



# MDS-535

## Activity - 1

Renjith Madhavan

## Contents

Overview and Task Summary.....	2
AWS User Account Creation .....	3
Creating and setting up an AWS Instance.....	5
AWS Management Console Screenshot .....	8
Installing dos2unix .....	10
Invoking “java” .....	11
Invoking “Hadoop” .....	12
Errors.....	12

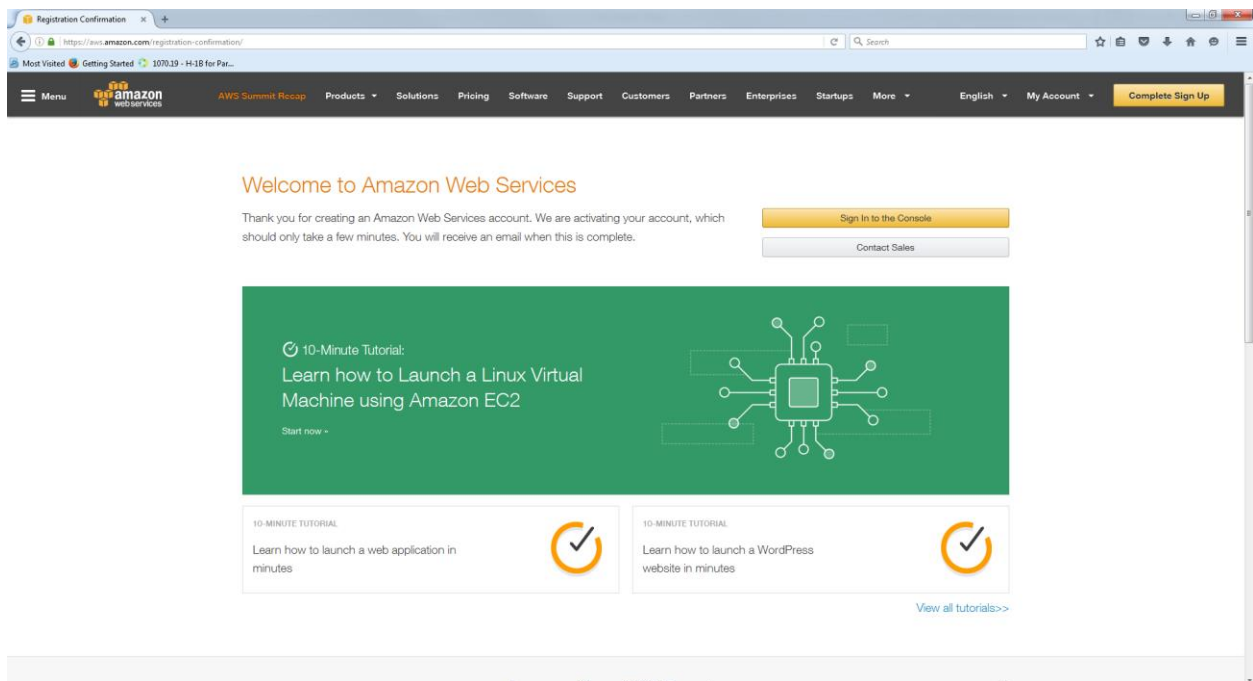
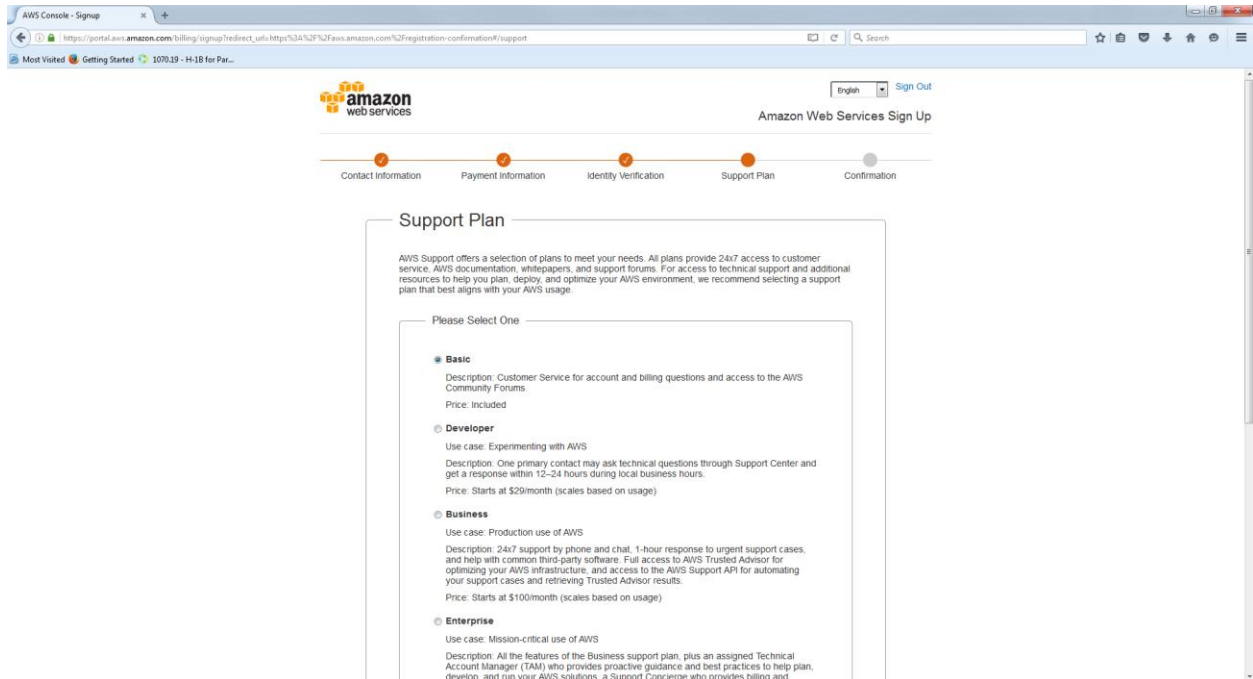
## Overview and Task Summary

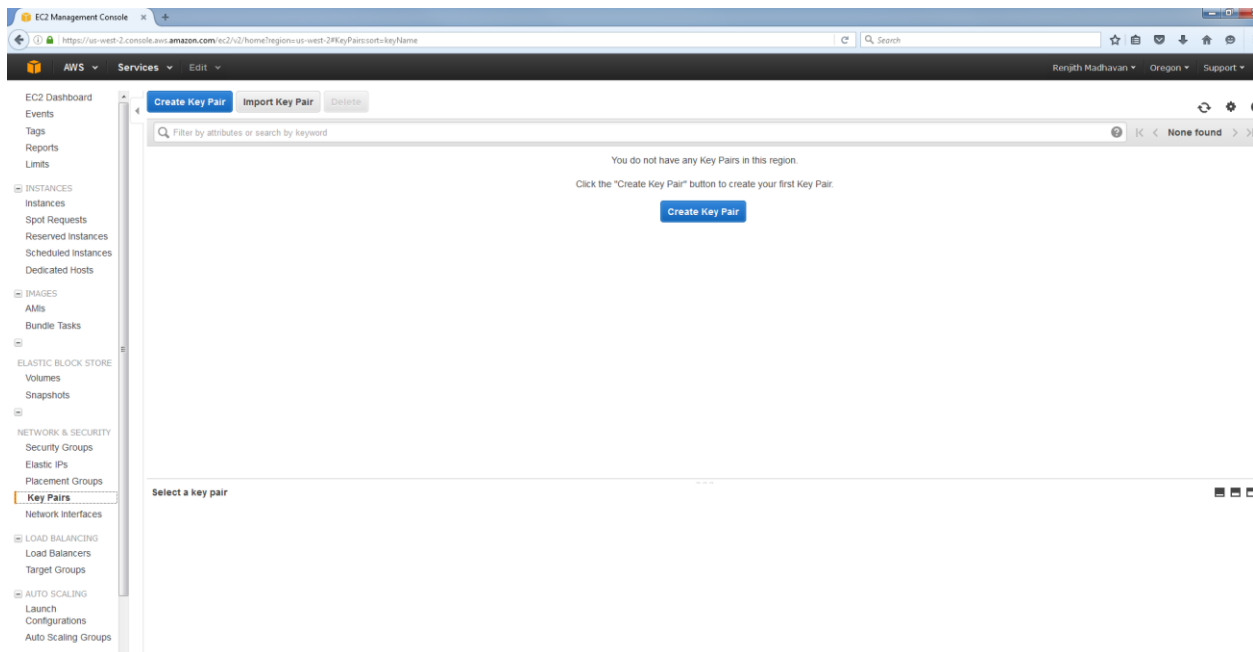
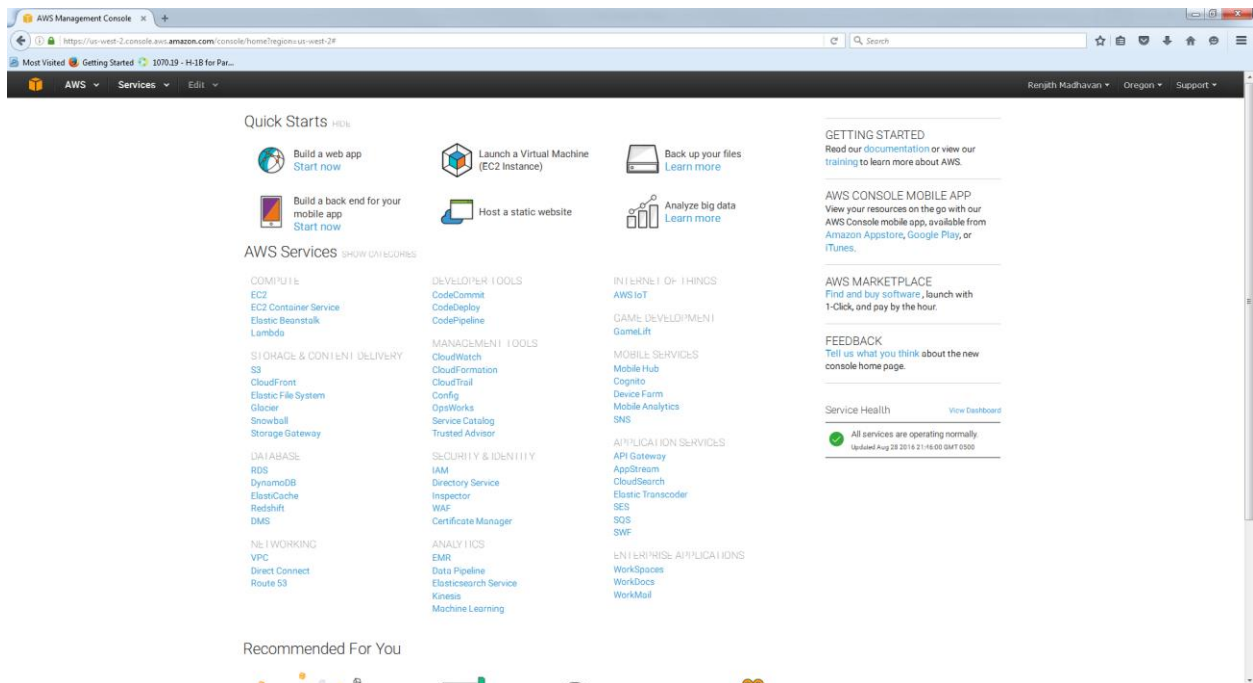
This is my first experience to install and setup an AWS instance for Hadoop. The how to setup document with instruction was very easy to follow with screenshots. I also captured my screenshots during installation and the final console. Once the setup was complete I copied the environment files `env.sh` and `bashScript24.sh` to the AWS instance. Also installed `dos2unix` and changed permissions for the files.

While sourcing the `env.sh`, there were error messages for directory exists but that can be ignored as those directories were already created as part of the `bashScript24.sh`.

Finally I was able to invoke `java` and Hadoop commands without errors confirming that the setup worked as intended.

# AWS User Account Creation





# Creating and setting up an AWS Instance

EC2 Management Console

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

Amazon Linux 2016.03.3 (HVM), SSD Volume Type - ami-712b611

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

Red Hat Enterprise Linux 7.2 (HVM), SSD Volume Type - ami-775e4f16

Red Hat Enterprise Linux version 7.2 (HVM), EBS General Purpose (SSD) Volume Type

Root device type: ebs Virtualization type: hvm

SUSE Linux Enterprise Server 12 SP1 (HVM), SSD Volume Type - ami-d2627db3

SUSE Linux Enterprise Server 12 Service Pack 1 (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

Ubuntu Server 14.04 LTS (HVM), SSD Volume Type - ami-d7320b27

Ubuntu Server 14.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).

Root device type: ebs Virtualization type: hvm

Microsoft Windows Server 2012 R2 Base - ami-1712d877

Microsoft Windows 2012 R2 Standard edition with 64-bit architecture. [English]

Root device type: ebs Virtualization type: hvm

Are you launching a database instance? Try Amazon RDS.

Amazon RDS

Amazon Relational Database Service (RDS) makes it easy to set up, operate, and scale your database on AWS by automating time-consuming database management tasks. With RDS, you can easily deploy Amazon Aurora, MariaDB, MySQL, Oracle, PostgreSQL, and SQL Server databases on AWS. Aurora is a MySQL-compatible, enterprise-class database at 1/10th the cost of commercial databases. [Learn more about RDS](#)

Launch a database using RDS

Microsoft Windows Server 2012 R2 with SQL Server Express - ami-c513d9a5

EC2 Management Console

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 GB memory, EBS only)

	Family	Type	vCPUs	Memory (GiB)	Instance Storage (GiB)	EBS-Optimized Available	Network Performance
<input type="checkbox"/>	General purpose	t2.nano	1	0.5	EBS only	-	Low to Moderate
<input checked="" type="checkbox"/>	General purpose	t2.micro	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
<input type="checkbox"/>	General purpose	m4.10xlarge	40	160	EBS only	Yes	10 Gigabit
<input type="checkbox"/>	General purpose	m3.medium	1	3.75	1 x 4 (SSD)	-	Moderate
<input type="checkbox"/>	General purpose	m3.large	2	7.5	1 x 32 (SSD)	-	Moderate
<input type="checkbox"/>	General purpose	m3.xlarge	4	15	2 x 40 (SSD)	Yes	High

Cancel Previous Review and Launch Next: Configure Instance Details

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 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

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 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](#)

Free tier eligible

**Ubuntu Server 14.04 LTS (HVM), SSD Volume Type - ami-d732f0b7**

Ubuntu Server 14.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).

Root Device Type: ebs Virtualization type: hvm

Instance Type

[Edit instance type](#)

Instance Type	ECUs	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance
t2.micro	Variable	1	1	EBS only	-	Low to Moderate

Security Groups

[Edit security groups](#)

Security group name launch-wizard-1

Description launch-wizard-1 created 2016-08-28T22:16:10.738-05:00

Type	Protocol	Port Range	Source
SSH	TCP	22	0.0.0.0/0

Instance Details

[Edit instance details](#)

Storage

[Edit storage](#)

Tags

[Edit tags](#)

Cancel Previous **Launch**

Launch Status

**Your instances are now launching**

The following instance launches have been initiated: [i-07af6baa0bd3860d5](#) [View launch log](#)

**Get notified of estimated charges**

[Create billing alerts](#) to get an email notification when estimated charges on your AWS bill exceed an amount you define (for example, if you exceed the free usage tier).

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances. Click **View instances** to monitor your instances' status. Once your instances are in the **running** state, you can **connect** to them from the instances screen. [Find out](#) how to connect to your instances.

- Here are some helpful resources to get you started
- How to connect to your Linux instance
  - Learn about AWS Free Usage Tier
  - Amazon EC2 User Guide
  - Amazon EC2 Discussion Forum

While your instances are launching you can also

- Create status check alarms to be notified when these instances fail status checks. (Additional charges may apply)
- Create and attach additional EBS volumes (Additional charges may apply)
- Manage security groups

View instances



# AWS Management Console Screenshot

The screenshot shows the AWS Management Console interface. The left sidebar contains navigation links for EC2 Dashboard, Events, Tags, Reports, Limits, INSTANCES, SPOT REQUESTS, RESERVED INSTANCES, SCHEDULED INSTANCES, DEDICATED HOSTS, 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, Target Groups, AUTO SCALING, Launch, Configurations, Auto Scaling Groups, and COMMANDS. The main content area displays a table of EC2 instances. The instance 'i-07af6baa0bd3860d5' is highlighted, and its details are shown in the 'Description' tab. The details include Instance ID, Instance state, Instance type, Private DNS, Private IPs, Secondary private IPs, VPC ID, Public DNS, Public IP, Elastic IPs, Availability zone, Security groups, Scheduled events, AMI ID, and Platform.

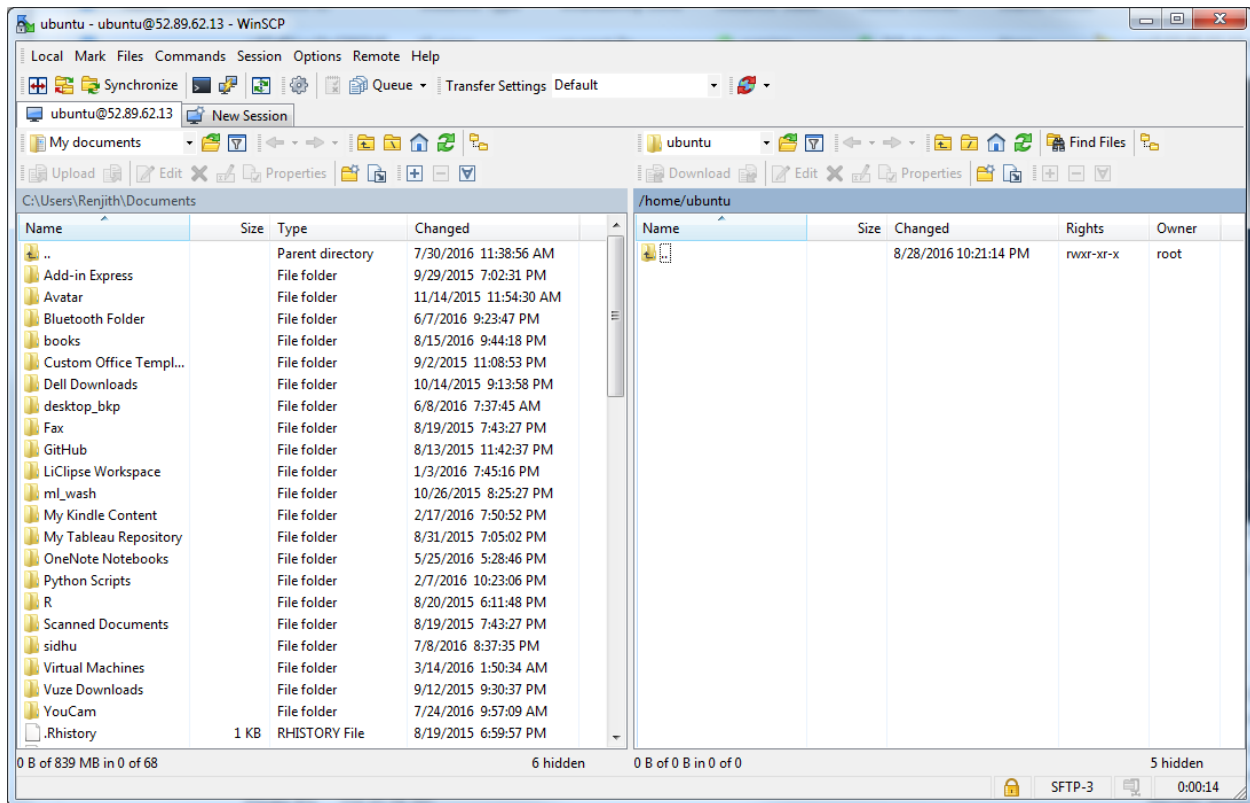
Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS	Public IP	Key Name	Monitoring	Launch Time
i-07af6baa0bd3860d5	i-07af6baa0bd3860d5	t2.micro	us-west-2a	running	Initializing	None	ec2-52-89-62-13.us-we...	52.89.62.13	rdara123	disabled	August 28, 2016 at 10:20:41...

Instance: i-07af6baa0bd3860d5 Public DNS: ec2-52-89-62-13.us-west-2.compute.amazonaws.com

Description	Status Checks	Monitoring	Tags
Instance ID: i-07af6baa0bd3860d5 Instance state: running Instance type: t2.micro Private DNS: ip-172-31-19-191.us-west-2.compute.internal Private IPs: 172.31.19.191 Secondary private IPs: - VPC ID: vpc-a9794bcd Subnet ID: subnet-a7944b8e	Public DNS: ec2-52-89-62-13.us-west-2.compute.amazonaws.com Public IP: 52.89.62.13 Elastic IPs: - Availability zone: us-west-2a Security groups: launch-wizard-1 view rules Scheduled events: No scheduled events AMI ID: ubuntu/images/hvm-ssd/ubuntu-trusty-14.04-amd64-server-20160714 (ami-d7326b7) Platform: -		

The screenshot shows the AWS Management Console interface with a terminal window open. The terminal displays the output of the 'whoami' command, showing the user is 'ubuntu' on the 'ip-172-31-19-191' instance. The terminal also shows the output of the 'uname -a' command, displaying the kernel version and architecture.

```
ubuntu@ip-172-31-19-191:~$ whoami
ubuntu
ubuntu@ip-172-31-19-191:~$ uname -a
Linux ip-172-31-19-191 3.13.0-92-generic #139-Ubuntu SMP Tue Jun 28 20:42:26 UTC
2016 x86_64 x86_64 x86_64 GNU/Linux
ubuntu@ip-172-31-19-191:~$
```



## Installing dos2unix

```
ubuntu@ip-172-31-19-191:~$ sudo apt-get install dos2unix
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  dos2unix
0 upgraded, 1 newly installed, 0 to remove and 0 not upgraded.
Need to get 68.7 kB of archives.
After this operation, 306 kB of additional disk space will be used.
Get:1 http://us-west-2.ec2.archive.ubuntu.com/ubuntu/ trusty/universe dos2unix amd64 6.0.4-1 [68.7 kB]
Fetched 68.7 kB in 0s (82.4 kB/s)
Selecting previously unselected package dos2unix.
(Reading database ... 51172 files and directories currently installed.)
Preparing to unpack .../dos2unix_6.0.4-1_amd64.deb ...
Unpacking dos2unix (6.0.4-1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Setting up dos2unix (6.0.4-1) ...
ubuntu@ip-172-31-19-191:~$ pwd
/home/ubuntu
ubuntu@ip-172-31-19-191:~$ df -h
```

```
ubuntu@ip-172-31-19-191:~$ dos2unix bashScript24.sh
dos2unix: converting file bashScript24.sh to Unix format ...
ubuntu@ip-172-31-19-191:~$ dos2unix env.sh
dos2unix: converting file env.sh to Unix format ...
ubuntu@ip-172-31-19-191:~$ chmod 777 bashScript24.sh
ubuntu@ip-172-31-19-191:~$ chmod 777 env.sh
ubuntu@ip-172-31-19-191:~$
```

## Invoking “java”

```
ubuntu@ip-172-31-31-237: ~$ java
Usage: java [-options] class [args...]
           (to execute a class)
or  java [-options] -jar jarfile [args...]
           (to execute a jar file)
where options include:
-d32          use a 32-bit data model if available
-d64          use a 64-bit data model if available
-server      to select the "server" VM
-zero        to select the "zero" VM
-javm        to select the "javm" VM
-avian       to select the "avian" VM
-dcevm       to select the "dcevm" VM
              The default VM is server.

-cp <class search path of directories and zip/jar files>
-classpath <class search path of directories and zip/jar files>
           A : separated list of directories, JAR archives,
           and ZIP archives to search for class files.
-D<name>=<value>
           set a system property
-verbose[:<class>[:<jni>]]
           enable verbose output
-version      print product version and exit
-version:<value>
           require the specified version to run
-showversion  print product version and continue
-jre-restrict-search | -no-jre-restrict-search
           include/exclude user private JREs in the version search
-? -help      print this help message
-X            print help on non-standard options
-ea[:<packagename>...][:<classname>]
           enable assertions with specified granularity
-da[:<packagename>...][:<classname>]
           disable assertions with specified granularity
-esa | -enablesystemassertions
           enable system assertions
-dsa | -disablesystemassertions
           disable system assertions
-agentlib:<libname>[=<options>]
           load native agent library <libname>, e.g. -agentlib:hprof
           see also, -agentlib:jdwp=help and -agentlib:hprof=help
-agentpath:<pathname>[=<options>]
           load native agent library by full pathname
-javaagent:<jarpath>[=<options>]
           load Java programming language agent, see java.lang.instrument
-splash:<imagepath>
           show splash screen with specified image
```

## Invoking “Hadoop”

```
ubuntu@ip-172-31-31-237:~$
-dsa | -disablesystemassertions
    disable system assertions
-agentlib:<libname>[=<options>]
    load native agent library <libname>, e.g. -agentlib:hprof
    see also, -agentlib:jdwp=help and -agentlib:hprof=help
-agentpath:<pathname>[=<options>]
    load native agent library by full pathname
-javaagent:<jarpath>[=<options>]
    load Java programming language agent, see java.lang.instrument
-splash:<imagepath>
    show splash screen with specified image
See http://www.oracle.com/technetwork/java/javase/documentation/index.html for more details.
ubuntu@ip-172-31-31-237:~$ hadoop
Usage: hadoop [--config confdir] COMMAND
where COMMAND is one of:
namenode -format      format the DFS filesystem
secondarynamenode    run the DFS secondary namenode
namenode              run the DFS namenode
datanode              run a DFS datanode
dfsadmin              run a DFS admin client
mradmin               run a Map-Reduce admin client
fsck                  run a DFS filesystem checking utility
fs                     run a generic filesystem user client
balancer              run a cluster balancing utility
oiv                   apply the offline fsimage viewer to an fsimage
fetchdt               fetch a delegation token from the NameNode
jobtracker            run the MapReduce job Tracker node
pipes                 run a Pipes job
tasktracker           run a MapReduce task Tracker node
historyserver         run job history servers as a standalone daemon
job                   manipulate MapReduce jobs
queue                 get information regarding JobQueues
version               print the version
jar <jar>              run a jar file
distcp <srcurl> <desturl> copy file or directories recursively
distcp2 <srcurl> <desturl> DistCp version 2
archive -archiveName NAME -p <parent path> <src> <dest> create a hadoop archive
classpath             prints the class path needed to get the
                      Hadoop jar and the required libraries
daemonlog              get/set the log level for each daemon
or
CLASSNAME             run the class named CLASSNAME
Most commands print help when invoked w/o parameters.
ubuntu@ip-172-31-31-237:~$ hadoop version
Hadoop 1.2.1
Subversion https://svn.apache.org/repos/asf/hadoop/common/branches/branch-1.2 -r 1503152
Compiled by mattf on Mon Jul 22 15:23:09 PDT 2013
From source with checksum 6923c86528809c4e1ef493bcb413a9a
This command was run using /home/ubuntu/hadoop-1.2.1/hadoop-core-1.2.1.jar
ubuntu@ip-172-31-31-237:~$
```

## Errors

File exists error while sourcing env.sh

Found that these directories are already created part of bashScript24.sh, hence can be ignored.

```
ubuntu@ip-172-31-31-237:~$
ubuntu@ip-172-31-31-237:~$ source env.sh
mkdir: cannot create directory âkmeansâ: File exists
mkdir: cannot create directory â/home/ubuntu/kmeans/reuters-sgmâ: File exists
mkdir: cannot create directory â/home/ubuntu/kmeans/reuters-outâ: File exists
ubuntu@ip-172-31-31-237:~$
ubuntu@ip-172-31-31-237:~$
```