.

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with navigation links like EC2 Dashboard, Events, Tags, Reports, Limits, Instances, Spot Requests, Reserved Instances, Images, AMIs, Bundle Tasks, Elastic Block Store, Volumes, Snapshots, Network & Security, Security Groups, Elastic IPs, Placement Groups, Key Pairs, and Network Interfaces. The main area displays a table of four EC2 instances. The fourth instance, with the ID i-9dcdf236, has a context menu open over it. The menu includes options like Connect, Get Windows Password, Launch More Like This, Instance State (with Start, Stop, and Terminate), Instance Settings, Image, Networking, and CloudWatch Monitoring. The 'Terminate' option under Instance State is highlighted with a red box. The entire context menu is also circled in red. To the right of the table, there are navigation icons and a message indicating '1 to 4 of 4' instances.

Name	Instance ID	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP	Key Name	Monitor
i-3c031ce...	i-3c031ce...	us-east-1c	running	2/2 checks ...	None	ec2-52-7-232-239.com...	52.7.232.239	ga	disabled
i-3d031cef...	i-3d031cef...	us-east-1c	running	2/2 checks ...	None	ec2-52-20-51-153.com...	52.20.51.153	ga	disabled
i-9b021d49...	i-9b021d49...	us-east-1c	running	2/2 checks ...	None	ec2-52-22-146-230.co...	52.22.146.230	ga	disabled
i-9dcdf236...	i-9dcdf236...	us-east-1c	running	2/2 checks ...	None	ec2-52-21-176-167.co...	52.21.176.167	ga	disabled

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

---

INTRO TO DATA SCIENCE

---

# GENERATING YOUR KEY PAIRS

The screenshot shows the AWS Management Console homepage with a navigation bar at the top. The navigation bar includes the AWS logo, a dropdown menu, a Services dropdown, and an Edit button. On the right side of the bar are the user name "Naeff", a dropdown for "N. Virginia", and a "Support" link.

The main content area is titled "Amazon Web Services" and lists various service categories:

- Compute**
  - EC2** Virtual Servers in the Cloud
  - Lambda** Run Code in Response to Events
  - EC2 Container Service** Run and Manage Docker Containers
- Storage & Content Delivery**
  - S3** Scalable Storage in the Cloud
  - Elastic File System** PREVIEW Fully Managed File System for EC2
  - Storage Gateway** Integrates On-Premises IT Environments with Cloud Storage
  - Glacier** Archive Storage in the Cloud
  - CloudFront** Global Content Delivery Network
- Database**
  - RDS** MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
  - DynamoDB** Predictable and Scalable NoSQL Data Store
  - ElastiCache** In-Memory Cache
  - Redshift** Managed Petabyte-Scale Data Warehouse Service
- Networking**
  - VPC** Isolated Cloud Resources
  - Direct Connect** Dedicated Network Connection to AWS
  - Route 53** Scalable DNS and Domain Name Registration
- Administration & Security**
  - Directory Service** Managed Directories in the Cloud
  - Identity & Access Management** Access Control and Key Management
  - Trusted Advisor** AWS Cloud Optimization Expert
  - CloudTrail** User Activity and Change Tracking
  - Config** Resource Configurations and Inventory
  - CloudWatch** Resource and Application Monitoring
  - Service Catalog** Personalized Catalog of AWS Resources
- Deployment & Management**
  - Elastic Beanstalk** AWS Application Container
  - OpsWorks** DevOps Application Management Service
  - CloudFormation** Templated AWS Resource Creation
  - CodeDeploy** Automated Deployments
  - CodeCommit** Managed Git Repositories
  - CodePipeline** Continuous Delivery
- Analytics**
  - EMR** Managed Hadoop Framework
  - Kinesis** Real-time Processing of Streaming Big Data
  - Data Pipeline** Orchestration for Data-Driven Workflows
  - Machine Learning** Build Smart Applications Quickly and Easily
- Application Services**
  - SQS** Message Queue Service
  - SWF** Workflow Service for Coordinating Applications
  - AppStream** Low Latency Application Streaming
  - Elastic Transcoder** Easy-to-use Scalable Media Transcoding
  - SES** Email Sending Service
  - CloudSearch** Managed Search Service
  - API Gateway** Build, Deploy and Manage APIs
- Mobile Services**
  - Cognito** User Identity and App Data Synchronization
  - Device Farm** Test Android, Fire OS, and iOS apps on real devices in the Cloud
  - Mobile Analytics** Collect, View and Export App Analytics
  - SNS** Push Notification Service
- Enterprise Applications**
  - WorkSpaces** Desktops in the Cloud
  - WorkDocs** Secure Enterprise Storage and Sharing Service
  - WorkMail** PREVIEW Secure Email and Calendaring Service
- My Account**
- Billing & Cost Management**
- Security Credentials** (This link is highlighted with a green oval)
- Sign Out**

On the right side of the page, there is a sidebar with the following text:

of resources that a group for each in your

Editor

Check our Documentation or view our training to learn more about AWS.

**AWS Console Mobile App**  
View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

**AWS Marketplace**  
Find and buy software, launch with 1-Click and pay by the hour.

**AWS Lambda**  
Run your code without managing servers. Try AWS Lambda for free today.

**Service Health**

All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

[Service Health Dashboard](#)

The screenshot shows the AWS Identity and Access Management (IAM) service interface. On the left, a navigation sidebar lists several options: Dashboard, Details, Groups, **Users**, Roles, Policies, Identity Providers, Account Settings, Credential Report, and Encryption Keys. The 'Users' option is highlighted with a green oval. The main content area has a dark header bar with the AWS logo, 'AWS Services', and 'Edit' dropdowns. Below the header are two buttons: 'Create New Users' (blue) and 'User Actions' (grey). A search bar is followed by a table with two columns: 'User Name' and 'Groups'. The table contains one row for 'Ruben', which is selected (indicated by a blue border around the row). The 'User Name' column shows a checkbox and 'Ruben', and the 'Groups' column shows '0'.

User Name	Groups
Ruben	0

The screenshot shows the AWS IAM (Identity and Access Management) service interface. The top navigation bar includes the AWS logo, a Services dropdown, and an Edit dropdown. On the left, a sidebar menu lists several options: Dashboard, Details, Groups, **Users** (which is selected and highlighted with a blue border), Roles, Policies, Identity Providers, Account Settings, Credential Report, and Encryption Keys. The main content area is titled "Create New Users" and "User Actions". It features a search bar and a table with two columns: "User Name" and "Groups". The table contains one row for a user named "Ruben", who has 0 groups assigned. A green oval highlights the "Create New Users" button.

User Name	Groups
Ruben	0

The screenshot shows the AWS IAM (Identity and Access Management) service interface. The left sidebar has a navigation menu with the following items:

- Dashboard
- Details
- Groups
- Users** (highlighted with a blue border)
- Roles
- Policies
- Identity Providers
- Account Settings
- Credential Report

A horizontal line separates this from the main content area.

The main content area has a dark header bar with the following buttons:

- AWS icon
- Services dropdown
- Edit dropdown

Below the header are two buttons:

- Create New Users** (blue button)
- User Actions dropdown

The main content area contains a search bar labeled "Search". Below it is a table with the following columns:

<input type="checkbox"/>	User Name <span>▲</span>	Groups
<input type="checkbox"/>	Ruben	0

At the bottom right of the main content area, there is a green oval highlighting a blue button labeled "Download Credentials". To its left is a "Close" link.

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

21

AWS Services

**Compute**

**EC2** Virtual Servers in the Cloud

Run and Manage Docker Containers

**Storage & Content Delivery**

- S3 Scalable Storage in the Cloud
- Elastic File System [PREVIEW](#) Fully Managed File System for EC2
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
- Glacier Archive Storage in the Cloud
- CloudFront Global Content Delivery Network

**Database**

- RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DYNAMODB Predictable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift Managed Petabyte-Scale Data Warehouse Service

**Networking**

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Scalable DNS and Domain Name Registration

**Administration & Security**

- Directory Service Managed Directories in the Cloud
- Identity & Access Management Access Control and Key Management
- Trusted Advisor AWS Cloud Optimization Expert
- CloudTrail User Activity and Change Tracking
- Config Resource Configurations and Inventory
- CloudWatch Resource and Application Monitoring
- Service Catalog Personalized Catalog of AWS Resources

**Deployment & Management**

- Elastic Beanstalk AWS Application Container
- OpsWorks DevOps Application Management Service
- CloudFormation Templated AWS Resource Creation
- CodeDeploy Automated Deployments
- CodeCommit Managed Git Repositories
- CodePipeline Continuous Delivery

**Analytics**

- EMR Managed Hadoop Framework
- Kinesis Real-time Processing of Streaming Big Data
- Data Pipeline Orchestration for Data-Driven Workflows
- Machine Learning Build Smart Applications Quickly and Easily

**Application Services**

- SQS Message Queue Service
- SWF Workflow Service for Coordinating Application Components
- AppStream Low Latency Application Streaming
- Elastic Transcoder Easy-to-use Scalable Media Transcoding
- SES Email Sending Service
- CloudSearch Managed Search Service
- API Gateway Build, Deploy and Manage APIs

**Mobile Services**

- Cognito User Identity and App Data Synchronization
- Device Farm Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics Collect, View and Export App Analytics
- SNS Push Notification Service

**Enterprise Applications**

- WorkSpaces Desktops in the Cloud
- WorkDocs Secure Enterprise Storage and Sharing Service
- WorkMail [PREVIEW](#) Secure Email and Calendaring Service

**Resource Groups**

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#) [Tag Editor](#)

**Additional Resources**

**Getting Started**  
Read our [documentation](#) or view our [training](#) to learn more about AWS.

**AWS Console Mobile App**  
View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

**AWS Marketplace**  
Find and buy software, launch with 1-Click and pay by the hour.

**AWS Lambda**  
Run your code without managing servers. Try AWS Lambda for free today.

**Service Health**

All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

[Service Health Dashboard](#)

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

**EC2 Dashboard**

- Events
- Tags
- Reports
- Limits

**INSTANCES**

- Instances
- Spot Requests
- Reserved Instances

**IMAGES**

- AMIs
- Bundle Tasks

**ELASTIC BLOCK STORE**

- Volumes
- Snapshots

**NETWORK & SECURITY**

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces

**LOAD BALANCING**

- Load Balancers

**AUTO SCALING**

- Launch Configurations
- Auto Scaling Groups

**Resources**

You are using the following Amazon EC2 resources in the US East (N. Virginia) region:

0 Running Instances	0 Elastic IPs
0 Volumes	0 Snapshots
2 Key Pairs	0 Load Balancers
0 Placement Groups	3 Security Groups

Automate application deployments to EC2 with [CodeDeploy](#).

**Create Instance**

To start using Amazon EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

**Launch Instance**

Note: Your instances will launch in the US East (N. Virginia) region

**Service Health**

**Service Status:**

- US East (N. Virginia): This service is operating normally

**Availability Zone Status:**

- us-east-1a: Availability zone is operating normally
- us-east-1c: Availability zone is operating normally
- us-east-1d: Availability zone is operating normally
- us-east-1e: Availability zone is operating normally

**Scheduled Events**

**US East (N. Virginia):**

No events

**Account Attributes**

Supported Platforms VPC

Default VPC [vpc-978576f2](#)

**Additional Information**

Getting Started Guide Documentation All EC2 Resources Forums Pricing Contact Us

**AWS Marketplace**

Find free software trial products in the AWS Marketplace from the [EC2 Launch Wizard](#). Or try these popular AMIs:

**Brocade 5400 Virtual Router/Firewall/VPN**

Provided by Brocade Rating ★★★★☆ Pay by the hour for software and AWS usage [View all Networking](#)

**Alert Logic Threat Manager for AWS**

Provided by Alert Logic, Inc. Rating ★★★★★ Pay by the hour for software and AWS usage [View all Security Software](#)

**TIBCO Spotfire Analytics Platform (Hourly)**

Provided by TIBCO Software, Inc.

The screenshot shows the AWS EC2 Key Pairs page. The top navigation bar includes the AWS logo, a Services dropdown, an Edit dropdown, and user account information (Naeff, N. Virginia, Support). The left sidebar lists various EC2 services: EC2 Dashboard, Events, Tags, Reports, Limits, Instances (with sub-options: Instances, Spot Requests, Reserved Instances), Images (with sub-options: AMIs, Bundle Tasks), Elastic Block Store (with sub-options: Volumes, Snapshots), Network & Security (with sub-options: Security Groups, Elastic IPs, Placement Groups, Key Pairs, Network Interfaces), Load Balancing (with sub-option: Load Balancers), and Auto Scaling (with sub-options: Launch Configurations, Auto Scaling Groups). The main content area displays a message: "You do not have any Key Pairs in this region. Click the 'Create Key Pair' button to create your first Key Pair." A prominent blue "Create Key Pair" button is centered below the message. A green circle highlights this button. At the bottom right of the main content area, there are three small square icons.

EC2 Dashboard | AWS | Services | Edit | Naeff | N. Virginia | Support

Create Key Pair | Import Key Pair | Delete

Events | Tags | Reports | Limits

INSTANCES

- Instances
- Spot Requests
- Reserved Instances

IMAGES

- AMIs
- Bundle Tasks

ELASTIC BLOCK STORE

- Volumes
- Snapshots

NETWORK & SECURITY

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs**
- Network Interfaces

LOAD BALANCING

- Load Balancers

AUTO SCALING

- Launch Configurations
- Auto Scaling Groups

Filter by attributes or search by keyword

You do not have any Key Pairs in this region.

Click the "Create Key Pair" button to create your first Key Pair.

Create Key Pair

Select a key pair

The screenshot shows the AWS EC2 dashboard with the 'Key Pairs' section selected. A modal window titled 'Create Key Pair' is open, prompting the user to enter a key pair name. The name 'ga' has been typed into the input field, which is highlighted with a green oval. The 'Create' button at the bottom right of the modal is also highlighted with a green oval. The background shows a message indicating no key pairs exist in the region, and a search bar at the top.

AWS Services Edit

EC2 Dashboard Events Tags Reports Limits

INSTANCES Instances Spot Requests Reserved Instances

IMAGES AMIs Bundle Tasks

ELASTIC BLOCK STORE Volumes Snapshots

NETWORK & SECURITY Security Groups Elastic IPs Placement Groups

**Key Pairs** Network Interfaces

LOAD BALANCING Load Balancers

AUTO SCALING Launch Configurations Auto Scaling Groups

Naeff N. Virginia Support

Create Key Pair Import Key Pair Delete

Filter by attributes or search by keyword

You do not have any Key Pairs in this region.

Click the "Create Key Pair" button to create your first Key Pair.

## Create Key Pair

Key pair name:

Cancel Create

Select a key pair

- ▶ *Download your key pair file (e.g., ga.pem)*
- ▶ *You can name it and save the .pem locally wherever you'd like.  
We save it in a dedicated folder (e.g., ~/.aws/)*
- ▶ *Set the proper permissions:* chmod 600 ~/.aws/ga.pem

---

INTRO TO DATA SCIENCE

---

# LAUNCHING AN EC2 INSTANCE

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

27

AWS Services

Compute

**EC2** Virtual Servers in the Cloud

Run and Manage Docker Containers

Storage & Content Delivery

- S3 Scalable Storage in the Cloud
- Elastic File System [PREVIEW](#) Fully Managed File System for EC2
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
- Glacier Archive Storage in the Cloud
- CloudFront Global Content Delivery Network

Database

- RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DYNAMODB Predictable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service Managed Directories in the Cloud
- Identity & Access Management Access Control and Key Management
- Trusted Advisor AWS Cloud Optimization Expert
- CloudTrail User Activity and Change Tracking
- Config Resource Configurations and Inventory
- CloudWatch Resource and Application Monitoring
- Service Catalog Personalized Catalog of AWS Resources

Deployment & Management

- Elastic Beanstalk AWS Application Container
- OpsWorks DevOps Application Management Service
- CloudFormation Templated AWS Resource Creation
- CodeDeploy Automated Deployments
- CodeCommit Managed Git Repositories
- CodePipeline Continuous Delivery

Analytics

- EMR Managed Hadoop Framework
- Kinesis Real-time Processing of Streaming Big Data
- Data Pipeline Orchestration for Data-Driven Workflows
- Machine Learning Build Smart Applications Quickly and Easily

Application Services

- SQS Message Queue Service
- SWF Workflow Service for Coordinating Application Components
- AppStream Low Latency Application Streaming
- Elastic Transcoder Easy-to-use Scalable Media Transcoding
- SES Email Sending Service
- CloudSearch Managed Search Service
- API Gateway Build, Deploy and Manage APIs

Mobile Services

- Cognito User Identity and App Data Synchronization
- Device Farm Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics Collect, View and Export App Analytics
- SNS Push Notification Service

Enterprise Applications

- WorkSpaces Desktops in the Cloud
- WorkDocs Secure Enterprise Storage and Sharing Service
- WorkMail [PREVIEW](#) Secure Email and Calendaring Service

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#) [Tag Editor](#)

Additional Resources

Getting Started

Read our [documentation](#) or view our [training](#) to learn more about AWS.

**AWS Console Mobile App**

View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

**AWS Marketplace**

Find and buy software, launch with 1-Click and pay by the hour.

**AWS Lambda**

Run your code without managing servers. Try AWS Lambda for free today.

Service Health

All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

[Service Health Dashboard](#)

AWS Services Edit Naeff N. Virginia Support

**EC2 Dashboard**

- Events
- Tags
- Reports
- Limits

**INSTANCES**

- Instances
- Spot Requests
- Reserved Instances

**IMAGES**

- AMIs
- Bundle Tasks

**ELASTIC BLOCK STORE**

- Volumes
- Snapshots

**NETWORK & SECURITY**

- Security Groups
- Elastic IPs
- Placement Groups
- Key Pairs
- Network Interfaces

**LOAD BALANCING**

- Load Balancers

**AUTO SCALING**

- Launch Configurations
- Auto Scaling Groups

**Resources**

You are using the following Amazon EC2 resources in the US East (N. Virginia) region:

0 Running Instances	0 Elastic IPs
0 Volumes	0 Snapshots
2 Key Pairs	0 Load Balancers
0 Placement Groups	3 Security Groups

Automate application deployments to EC2 with [CodeDeploy](#).

**Create Instance**

To get started with EC2 you will want to launch a virtual server, known as an Amazon EC2 instance.

**Launch Instance**

Note: Your instances will launch in the US East (N. Virginia) region

**Service Health**

**Service Status:**

- US East (N. Virginia): This service is operating normally

**Availability Zone Status:**

- us-east-1a: Availability zone is operating normally
- us-east-1c: Availability zone is operating normally
- us-east-1d: Availability zone is operating normally
- us-east-1e: Availability zone is operating normally

[Service Health Dashboard](#)

**Scheduled Events**

**US East (N. Virginia):**

No events

**Account Attributes**

Supported Platforms VPC

Default VPC [vpc-978576f2](#)

**Additional Information**

Getting Started Guide Documentation All EC2 Resources Forums Pricing Contact Us

**AWS Marketplace**

Find **free software** trial products in the AWS Marketplace from the [EC2 Launch Wizard](#). Or try these popular AMIs:

**Brocade 5400 Virtual Router/Firewall/VPN**

Provided by Brocade Rating ★★★★☆ Pay by the hour for software and AWS usage [View all Networking](#)

**Alert Logic Threat Manager for AWS**

Provided by Alert Logic, Inc. Rating ★★★★★ Pay by the hour for software and AWS usage [View all Security Software](#)

**TIBCO Spotfire Analytics Platform (Hourly)**

Provided by TIBCO Software, Inc.

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

**Step 1: Choose an Amazon Machine Image (AMI)**

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

**Quick Start**

- My AMIs
- AWS Marketplace
- Community AMIs
- Free tier only ⓘ

		1 to 22 of 22 AMIs	
	<b>Amazon Linux AMI 2015.03.1 (HVM), SSD Volume Type - ami-0d4cf66</b>		64-bit
	The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.		
	Root device type: ebs Virtualization type: hvm		
	<b>Red Hat Enterprise Linux 7.1 (HVM), SSD Volume Type - ami-12663b7a</b>		64-bit
	Red Hat Enterprise Linux version 7.1 (HVM), EBS General Purpose (SSD) Volume Type		
	Root device type: ebs Virtualization type: hvm		
	<b>SUSE Linux Enterprise Server 12 (HVM), SSD Volume Type - ami-aeb532c6</b>		64-bit
	SUSE Linux Enterprise Server 12 (HVM), EBS General Purpose (SSD) Volume Type. Public Cloud, Advanced Systems Management, Web and Scripting, and Legacy modules enabled.		
	Root device type: ebs Virtualization type: hvm		
	<b>Ubuntu Server 14.04 LTS (HVM), SSD Volume Type - ami-d05e75b8</b>		64-bit
	Ubuntu Server 14.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical ( <a href="http://www.ubuntu.com/cloud/services">http://www.ubuntu.com/cloud/services</a> ).		
	Root device type: ebs Virtualization type: hvm		
	<b>Microsoft Windows Server 2012 R2 Base - ami-cd9339a6</b>		64-bit
	Microsoft Windows 2012 R2 Standard edition with 64-bit architecture. [English]		
	Root device type: ebs Virtualization type: hvm		



1. Choose AMI   2. Choose Instance Type   3. Configure Instance   4. Add Storage   5. Tag Instance   6. Configure Security Group   7. Review

## Step 2: Choose an Instance Type

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Filter by: All instance types ▾   Current generation ▾   Show/Hide Columns

Currently selected: t2.micro (Variable ECUs, 1 vCPUs, 2.5 GHz, Intel Xeon Family, 1 GiB memory, EBS only)

	Family	Type	vCPUs ⓘ	Memory (GiB)	Instance Storage (GB) ⓘ	EBS-Optimized Available ⓘ	Network Performance ⓘ
<input checked="" type="checkbox"/>	General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate
<input type="checkbox"/>	General purpose	t2.small	1	2	EBS only	-	Low to Moderate
<input type="checkbox"/>	General purpose	t2.medium	2	4	EBS only	-	Low to Moderate
<input type="checkbox"/>	General purpose	t2.large	2	8	EBS only	-	Low to Moderate
<input type="checkbox"/>	General purpose	m4.large	2	8	EBS only	Yes	Moderate
<input type="checkbox"/>	General purpose	m4.xlarge	4	16	EBS only	Yes	High
<input type="checkbox"/>	General purpose	m4.2xlarge	8	32	EBS only	Yes	High
<input type="checkbox"/>	General purpose	m4.4xlarge	16	64	EBS only	Yes	High

Cancel Previous **Review and Launch** Next: Configure Instance Details

AWS Services Edit

Naeff N. Virginia Support

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

## Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

**A Improve your instances' security. Your security group, launch-wizard-1, is open to the world.**

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only. You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)

▶ AMI Details [Edit AMI](#)

▶ Instance Type [Edit instance type](#)

▶ Security Groups [Edit security groups](#)

▶ Instance Details [Edit instance details](#)

▶ Storage [Edit storage](#)

▶ Tags [Edit tags](#)

Cancel Previous **Launch**

AWS Services Edit Naeff N. Virginia Support

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

## Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

**⚠ Improve your instances' security. Your security group, launch-wizard-1, is open to the world.**

Your instances may be accessible from any IP address. We recommend that you update your security group's rules to allow access from known IP addresses only. You can also open additional ports in your security group.

**Select an existing key pair or create a new key pair**

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about [removing existing key pairs from a public AMI](#).

Choose an existing key pair

Select a key pair

ga

I acknowledge that I have access to the selected private key file (ga.pem), and that without this file, I won't be able to log into my instance.

CANCEL Launch Instances

Edit AMI Edit instance type Edit security groups Edit instance details Edit storage Edit tags

Cancel Previous Launch

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

33

AWS Services

Compute

**EC2** Virtual Servers in the Cloud

Run and Manage Docker Containers

Storage & Content Delivery

- S3 Scalable Storage in the Cloud
- Elastic File System [PREVIEW](#) Fully Managed File System for EC2
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
- Glacier Archive Storage in the Cloud
- CloudFront Global Content Delivery Network

Database

- RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DYNAMODB Predictable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service Managed Directories in the Cloud
- Identity & Access Management Access Control and Key Management
- Trusted Advisor AWS Cloud Optimization Expert
- CloudTrail User Activity and Change Tracking
- Config Resource Configurations and Inventory
- CloudWatch Resource and Application Monitoring
- Service Catalog Personalized Catalog of AWS Resources

Deployment & Management

- Elastic Beanstalk AWS Application Container
- OpsWorks DevOps Application Management Service
- CloudFormation Templated AWS Resource Creation
- CodeDeploy Automated Deployments
- CodeCommit Managed Git Repositories
- CodePipeline Continuous Delivery

Analytics

- EMR Managed Hadoop Framework
- Kinesis Real-time Processing of Streaming Big Data
- Data Pipeline Orchestration for Data-Driven Workflows
- Machine Learning Build Smart Applications Quickly and Easily

Application Services

- SQS Message Queue Service
- SWF Workflow Service for Coordinating Application Components
- AppStream Low Latency Application Streaming
- Elastic Transcoder Easy-to-use Scalable Media Transcoding
- SES Email Sending Service
- CloudSearch Managed Search Service
- API Gateway Build, Deploy and Manage APIs

Mobile Services

- Cognito User Identity and App Data Synchronization
- Device Farm Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics Collect, View and Export App Analytics
- SNS Push Notification Service

Enterprise Applications

- WorkSpaces Desktops in the Cloud
- WorkDocs Secure Enterprise Storage and Sharing Service
- WorkMail [PREVIEW](#) Secure Email and Calendaring Service

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#) [Tag Editor](#)

Additional Resources

**Getting Started**  
Read our [documentation](#) or view our [training](#) to learn more about AWS.

**AWS Console Mobile App**  
View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

**AWS Marketplace**  
Find and buy software, launch with 1-Click and pay by the hour.

**AWS Lambda**  
Run your code without managing servers. Try AWS Lambda for free today.

**Service Health**

All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

[Service Health Dashboard](#)

# AMAZON WEB SERVICES

34

AWS Services Edit

Naeff N. Virginia Support

EC2 Dashboard Events Tags Reports Limits Instances Spot Requests Reserved Instances

Images AMIs Bundle Tasks

Elastic Block Store Volumes Snapshots

Network & Security Security Groups Elastic IPs Placement Groups Key Pairs Network Interfaces

Load Balancing Load Balancers

Auto Scaling Launch Configurations Auto Scaling Groups

Launch Instance Connect Actions

Filter by tags and attributes or search by keyword

1 to 4 of 4

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP	Key Name	Monitor
i-3c031cee	i-3c031cee	m3.xlarge	us-east-1d	running	2/2 checks ...	None	ec2-52-7-232-239.com...	52.7.232.239	ga	disabled
i-3d031cef	i-3d031cef	m3.xlarge	us-east-1d	running	2/2 checks ...	None	ec2-52-20-51-153.com...	52.20.51.153	ga	disabled
i-9b021d49	i-9b021d49	m3.xlarge	us-east-1d	running	2/2 checks ...	None	ec2-52-22-146-230.co...	52.22.146.230	ga	disabled
i-9cdcf236	i-9cdcf236	t2.micro	us-east-1c	running	2/2 checks ...	None	ec2-52-21-176-167.co...	52.21.176.167	ga	disabled

Select an instance above

Instances

The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with navigation links for EC2 Dashboard, Events, Tags, Reports, Limits, Instances, Images, AMIs, Bundle Tasks, Elastic Block Store, Volumes, Snapshots, Network & Security, Security Groups, Elastic IPs, Placement Groups, Key Pairs, Network Interfaces, Load Balancing, and Auto Scaling.

The main area displays a table of instances with columns: Name, Instance ID, Instance Type, and Availability. There are four instances listed:

Name	Instance ID	Instance Type	Availability
i-3c031cee	i-3c031cee	m3.xlarge	us-east-1d
i-3d031cef	i-3d031cef	m3.xlarge	us-east-1d
i-9b021d49	i-9b021d49	m3.xlarge	us-east-1d
i-9cdcf236	i-9cdcf236	t2.micro	us-east-1c

A context menu is open over the first instance (i-3c031cee), with the 'Connect' option circled in green. Another green circle highlights the 'Connect' button in the top navigation bar. A large green box highlights the 'Connect' section of the interface, which includes options like 'Get Windows Password', 'Launch More Like This', 'Instance State', 'Instance Settings', 'Image', 'Networking', and 'CloudWatch Monitoring'.

At the bottom, a message says "Select an instance above".

AWS Services Edit

EC2 Dashboard Events Tags Reports Limits

INSTANCES Instances Spot Requests Reserved Instances

IMAGES AMIs Bundle Tasks

ELASTIC BLOCK STORE Volumes Snapshots

NETWORK & SECURITY Security Groups Elastic IPs Placement Groups Key Pairs Network Interfaces

LOAD BALANCING Load Balancers

AUTO SCALING Launch Configurations Auto Scaling Groups

Launch Instance Connect

Filter by tags and attributes

Name	Instance
i-3c031c	
i-3d031c	
i-9b021d	
i-9ddcf23	

## Connect To Your Instance

I would like to connect with  A standalone SSH client  A Java SSH Client directly from my browser (Java required)

To access your instance:

1. Open an SSH client. (find out how to [connect using PuTTY](#))
2. Locate your private key file (ga.pem). The wizard automatically detects the key you used to launch the instance.
3. Your key must not be publicly viewable for SSH to work. Use this command if needed:  
`chmod 400 ga.pem`
4. Connect to your instance using its Public IP:  
**52.21.176.167**

Example:

```
ssh -i "ga.pem" ec2-user@52.21.176.167
```

Please note that in most cases the username above will be correct, however please ensure that you read your AMI usage instructions to ensure that the AMI owner has not changed the default AMI username.

If you need any assistance connecting to your instance, please see our [connection documentation](#).

Select an instance above

**Close**

Naeff N. Virginia Support

IP Key Name Monitor

IP	Key Name	Monitor
2.239	ga	disabled
1.153	ga	disabled
46.230	ga	disabled
76.167	ga	disabled

- ▶ *Log into your instance:*

```
$ ssh -i ~/.aws/ga.pem ec2-user@52.21.176.167
```

- ▶ *Log into your instance:*

```
$ ssh -i ~/.aws/ga.pem ec2-user@52.21.176.167
```

- ▶ *Log out of your instance:*

```
$ exit
```

The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with navigation links like EC2 Dashboard, Events, Tags, Reports, Limits, Instances, Spot Requests, Reserved Instances, Images, AMIs, Bundle Tasks, Elastic Block Store, Volumes, Snapshots, Network & Security, Security Groups, Elastic IPs, Placement Groups, Key Pairs, and Network Interfaces. The main area displays a table of four EC2 instances. The fourth instance, with the ID i-9dcdf236, has a context menu open over it. The menu includes options like Connect, Get Windows Password, Launch More Like This, Instance State (with Start, Stop, and Terminate), Instance Settings, Image, Networking, and CloudWatch Monitoring. The 'Terminate' option under Instance State is highlighted with a green circle. The entire context menu is also circled in green. To the right of the table, there are navigation icons and a message indicating '1 to 4 of 4' instances.

Name	Instance ID	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP	Key Name	Monitor
i-3c031ce...	i-3c031ce...	us-east-1c	running	2/2 checks ...	None	ec2-52-7-232-239.com...	52.7.232.239	ga	disabled
i-3d031cef...	i-3d031cef...	us-east-1c	running	2/2 checks ...	None	ec2-52-20-51-153.com...	52.20.51.153	ga	disabled
i-9b021d49...	i-9b021d49...	us-east-1c	running	2/2 checks ...	None	ec2-52-22-146-230.co...	52.22.146.230	ga	disabled
i-9dcdf236...	i-9dcdf236...	us-east-1c	running	2/2 checks ...	None	ec2-52-21-176-167.co...	52.21.176.167	ga	disabled

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

---

INTRO TO DATA SCIENCE

---

# LAUNCHING A SPARK CLUSTER

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

41

AWS Services Edit ▾

Naeff ▾ N. Virginia ▾ Support ▾

### Amazon Web Services

Compute

- EC2** Virtual Servers in the Cloud
- Lambda** Run Code in Response to Events
- EC2 Container Service** Run and Manage Docker Containers

Storage & Content Delivery

- S3** Scalable Storage in the Cloud
- Elastic File System** PREVIEW Fully Managed File System for EC2
- Storage Gateway** Integrates On-Premises IT Environments with Cloud Storage
- Glacier** Archive Storage in the Cloud
- CloudFront** Global Content Delivery Network

Database

- RDS** MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DYNAMODB** Predictable and Scalable NoSQL Data Store
- ElastiCache** In-Memory Cache
- Redshift** Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC** Isolated Cloud Resources
- Direct Connect** Dedicated Network Connection to AWS
- Route 53** Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service** Managed Directories in the Cloud
- Identity & Access Management** Access Control and Key Management
- Trusted Advisor** AWS Cloud Optimization Expert
- CloudTrail** User Activity and Change Tracking
- Config** Resource Configurations and Inventory
- CloudWatch** Resource and Application Monitoring
- Service Catalog** Personalized Catalog of AWS Resources

Deployment & Management

- Elastic Beanstalk** AWS Application Container
- OpsWorks** DevOps Application Management Service
- CloudFormation** Templated AWS Resource Creation
- CodeDeploy** Automated Deployments
- CodeCommit** Managed Git Repositories
- CodePipeline** Continuous Delivery

Analytics

- EMR** Managed Hadoop Framework

Machine Learning

Build Smart Applications Quickly and Easily

Application Services

- SQS** Message Queue Service
- SWF** Workflow Service for Coordinating Application Components
- AppStream** Low Latency Application Streaming
- Elastic Transcoder** Easy-to-use Scalable Media Transcoding
- SES** Email Sending Service
- CloudSearch** Managed Search Service
- API Gateway** Build, Deploy and Manage APIs

Mobile Services

- Cognito** User Identity and App Data Synchronization
- Device Farm** Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics** Collect, View and Export App Analytics
- SNS** Push Notification Service

Enterprise Applications

- WorkSpaces** Desktops in the Cloud
- WorkDocs** Secure Enterprise Storage and Sharing Service
- WorkMail** PREVIEW Secure Email and Calendaring Service

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

Create a Group Tag Editor

### Additional Resources

Getting Started

Read our [documentation](#) or view our [training](#) to learn more about AWS.

AWS Console Mobile App

View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

AWS Marketplace

Find and buy software, launch with 1-Click and pay by the hour.

AWS Lambda

Run your code without managing servers. Try AWS Lambda for free today.

Service Health

All services operating normally.

Updated: Aug 25 2015 15:57:00 GMT-0400

[Service Health Dashboard](#)

# AMAZON WEB SERVICES

## Quick cluster configuration [Go to advanced options](#)

Cluster name

Logging  Enable

Copy the cluster's log files automatically to S3.

S3 folder

s3://<bucket-name>/<folder>/

Launch mode  Cluster

Step execution

With Cluster, EMR creates a cluster with a set of specified applications. With Step execution, EMR will create a cluster, execute added steps and terminate when done.

Cluster name

## Software configuration

Vendor  Amazon  MapR

Release

Release

Applications  All Applications: Hadoop 2.6.0, Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, and Spark 1.4.1

Core Hadoop: Hadoop 2.6.0, Hive 1.0.0, and Pig 0.14.0

Spark: Spark 1.4.1 on Hadoop 2.6.0 YARN

All Applications: Hadoop 2.6.0, Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, and Spark 1.4.1

## Hardware configuration

Instance type

Number of instances  (1 master and 2 core nodes)

EC2 key pair

## Security and access

EC2 key pair

Use an existing EC2 key pair to SSH into the master node of the EMR cluster. [Learn how to create an EC2 key pair.](#)

Permissions  Default

[View EMR role policy](#)

[View EC2 instance profile](#)

Custom

IAM roles grant EMR and your cluster's EC2 instances access to AWS services. If the roles don't exist, they are created for you using AWS managed policies. [Learn more](#)

Select custom IAM roles to tailor permissions for your cluster. [Learn more](#)

[Cancel](#) [Create cluster](#)

# AMAZON WEB SERVICES

Quick cluster configuration [Go to advanced options](#)

Cluster name  Logging  Enable Copy the cluster's log files automatically to S3.  
 S3 folder  Launch mode  Cluster With Cluster, EMR creates a cluster with a set of specified applications. With Step execution, EMR will create a cluster, execute added steps and terminate when done.  
 Step execution

#### Software configuration

Vendor  Amazon  MapR  
 Release   
 Applications  All Applications: Hadoop 2.6.0, Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, and Spark 1.4.1  
 Core Hadoop: Hadoop 2.6.0, Hive 1.0.0, and Pig 0.14.0  
 Spark: Spark 1.4.1 on Hadoop 2.6.0 YARN

#### Hardware configuration

Instance type   
 Number of instances  (1 master and 2 core nodes)

#### Security and access

EC2 key pair  Use an existing EC2 key pair to SSH into the master node of the EMR cluster. [Learn how to create an EC2 key pair.](#)  
 Permissions  Default  
[View EMR role policy](#) [View EC2 instance profile](#)  
 Custom Select custom IAM roles to tailor permissions for your cluster. [Learn more](#)

[Cancel](#) [Create cluster](#)

## Nota bene: Billing

*Note that larger instances are billed as multiple “normalized instances”.*

Instance Size	Normalization Factor
Small	1
Medium	2
Large	4
Xlarge	8
2xlarge	16
4xlarge	32
8xlarge	64

Quick cluster configuration [Go to advanced options](#)

Cluster name  Logging  Enable Copy the cluster's log files automatically to S3.  
 S3 folder  Launch mode  Cluster With Cluster, EMR creates a cluster with a set of specified applications. With Step execution, EMR will create a cluster, execute added steps and terminate when done.  
 Step execution

#### Software configuration

Vendor  Amazon  MapR  
 Release   
 Applications  All Applications: Hadoop 2.6.0, Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, and Spark 1.4.1  
 Core Hadoop: Hadoop 2.6.0, Hive 1.0.0, and Pig 0.14.0  
 Spark: Spark 1.4.1 on Hadoop 2.6.0 YARN

#### Hardware configuration

Instance type

Number of instances  (1 master and 2 core nodes)

#### Security and access

EC2 key pair

Use an existing EC2 key pair to SSH into the master node of the EMR cluster. [Learn how to create an EC2 key pair.](#)

Permissions  Default

[View EMR role policy](#)  
[View EC2 instance profile](#)

Custom

[Cancel](#) [Create cluster](#)

## Nota bene: Billing

*Note that larger instances are billed as multiple “normalized instances”. Also note that partial hours are billed as a full hour.*

Instance Size	Normalization Factor	
	Elapsed time	Normalized instance hours
Small		1
Medium		
Large	5 hours, 7 minutes	144
Xlarge	32 minutes	24
2xlarge	2 days, 5 hours	1296
4xlarge		
8xlarge	20 minutes	24

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

Elastic MapReduce Cluster List Cluster Details EMR Help

Add step Resize Clone Terminate

Cluster: ga-cluster Starting Configuring cluster software C

**Connections:** Enable Web Connection – Resource Manager ... (View All)

**Master public DNS:** ec2-52-22-32-35.compute-1.amazonaws.com SSH

**Tags:** -- View All / Edit

Summary	Configuration Details	Network and Hardware	Security and Access
ID: j-2G46O8FKD36WD Creation date: 2015-08-25 16:45 (UTC-4) Elapsed time: 4 minutes Auto-terminate: No Termination Off Change protection:	Release label: emr-4.0.0 Hadoop Amazon 2.6.0 distribution: Applications: Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, Spark 1.4.1 Log URI: s3://aws-logs-430081998647-us-east-1/elasticmapreduce/  EMRFS Disabled consistent view:	Availability zone: us-east-1e Subnet ID: subnet-2623421c Master: Bootstrapping 1 m3.xlarge Core: Provisioning 2 m3.xlarge Task: --	Key name: ga EC2 instance EMR_EC2_DefaultRole profile: EMR role: EMR_DefaultRole Visible to all All Change users: Security groups sg-ee631389 for Master: (ElasticMapReduce-master) Security groups sg-ed63138a for Core & Task: (ElasticMapReduce-slave)

- ▶ Monitoring
- ▶ Hardware
- ▶ Steps
- ▶ Bootstrap Actions

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

Elastic MapReduce Cluster List Cluster Details EMR Help

Add step Resize Change Template

Cluster: ga-cluster Starting Configuring cluster software C

**Connections:**  
Master public DNS: ec2-52-22-32-35.compute-1.amazonaws.com SSH

**Tags:** -- View All / Edit

Summary	Configuration Details	Network and Hardware	Security and Access
ID: j-2G46O8FKD36WD Creation date: 2015-08-25 16:45 (UTC-4) Elapsed time: 4 minutes Auto-terminate: No Termination Off Change protection:	Release label: emr-4.0.0 Hadoop Amazon 2.6.0 distribution: Applications: Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, Spark 1.4.1 Log URI: s3://aws-logs-430081998647-us-east-1/elasticmapreduce/  EMRFS Disabled consistent view:	Availability zone: us-east-1e Subnet ID: subnet-2623421c Master: Bootstrapping 1 m3.xlarge Core: Provisioning 2 m3.xlarge Task: --	Key name: ga EC2 instance EMR_EC2_DefaultRole profile: EMR role: EMR_DefaultRole Visible to all All Change users: Security groups sg-ee631389 for Master: (ElasticMapReduce-master) Security groups sg-ed63138a for Core & Task: (ElasticMapReduce-slave)

- ▶ Monitoring
- ▶ Hardware
- ▶ Steps
- ▶ Bootstrap Actions

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

Elastic MapReduce Cluster List Cluster Details EMR Help

Add step Resize Clone Terminate

Cluster: ga-cluster Waiting after step completed C

**Connections:**  
Master public DNS: ec2-52-22-32-35.compute-1.amazonaws.com SSH

**Tags:** -- View All / Edit

Summary	Configuration Details	Network and Hardware	Security and Access
ID: j-2G46O8FKD36WD Creation date: 2015-08-25 16:45 (UTC-4) Elapsed time: 4 minutes Auto-terminate: No Termination Off Change protection:	Release label: emr-4.0.0 Hadoop Amazon 2.6.0 distribution: Applications: Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, Spark 1.4.1 Log URI: s3://aws-logs-430081998647-us-east-1/elasticmapreduce/  EMRFS Disabled consistent view:	Availability zone: us-east-1e Subnet ID: subnet-2623421c Master: Bootstrapping 1 m3.xlarge Core: Provisioning 2 m3.xlarge Task: --	Key name: ga EC2 instance EMR_EC2_DefaultRole profile: EMR role: EMR_DefaultRole Visible to all All Change users: Security groups sg-ee631389 for Master: (ElasticMapReduce-master) Security groups sg-ed63138a for Core & Task: (ElasticMapReduce-slave)

- ▶ Monitoring
- ▶ Hardware
- ▶ Steps
- ▶ Bootstrap Actions

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

Elastic MapReduce Cluster List Cluster Details EMR Help

Add step Resize Clone Terminate

Cluster: ga-cluster Starting Configuring cluster software C

**Master public DNS:** ec2-52-22-32-35.compute-1.amazonaws.com [SSH](#)

Summary	Configuration Details	Network and Hardware	Security and Access
ID: j-2G46O8FKD36WD Creation date: 2015-08-25 16:45 (UTC-4) Elapsed time: 4 minutes Auto-terminate: No Termination Off <a href="#">Change</a> protection:	Release label: emr-4.0.0 Hadoop Amazon 2.6.0 distribution: Applications: Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, Spark 1.4.1 Log URI: s3://aws-logs-430081998647-us-east-1/elasticmapreduce/  EMRFS Disabled consistent view:	Availability zone: us-east-1e Subnet ID: subnet-2623421c Master: <b>Bootstrapping</b> 1 m3.xlarge Core: <b>Provisioning</b> 2 m3.xlarge Task: --	Key name: ga EC2 instance <a href="#">EMR_EC2_DefaultRole</a> profile: EMR role: <a href="#">EMR_DefaultRole</a> Visible to all All <a href="#">Change</a> users: Security groups <a href="#">sg-ee631389</a> for Master: (ElasticMapReduce-master) Security groups <a href="#">sg-ed63138a</a> for Core & Task: (ElasticMapReduce-slave)

- ▶ Monitoring
- ▶ Hardware
- ▶ Steps
- ▶ Bootstrap Actions

# AMAZON WEB SERVICES

AWS Services Edit Naeff N. Virginia Support

Elastic MapReduce Cluster List Cluster Details EMR Help

Add step Resize Clone Terminate

Cluster: ga-cluster Starting Configuring cluster software C

**Master public DNS:** ec2-52-22-32-35.compute-1.amazonaws.com **SSH**

**Summary**  
 ID: j-2G46O8FKD36WD  
 Creation date: 2015-08-25 16:45 (UTC-4)  
 Elapsed time: 4 minutes  
 Auto-terminate: No  
 Termination Off Change protection:

**Configuration Details**  
 Release label: emr-4.0.0  
 Hadoop Amazon 2.6.0 distribution:  
 Applications: Hive 1.0.0, Mahout 0.10.0, Pig 0.14.0, Spark 1.4.1  
 Log URI: s3://aws-logs-430081998647-us-east-1/elasticmapreduce/

**Network and Hardware**  
 Availability zone: us-east-1e  
 Subnet ID: subnet-2623421c  
 Master: Bootstrapping 1 m3.xlarge  
 Core: Provisioning 2 m3.xlarge  
 Task: --

**Security and Access**  
 Key name: ga  
 EC2 instance EMR\_EC2\_DefaultRole profile:  
 EMR role: EMR\_DefaultRole  
 Visible to all All Change users:  
 Security groups sg-ee631389 for Master: (ElasticMapReduce-master)  
 Security groups sg-ed63138a for Core & Task: (ElasticMapReduce-slave)

▶ Monitoring

▶ Hardware

▶ Steps

▶ Bootstrap Actions

The screenshot shows the AWS Elastic MapReduce Cluster Details page. A modal dialog titled "SSH" is open, providing instructions on how to connect to the master node using SSH. The dialog includes a "Learn more" link and tabs for "Windows" and "Mac / Linux". It lists three steps: opening a terminal window, running an SSH command, and dismissing a security warning. The "Mac / Linux" tab is selected.

AWS Services Edit Naeff N. Virginia Support

Elastic MapReduce Cluster List Cluster Details EMR Help

Add step Cluster: ga-d

Connections: Master public Tags: Summary

Creation d Elapsed ti Auto-termi Terminal protect

Monitoring Hardware Steps Bootstrap Actions

**SSH**

Connect to the Master Node Using SSH

You can connect to the Amazon EMR master node using SSH to run interactive queries, examine log files, submit Linux commands, and so on.

[Learn more](#)

Windows Mac / Linux

1. Open a terminal window. On Mac OS X, choose Applications > Utilities > Terminal. On other Linux distributions, terminal is typically found at Applications > Accessories > Terminal.
2. To establish a connection to the master node, type the following command. Replace ~/ga.pem with the location and filename of the private key file (.pem) used to launch the cluster.

```
ssh hadoop@ec2-52-22-32-35.compute-1.amazonaws.com -i ~/ga.pem
```
3. Type yes to dismiss the security warning.

**Close**

## *Log into your master node*

```
$ ssh -i ~/.aws/ga.pem hadoop@ec2-52-22-146-230.compute-1.amazonaws.com
```

---

*Log into your master node*

```
$ ssh -i ~/.aws/ga.pem hadoop@ec2-52-22-146-230.compute-1.amazonaws.com
```

```
The authenticity of host 'ec2-52-22-146-230.compute-1.amazonaws.com (52.22.146.230)' can't be established.
```

```
RSA key fingerprint is 72:6c:a8:8f:01:32:eb:90:33:70:97:24:b1:b2:39:0d.
```

```
Are you sure you want to continue connecting (yes/no)?
```

## *Log into your master node*

```
$ ssh -i ~/.aws/ga.pem hadoop@ec2-52-22-146-230.compute-1.amazonaws.com
```

The authenticity of host 'ec2-52-22-146-230.compute-1.amazonaws.com (52.22.146.230)' can't be established.

RSA key fingerprint is 72:6c:a8:8f:01:32:eb:90:33:70:97:24:b1:b2:39:0d.

Are you sure you want to continue connecting (yes/no)? yes

## *Log into your master node*

```
$ ssh -i ~/.aws/ga.pem hadoop@ec2-52-22-146-230.compute-1.amazonaws.com
```

The authenticity of host 'ec2-52-22-146-230.compute-1.amazonaws.com (52.22.146.230)' can't be established.

RSA key fingerprint is 72:6c:a8:8f:01:32:eb:90:33:70:97:24:b1:b2:39:0d.

Are you sure you want to continue connecting (yes/no)? yes

[...]

```
[hadoop@ip-172-31-22-179 ~]$
```

*Please refer to the repo for instructions for next steps*

AWS Services
Naeff ▾ N. Virginia ▾ Support ▾

---

**Compute**

- EC2** Virtual Servers in the Cloud
- Lambda** Run Code in Response to Events
- EC2 Container Service** Run and Manage Docker Containers

**Storage & Content Delivery**

- S3** Scalable Storage in the Cloud
- Elastic File System** PREVIEW Fully Managed File System for EC2
- Storage Gateway** Integrates On-Premises IT Environments with Cloud Storage
- Glacier** Archive Storage in the Cloud
- CloudFront** Global Content Delivery Network

**Database**

- RDS** MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DynamoDB** Predictable and Scalable NoSQL Data Store
- ElastiCache** In-Memory Cache
- Redshift** Managed Petabyte-Scale Data Warehouse Service

**Networking**

- VPC** Isolated Cloud Resources
- Direct Connect** Dedicated Network Connection to AWS
- Route 53** Scalable DNS and Domain Name Registration

**Administration & Security**

- Directory Service** Managed Directories in the Cloud
- Identity & Access Management** Access Control and Key Management
- Trusted Advisor** AWS Cloud Optimization Expert
- CloudTrail** User Activity and Change Tracking
- Config** Resource Configurations and Inventory
- CloudWatch** Resource and Application Monitoring
- Service Catalog** Personalized Catalog of AWS Resources

**Deployment & Management**

- Elastic Beanstalk** AWS Application Container
- OpsWorks** DevOps Application Management Service
- CloudFormation** Templated AWS Resource Creation
- CodeDeploy** Automated Deployments
- CodeCommit** Managed Git Repositories
- CodePipeline** Continuous Delivery

**Analytics**

- EMR** Managed Hadoop Framework

**Machine Learning**  
Build Smart Applications Quickly and Easily

**Application Services**

- SQS** Message Queue Service
- SWF** Workflow Service for Coordinating Application Components
- AppStream** Low Latency Application Streaming
- Elastic Transcoder** Easy-to-use Scalable Media Transcoding
- SES** Email Sending Service
- CloudSearch** Managed Search Service
- API Gateway** Build, Deploy and Manage APIs

**Mobile Services**

- Cognito** User Identity and App Data Synchronization
- Device Farm** Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics** Collect, View and Export App Analytics
- SNS** Push Notification Service

**Enterprise Applications**

- WorkSpaces** Desktops in the Cloud

**Resource Groups**

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#)
[Tag Editor](#)

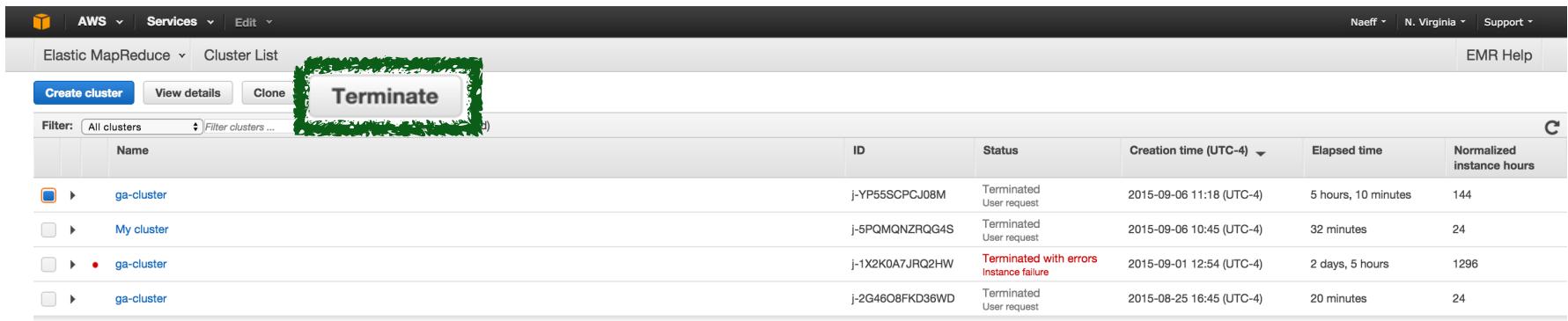
**Additional Resources**

**Getting Started**  
Read our [documentation](#) or view our [training](#) to learn more about AWS.

**AWS Console Mobile App**  
View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

**AWS Marketplace**  
Find and buy software, launch with 1-Click and pay by the hour.

Make sure to kill your instances to prevent adding minutes and \$\$ to your bill



The screenshot shows the AWS Elastic MapReduce Cluster List interface. At the top, there are navigation links for AWS Services and Edit, along with user information Naeff, N. Virginia, and Support. Below the header, there are buttons for Create cluster, View details, and Clone. A 'Terminate' button is highlighted with a green box. A filter bar allows filtering by cluster name. The main table lists four clusters:

Name	ID	Status	Creation time (UTC-4)	Elapsed time	Normalized instance hours
ga-cluster	j-YP55SCPCJ08M	Terminated User request	2015-09-06 11:18 (UTC-4)	5 hours, 10 minutes	144
My cluster	j-5PQMQNZRQG4S	Terminated User request	2015-09-06 10:45 (UTC-4)	32 minutes	24
ga-cluster	j-1X2K0A7JQRQ2HW	Terminated with errors Instance failure	2015-09-01 12:54 (UTC-4)	2 days, 5 hours	1296
ga-cluster	j-2G46O8FKD36WD	Terminated User request	2015-08-25 16:45 (UTC-4)	20 minutes	24

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

# AMAZON WEB SERVICES

## CONSOLE.AWS.AMAZON.COM

58

AWS Services

Compute

**EC2** Virtual Servers in the Cloud

Run and Manage Docker Containers

Storage & Content Delivery

- S3 Scalable Storage in the Cloud
- Elastic File System PREVIEW Fully Managed File System for EC2
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage
- Glacier Archive Storage in the Cloud
- CloudFront Global Content Delivery Network

Database

- RDS MySQL, Postgres, Oracle, SQL Server, and Amazon Aurora
- DynamoDB Predictable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Scalable DNS and Domain Name Registration

Administration & Security

- Directory Service Managed Directories in the Cloud
- Identity & Access Management Access Control and Key Management
- Trusted Advisor AWS Cloud Optimization Expert
- CloudTrail User Activity and Change Tracking
- Config Resource Configurations and Inventory
- CloudWatch Resource and Application Monitoring
- Service Catalog Personalized Catalog of AWS Resources

Deployment & Management

- Elastic Beanstalk AWS Application Container
- OpsWorks DevOps Application Management Service
- CloudFormation Templated AWS Resource Creation
- CodeDeploy Automated Deployments
- CodeCommit Managed Git Repositories
- CodePipeline Continuous Delivery

Analytics

- EMR Managed Hadoop Framework
- Kinesis Real-time Processing of Streaming Big Data
- Data Pipeline Orchestration for Data-Driven Workflows
- Machine Learning Build Smart Applications Quickly and Easily

Application Services

- SQS Message Queue Service
- SWF Workflow Service for Coordinating Application Components
- AppStream Low Latency Application Streaming
- Elastic Transcoder Easy-to-use Scalable Media Transcoding
- SES Email Sending Service
- CloudSearch Managed Search Service
- API Gateway Build, Deploy and Manage APIs

Mobile Services

- Cognito User Identity and App Data Synchronization
- Device Farm Test Android, Fire OS, and iOS apps on real devices in the Cloud
- Mobile Analytics Collect, View and Export App Analytics
- SNS Push Notification Service

Enterprise Applications

- WorkSpaces Desktops in the Cloud
- WorkD Secure
- Work Secure L

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

Create a Group Tag Editor

Additional Resources

Getting Started

Read our [documentation](#) or view our [training](#) to learn more about AWS.

AWS Console Mobile App

View your resources on the go with our AWS Console mobile app, available from [Amazon Appstore](#), [Google Play](#), or [iTunes](#).

AWS Marketplace

Find and buy software, launch with 1-Click and pay by the hour.

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with various navigation options: EC2 Dashboard, Events, Tags, Reports, Limits, Instances (which is circled in green), Spot Requests, Reserved Instances, IMAGES, AMIs, Bundle Tasks, ELASTIC BLOCK STORE, Volumes, Snapshots, and NETWORK & SECURITY. The main area displays a table of four EC2 instances. The fourth instance, with ID i-9dcdf236, has a context menu open over it. The menu includes options like Connect, Get Windows Password, Launch More Like This, Instance State (with Start, Stop, and Terminate options), Instance Settings, Image, Networking, and CloudWatch Monitoring. The 'Terminate' option is highlighted with a green circle. The table columns include Name, Instance ID, Availability Zone, Instance State, Status Checks, Alarm Status, Public DNS, Public IP, Key Name, and Monitor. The fourth instance is listed as running.

Name	Instance ID	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP	Key Name	Monitor
i-3c031ce...	i-3c031ce...	us-east-1c	running	2/2 checks ...	None	ec2-52-7-232-239.com...	52.7.232.239	ga	disabled
i-3d031cef...	i-3d031cef...	us-east-1c	running	2/2 checks ...	None	ec2-52-20-51-153.com...	52.20.51.153	ga	disabled
i-9b021d49...	i-9b021d49...	us-east-1c	running	2/2 checks ...	None	ec2-52-22-146-230.co...	52.22.146.230	ga	disabled
i-9dcdf236...	i-9dcdf236...	us-east-1c	running	2/2 checks ...	None	ec2-52-21-176-167.co...	52.21.176.167	ga	disabled

*Make sure to kill your instances to prevent adding minutes and \$\$ to your bill*

---

**INTRO TO DATA SCIENCE**

---

**DISCUSSION**