

Project: calculate Pi

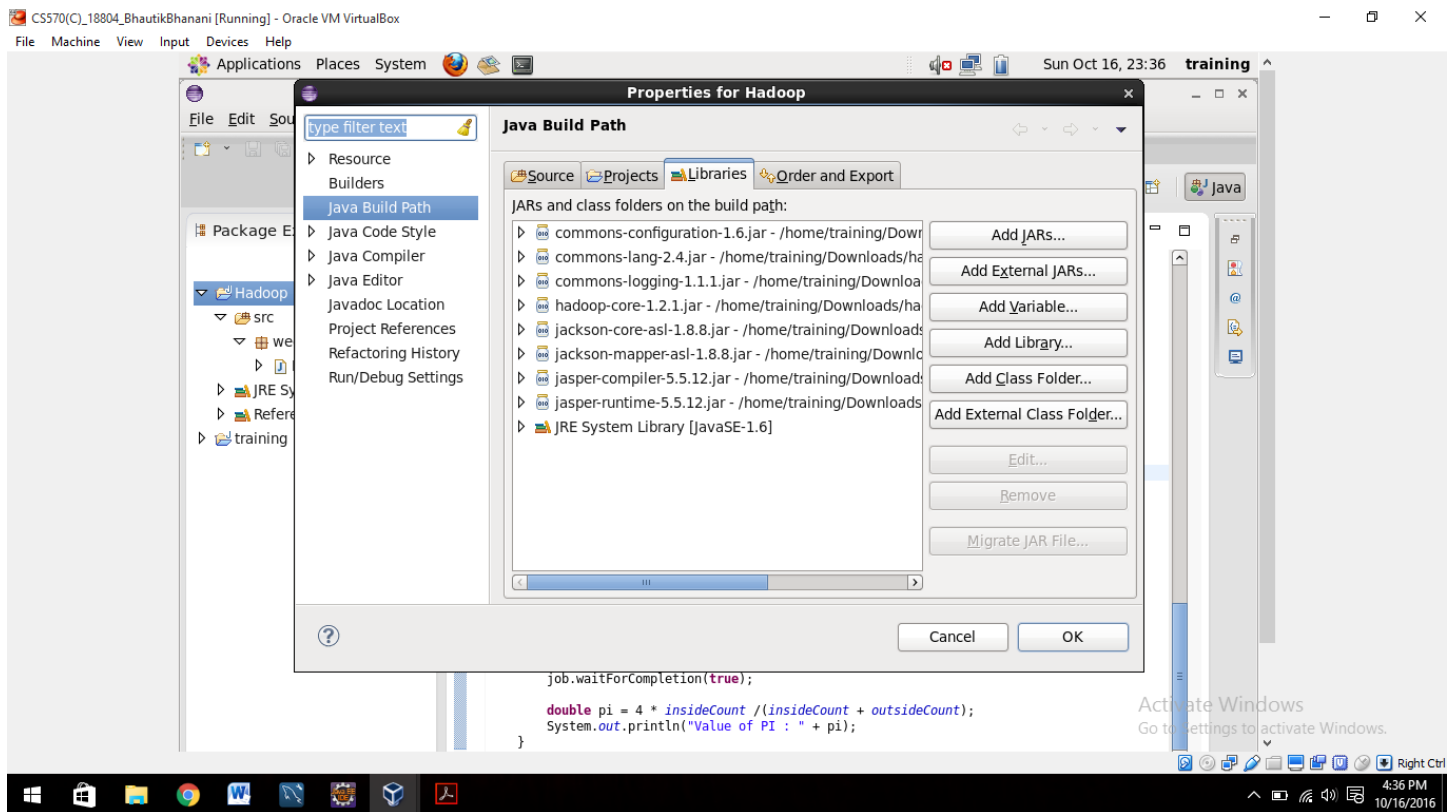
Solution:

Step 1: Open Cloudera and open Eclipse in it. Create one class under Hadoop project and name it 'Pi'.

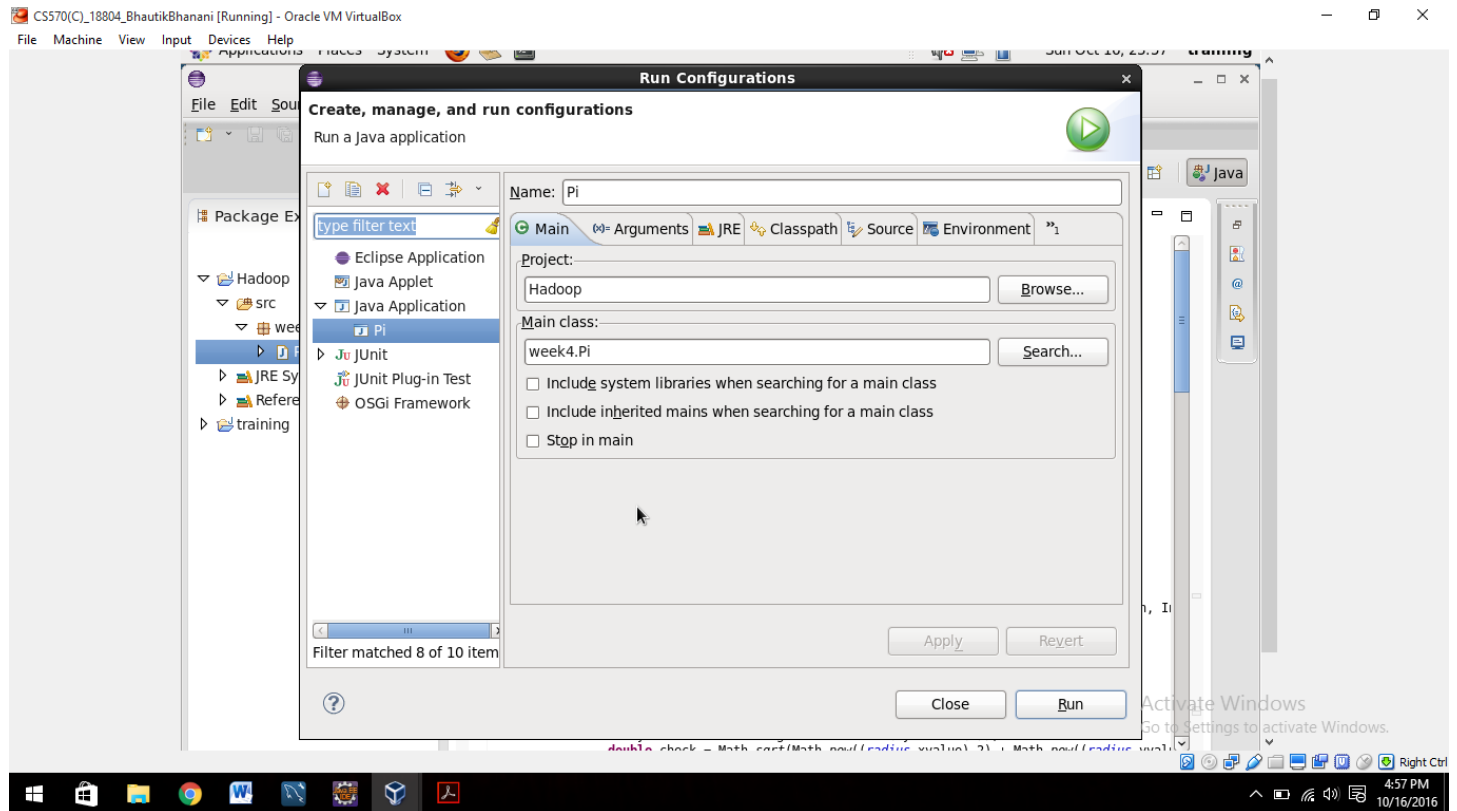
Step 2: Create file Pi.java and write a bellow code to calculate 'Pi' value and create driver function and configure job.

```
// copy from repository
```

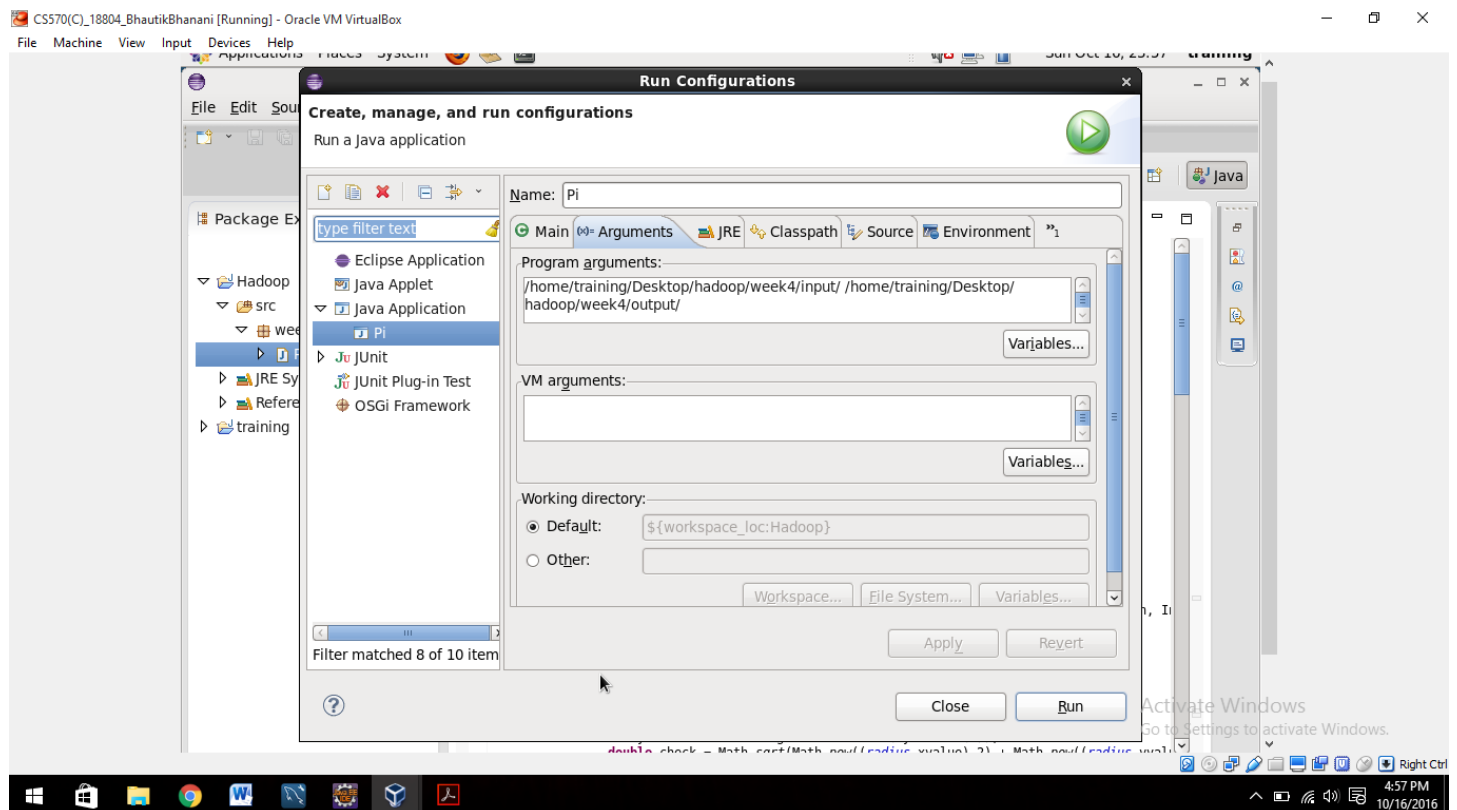
Step 3: Now import necessary hadoop jar files into "Java Build Path" of your project.



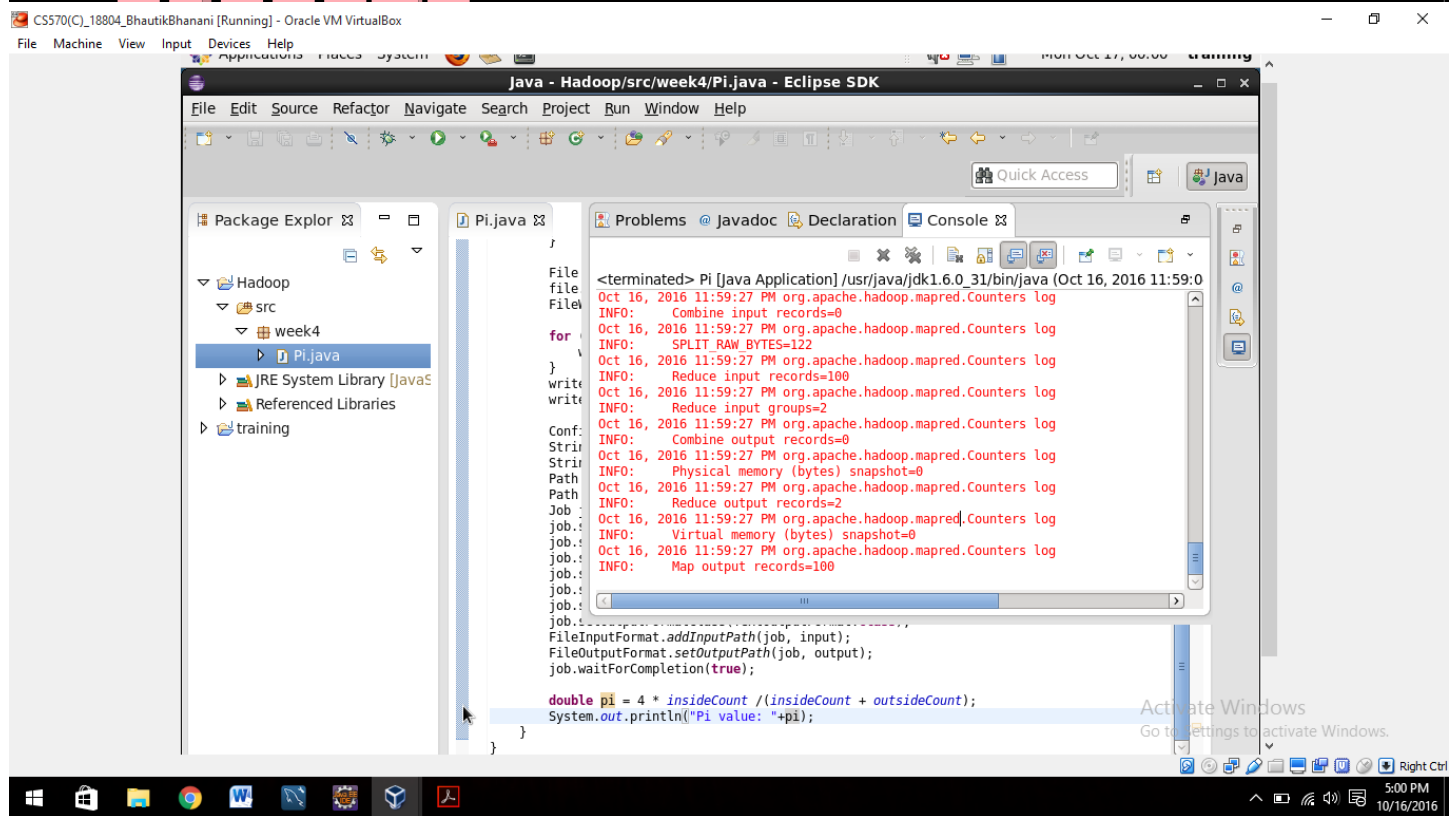
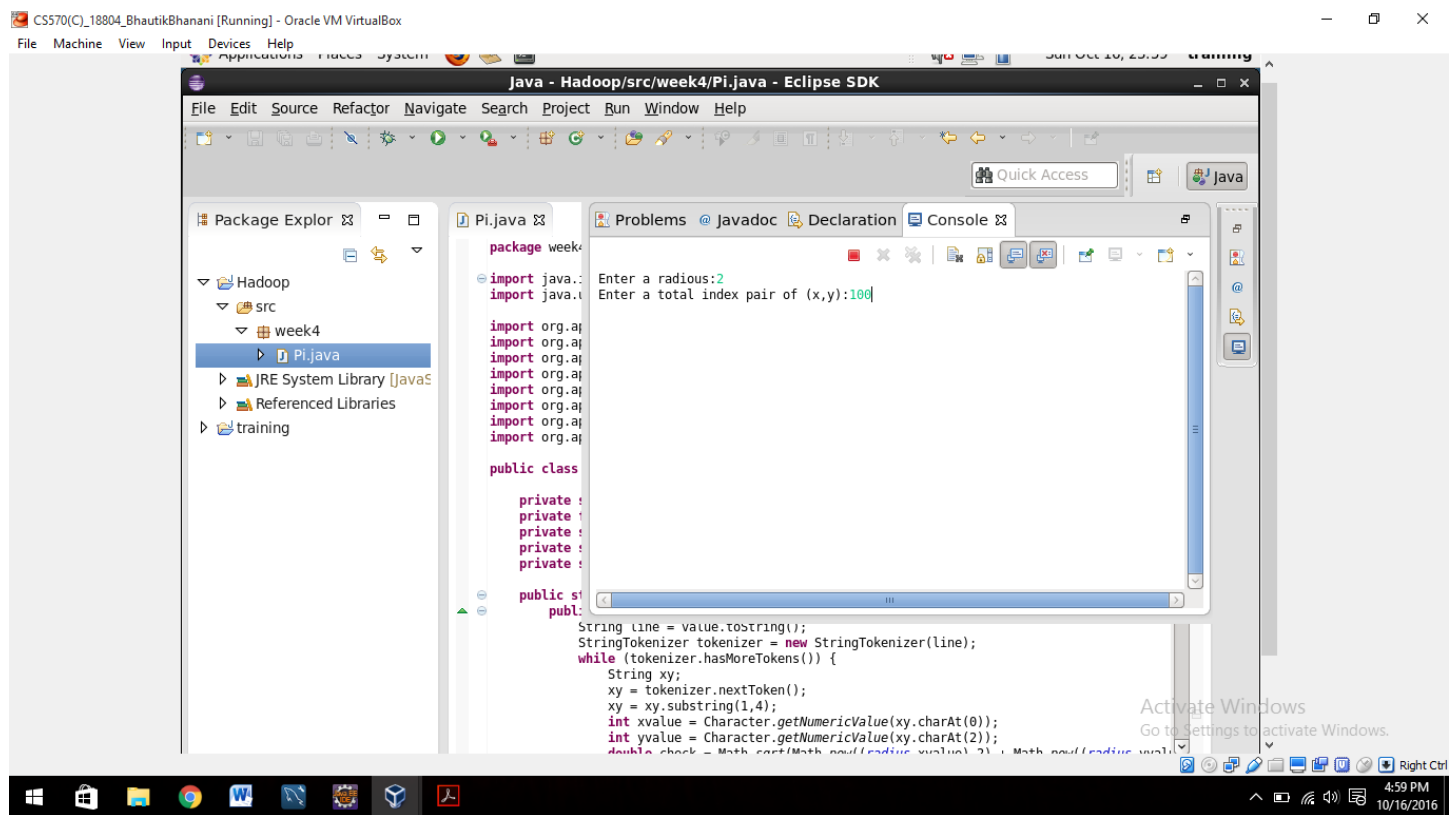
Step 4: Now run this program for once, so this class comes under “Java Application” in Run Configuration window.



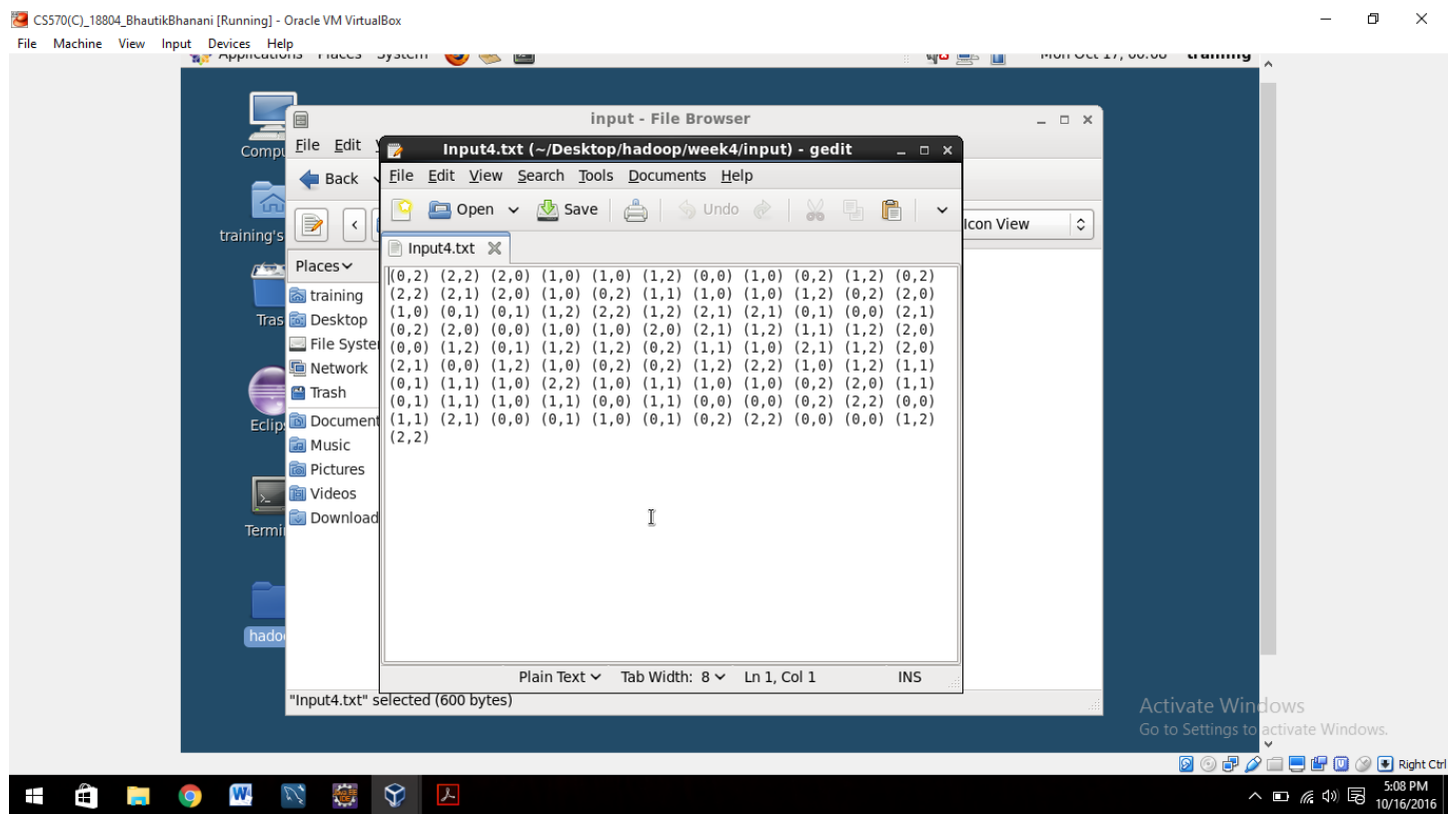
Step 5: Now add argument under “Arguments” window and for first argument give input folder and for second argument, give output folder.



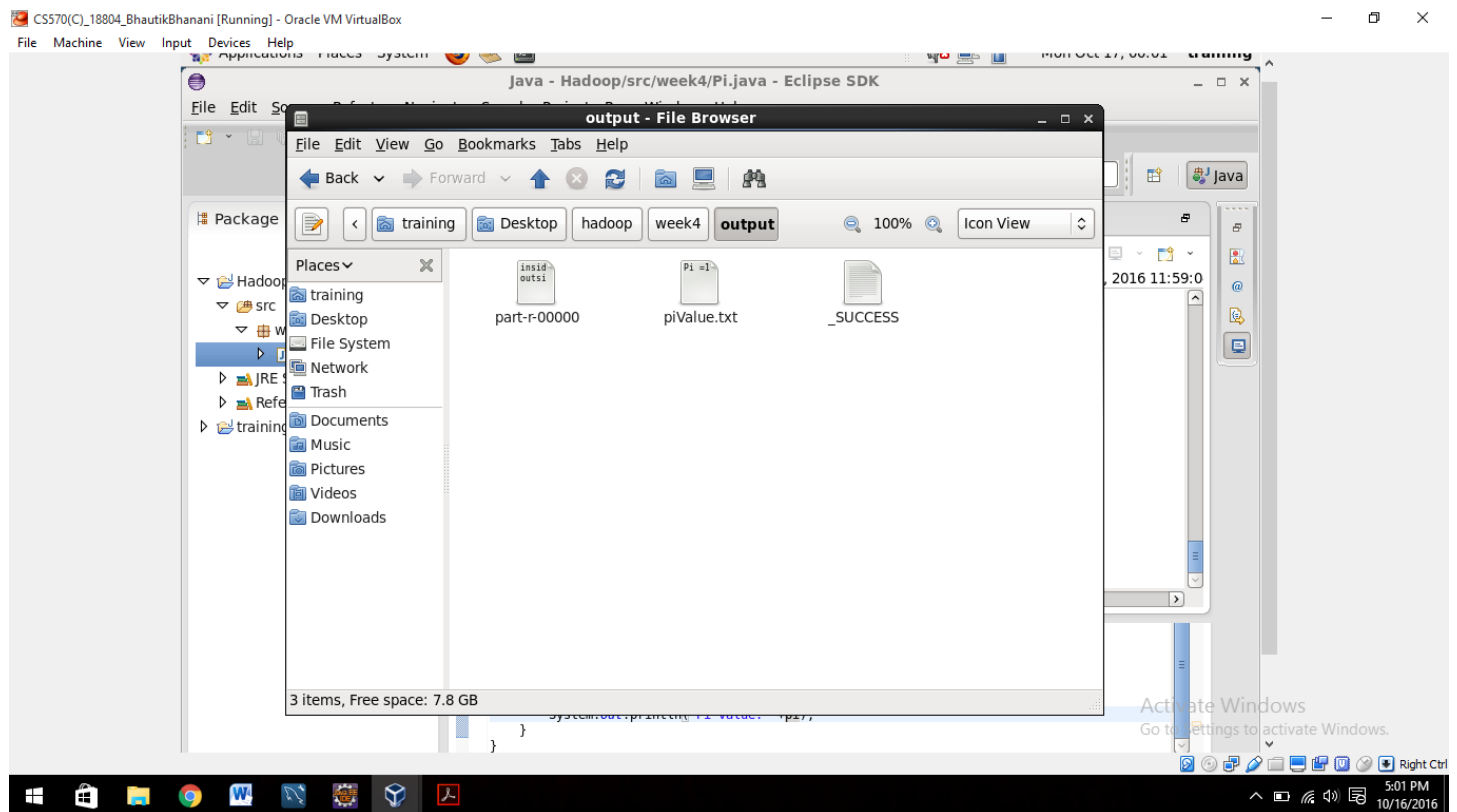
Step 6: Now run the program.



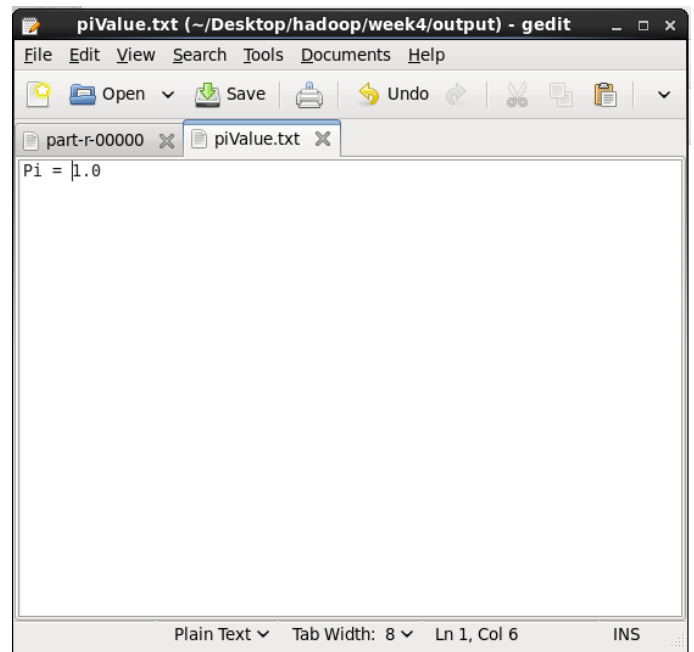
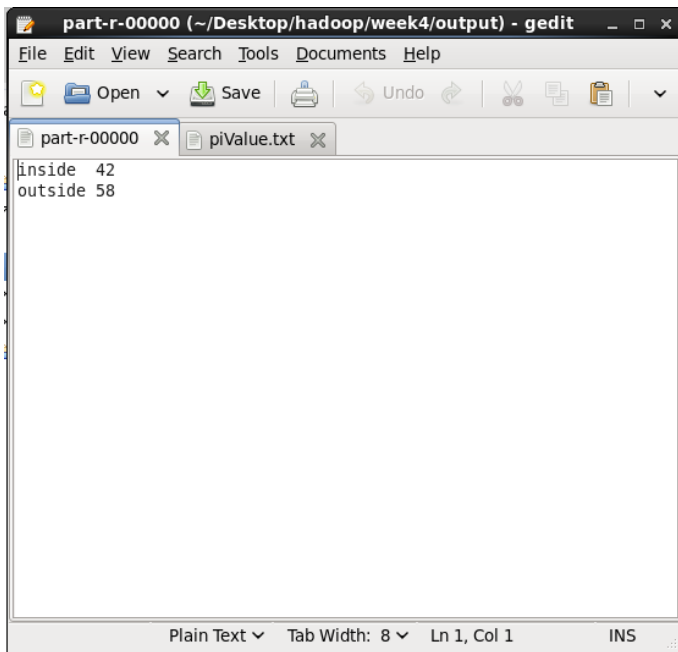
Step 7: Now first check input4.txt file under input folder.



Now check output folder.



Open part-r-0000 and piValue.txt files.



Step 8: Now change driver function as per bellow code. Here I am passing radius and number of (x,y) values in arguments for AWS.

```
public static void main(String[] args) throws Exception{
    radius = Integer.parseInt(args[0]);
    int index = Integer.parseInt(args[1]);

    int numX[] = new int[index];
    int numY[] = new int[index];

    for (int i = 0; i < index; i++){
        numX[i] = (int) (Math.random() * (radius + 1));
        numY[i] = (int) (Math.random() * (radius + 1));
        System.out.println(numX[i] + "," + numY[i]);
    }

    File file = new File(args[2] + "/Input4.txt");
    file.createNewFile();
    FileWriter writer = new FileWriter(file);

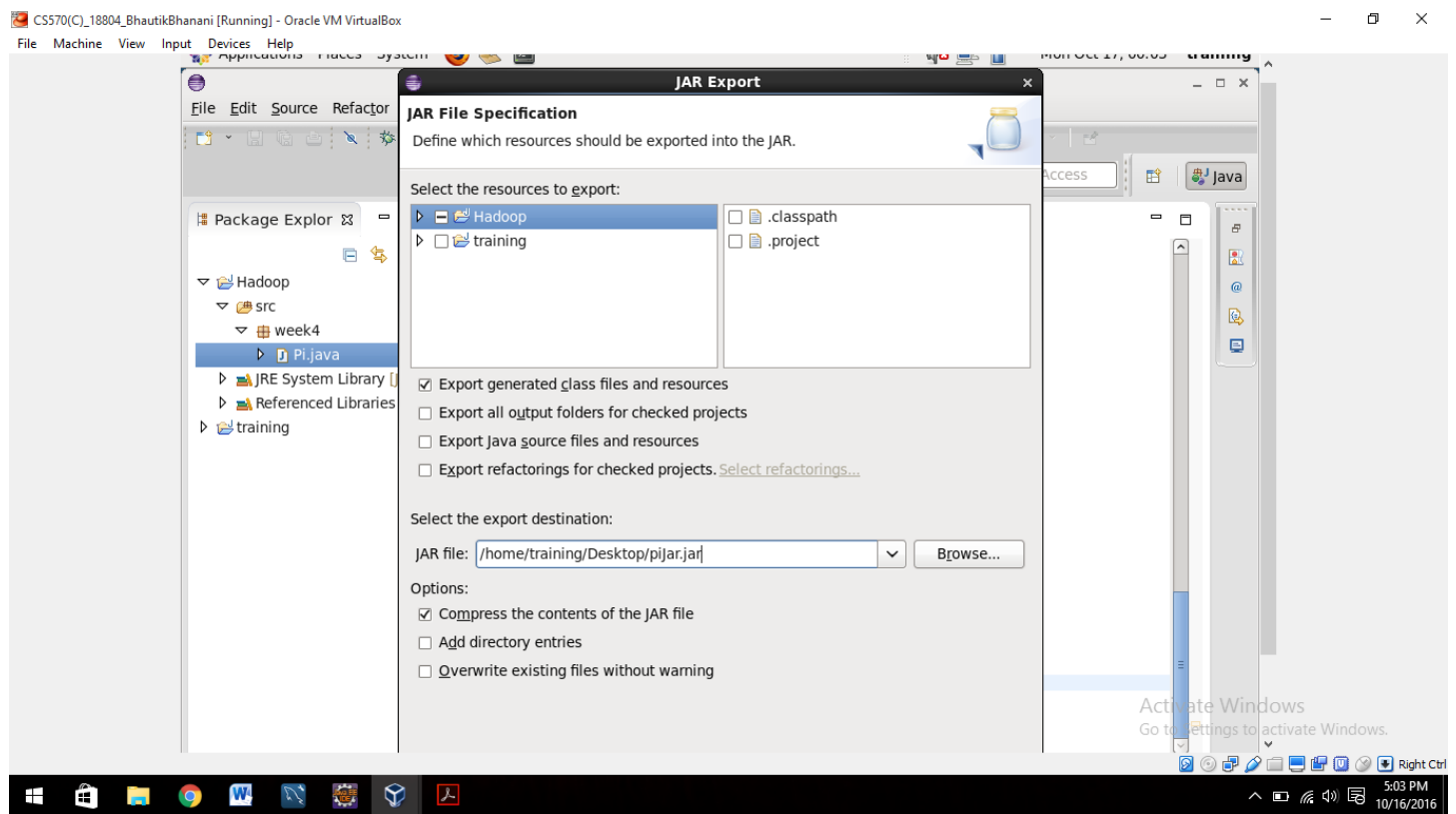
    for (int i = 0; i < index; i++){
        writer.write("(" + numX[i] + "," + numY[i] + ") ");
    }
    writer.flush();
    writer.close();
}
```

```
Configuration conf = new Configuration();
String Input = args[2];
String Output = args[3];
Path input = new Path(Input);
Path output = new Path(Output);
Job job = new Job(conf, "Pi");
job.setJarByClass(Pi.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
job.setMapperClass(Map.class);
job.setReducerClass(Reduce.class);
job.setNumReduceTasks(1);
job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);
FileInputFormat.addInputPath(job, input);
FileOutputFormat.setOutputPath(job, output);
job.waitForCompletion(true);

double pi = 4 * insideCount / (insideCount + outsideCount);
File file1 = new File(args[3] + "/piValue.txt");
FileWriter fw = new FileWriter(file1);
BufferedWriter bw = new BufferedWriter(fw);
bw.write(String.valueOf("Pi value = " + pi));
bw.close();
```

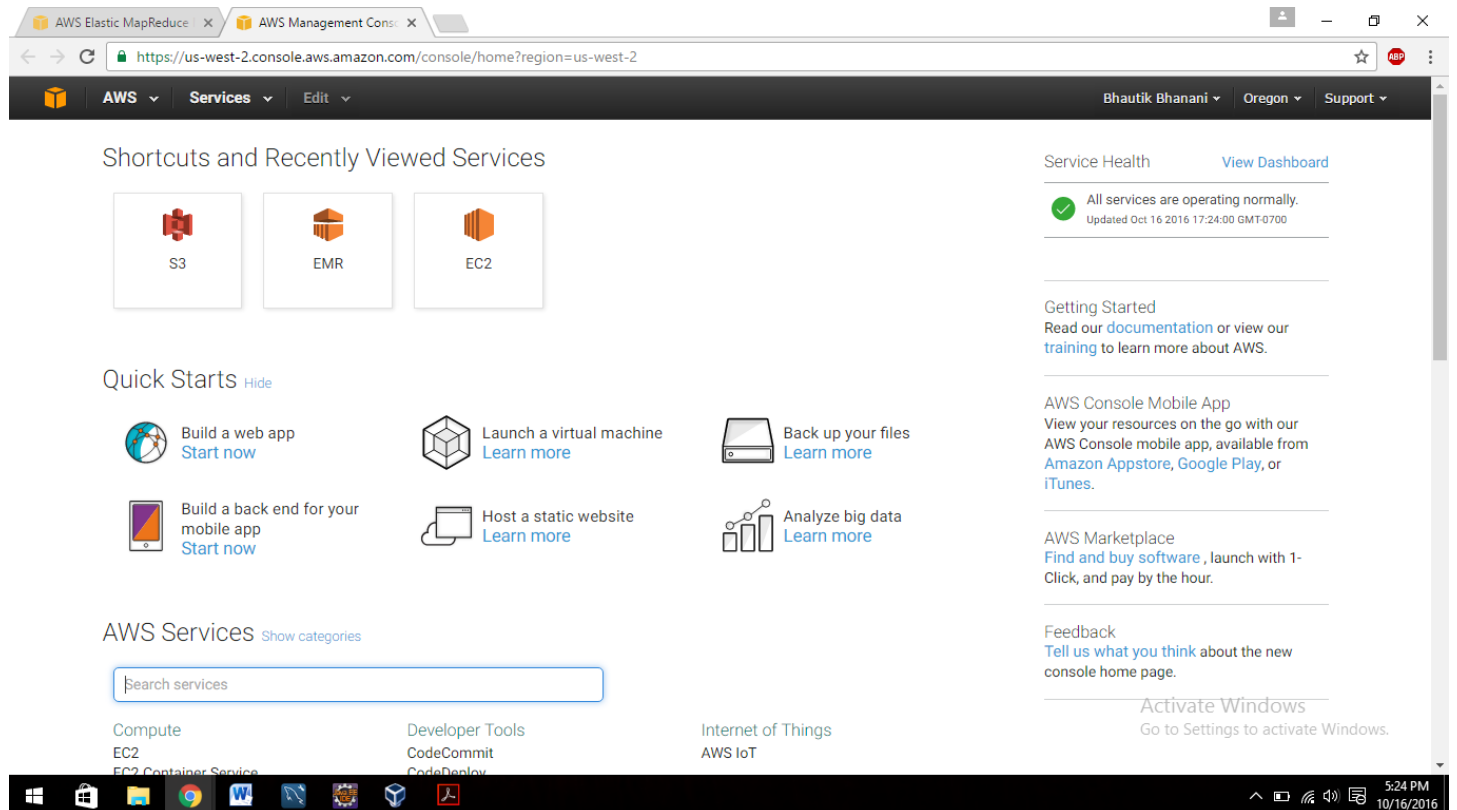
```
}
```

Step 9: Now export Pi.class into jar file.

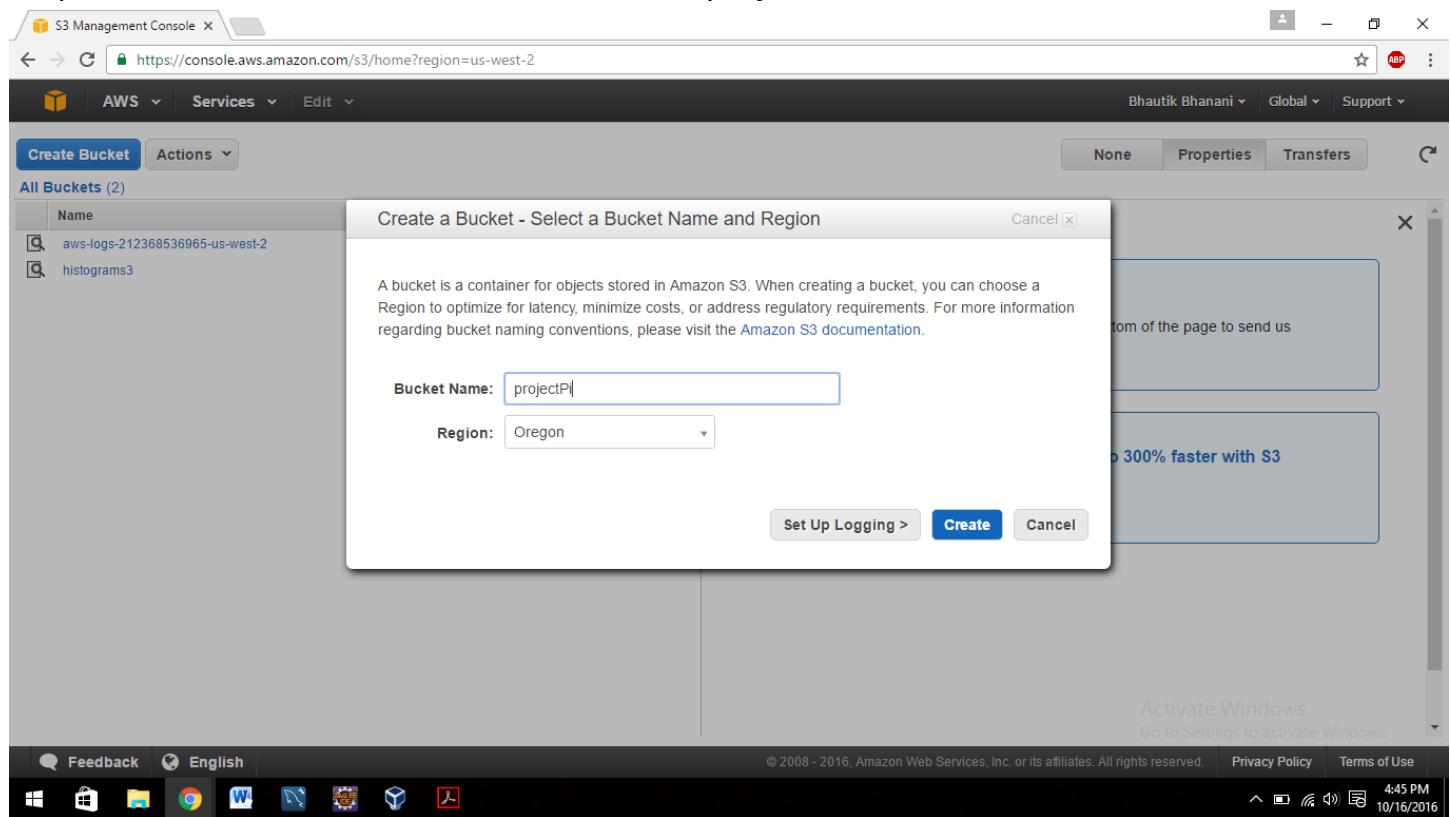


Now run Pi project on AWS

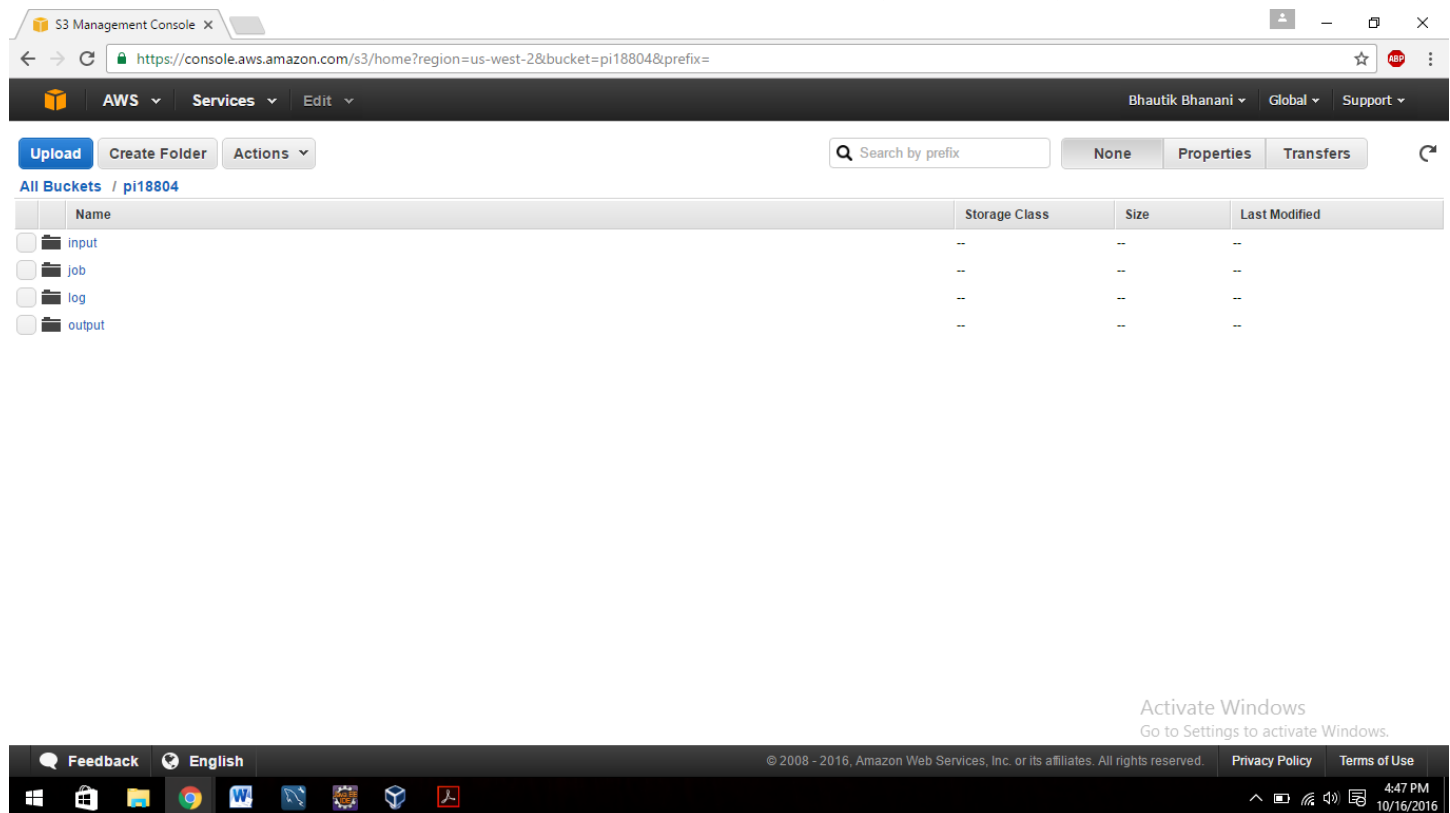
Step 1: Login to your AWS account and AWS dashboard.



Step 2: Go to **S3**, and create one bucket for Pi project.



Step 3: Under your newly created bucket, create following folders: input, output, jobs and log.



Step 4: Upload your exported jar file under **job** folder.

The screenshot shows the AWS S3 Management Console interface. The browser address bar displays the URL: <https://console.aws.amazon.com/s3/home?region=us-west-2&bucket=pi18804&prefix=job/>. The console header includes the AWS logo, navigation tabs (Services, Edit), and user information (Bhautik Bhanani, Global, Support). Below the header, there are buttons for 'Upload', 'Create Folder', and 'Actions'. A search bar labeled 'Search by prefix' is present. The main content area shows a table of objects in the bucket 'pi18804' under the prefix 'job'. The table has columns for Name, Storage Class, Size, and Last Modified. One object is listed: 'piJar.jar' with a Storage Class of 'Standard', a Size of '4.9 KB', and a Last Modified date of 'Sun Oct 16 16:53:37 GMT-700 2016'. At the bottom of the console, there is a Windows taskbar with various application icons and a system tray showing the time as 4:54 PM on 10/16/2016. An 'Activate Windows' watermark is visible in the bottom right corner.

Name	Storage Class	Size	Last Modified
piJar.jar	Standard	4.9 KB	Sun Oct 16 16:53:37 GMT-700 2016

Step 5: Now go to **EC2** from AWS dashboard.

The screenshot shows the AWS EC2 Management Console interface. The browser address bar displays the URL: <https://us-west-2.console.aws.amazon.com/ec2/v2/home?region=us-west-2#>. The console header includes the AWS logo, navigation tabs (Services, Edit), and user information (Bhautik Bhanani, Oregon, Support). On the left, there is a sidebar with navigation links for 'EC2 Dashboard', 'Events', 'Tags', 'Reports', 'Limits', 'INSTANCES', 'IMAGES', 'ELASTIC BLOCK STORE', and 'NETWORK & SECURITY'. The main content area is divided into several sections: 'Resources' (listing 0 Running Instances, 0 Elastic IPs, 0 Dedicated Hosts, 0 Snapshots, 0 Volumes, 0 Load Balancers, 1 Key Pairs, 3 Security Groups, and 0 Placement Groups), 'Create Instance' (with a 'Launch Instance' button), 'Service Health' (showing 'US West (Oregon): This service is operating normally'), 'Scheduled Events' (showing 'No events'), 'Account Attributes' (listing 'Supported Platforms', 'Default VPC', 'Resource ID length management', and 'Additional Information'), and 'AWS Marketplace' (listing 'Find free software trial products' and 'Tableau Server (10 users)'). At the bottom, there is a Windows taskbar with various application icons and a system tray showing the time as 4:48 PM on 10/16/2016. An 'Activate Windows' watermark is visible in the bottom right corner.

Step 6: Create key pair for your project.

EC2 Management Console

https://us-west-2.console.aws.amazon.com/ec2/v2/home?region=us-west-2#KeyPairs:sort=keyName

AWS Services Edit

Bhautik Bhanani Oregon Support

EC2 Dashboard Events Tags Reports Limits

INSTANCES 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

Create Key Pair Import Key Pair Delete

Filter by attributes or search by keyword

1 to 1 of 1

Key pair name Fingerprint

histogramKeyPair c4:fa:7b:a9:65:40:66:85:e5:ca:38:e7:ab:ca:eb:71:58:d9:5c:cc

Create Key Pair

Key pair name: piProjectKeyPair

Cancel Create

Key Pair: histogramKeyPair

Key pair name histogramKeyPair

Fingerprint c4:fa:7b:a9:65:40:66:85:e5:ca:38:e7:ab:ca:eb:71:58:d9:5c:cc

Feedback English

© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

4:48 PM 10/16/2016

Step 7: Go to EMR from AWS dashboard.

AWS Elastic MapReduce

https://us-west-2.console.aws.amazon.com/elasticmapreduce/home?region=us-west-2

AWS Services Edit

Bhautik Bhanani Oregon Support

Amazon EMR

Cluster list Security configurations VPC subnets Help

Create cluster View details Clone Terminate

Filter: All clusters Filter clusters ... 4 clusters (all loaded)

	Name	ID	Status	Creation time (UTC-7)	Elapsed time	Normalize instance I
<input type="checkbox"/>	PiProject	j-4B6IWMR4JRPD	Waiting Cluster ready	2016-10-16 16:51 (UTC-7)	37 minutes	24
<input type="checkbox"/>	My cluster	j-2G63WFG6XULCT	Terminated User request	2016-10-05 20:20 (UTC-7)	1 hour, 27 minutes	48
<input type="checkbox"/>	My cluster	j-V7Y811N4K8FU	Terminated User request	2016-10-04 16:03 (UTC-7)	7 hours, 32 minutes	192
<input type="checkbox"/>	Histogram	j-3CIWR1WMHRTUS	Terminated User request	2016-09-30 18:11 (UTC-7)	2 minutes	0

Activate Windows
Go to Settings to activate Windows.

Feedback English

© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

5:29 PM 10/16/2016

Step 8: Create new cluster of Pi project.

AWS Elastic MapReduce

https://us-west-2.console.aws.amazon.com/elasticmapreduce/home?region=us-west-2#quick-create:

How would you rate your experience with the Elastic MapReduce management console? ☆ ☆ ☆ ☆ ☆

AWS Services Edit Bhautik Bhanani Oregon Support

Create Cluster - Quick Options [Go to advanced options](#)

General Configuration

Cluster name

☒ Logging ⓘ

S3 folder

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Vendor ☒ Amazon ☐ MapR

Release ⓘ

Applications

- ☒ Core Hadoop: Hadoop 2.7.2 with Ganglia 3.7.2, Hive 2.1.0, Hue 3.10.0, Mahout 0.12.2, Pig 0.16.0, and Tez 0.8.4
- ☐ HBase: HBase 1.2.2 with Ganglia 3.7.2, Hadoop 2.7.2, Hive 2.1.0, Hue 3.10.0, Phoenix 4.7.0, and ZooKeeper 3.4.8
- ☐ Presto: Presto 0.150 with Hadoop 2.7.2 HDFS and Hive 2.1.0 Metastore
- ☐ Spark: Spark 2.0.0 on Hadoop 2.7.2 YARN with Ganglia 3.7.2 and Zeppelin 0.6.1

Activate Windows
Go to Settings to activate Windows.

Feedback English © 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 4:50 PM 10/16/2016

AWS Elastic MapReduce

https://us-west-2.console.aws.amazon.com/elasticmapreduce/home?region=us-west-2#quick-create:

How would you rate your experience with the Elastic MapReduce management console? ☆ ☆ ☆ ☆ ☆

AWS Services Edit Bhautik Bhanani Oregon Support

☐ Presto: Presto 0.150 with Hadoop 2.7.2 HDFS and Hive 2.1.0 Metastore

☐ Spark: Spark 2.0.0 on Hadoop 2.7.2 YARN with Ganglia 3.7.2 and Zeppelin 0.6.1

Hardware configuration

Instance type

Number of instances (1 master and 2 core nodes)

Security and access

EC2 key pair ⓘ [Learn how to create an EC2 key pair.](#)

Permissions ☒ Default ☐ Custom

Use default IAM roles. If roles are not present, they will be automatically created for you with managed policies for automatic policy updates.

EMR role [EMR_DefaultRole](#) ⓘ

EC2 instance profile [EMR_EC2_DefaultRole](#) ⓘ

Cancel [Create cluster](#)

Activate Windows
Go to Settings to activate Windows.

Feedback English © 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 4:50 PM 10/16/2016

How would you rate your experience with the Elastic MapReduce management console? ☆ ☆ ☆ ☆ ☆

AWS Services Edit Bhautik Bhanani Oregon Support

Amazon EMR

- Cluster list
- Security configurations
- VPC subnets
- Help

Cluster: PiProject Starting Configuring cluster software

Connections: [Enable Web Connection](#) – Hue, Ganglia, Resource Manager ... (View All)

Master public DNS: ec2-54-70-81-206.us-west-2.compute.amazonaws.com [SSH](#)

Tags: -- [View All / Edit](#)

Summary	Configuration Details	Network and Hardware
ID: j-4B6IWMR4JRPD Creation date: 2016-10-16 16:51 (UTC-7) Elapsed time: 2 minutes Auto-terminate: No Termination protection: Off Change	Release label: emr-5.0.0 Hadoop distribution: Amazon 2.7.2 Applications: Ganglia 3.7.2, Hive 2.1.0, Hue 3.10.0, Mahout 0.12.2, Pig 0.16.0, Tez 0.8.4 Log URI: s3://pi18804/log/ EMRFS consistent view: Disabled	Availability zone: us-west-2c Subnet ID: subnet-68f66230 Master: Bootstrapping 1 m3.xlarge Core: Provisioning 2 m3.xlarge Task: --

Security and Access

Key name: piProjectKeyPair

EC2 instance profile: EMR_EC2_DefaultRole

Activate Windows
Go to Settings to activate Windows.

Feedback English © 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 4:54 PM 10/16/2016

Step 9: When the cluster is in running state, create on step under Steps tab.

AWS Services Edit Bhautik Bhanani Oregon Support

Amazon EMR

- Cluster list
- Security configurations
- VPC subnets
- Help

Security sg-a88229d1
groups for (ElasticMapReduce-Master: master)

Security sg-aa8229d3
groups for (ElasticMapReduce-slave)

Core & Task:

- Monitoring
- Hardware
- Steps
 - [Add step](#) [Clone step](#)
 - [View all interactive jobs](#) | [View all jobs](#)
- Configurations
- Bootstrap Actions

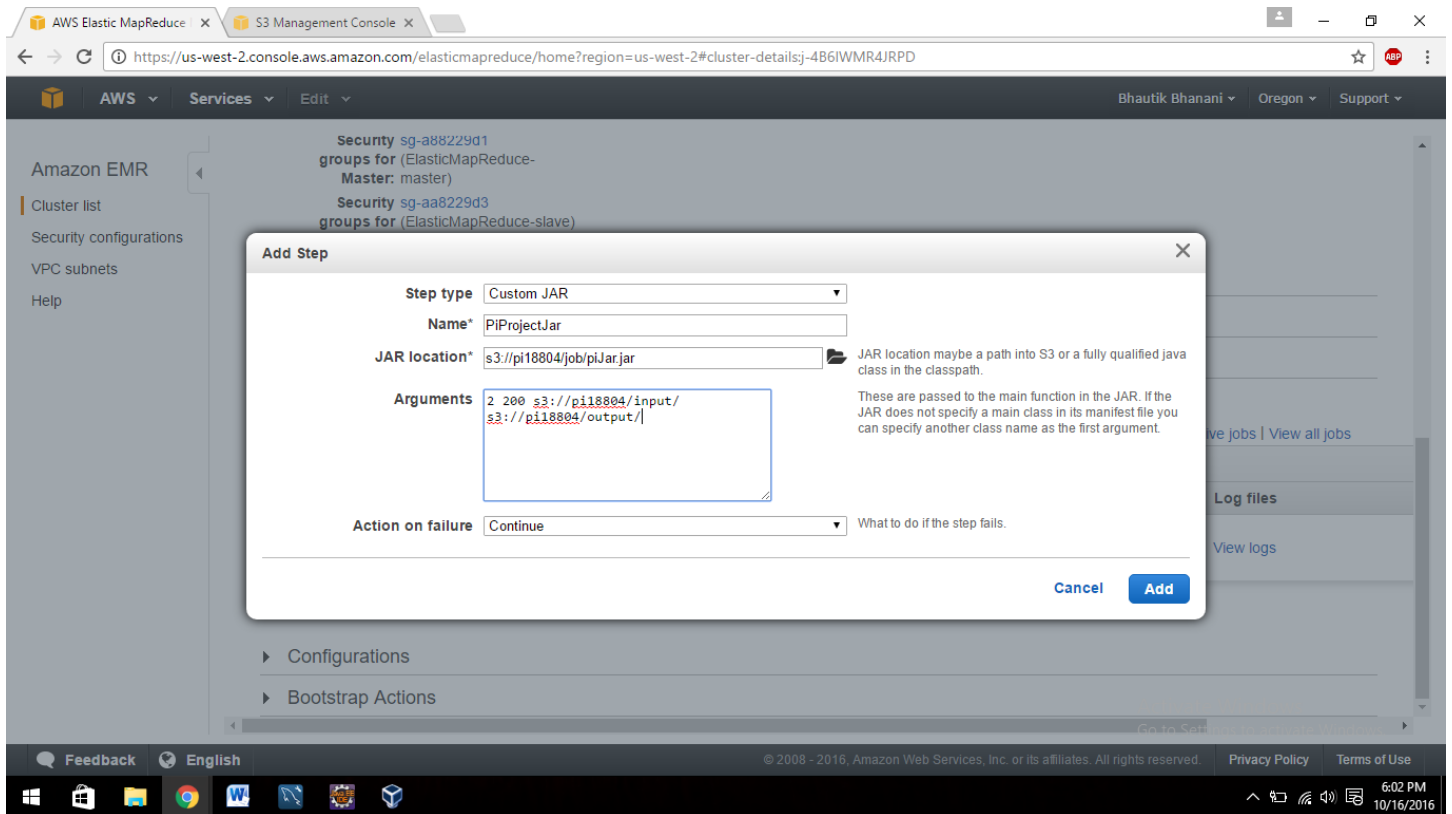
Steps

Filter: All steps Filter steps ... 1 step (all loaded)

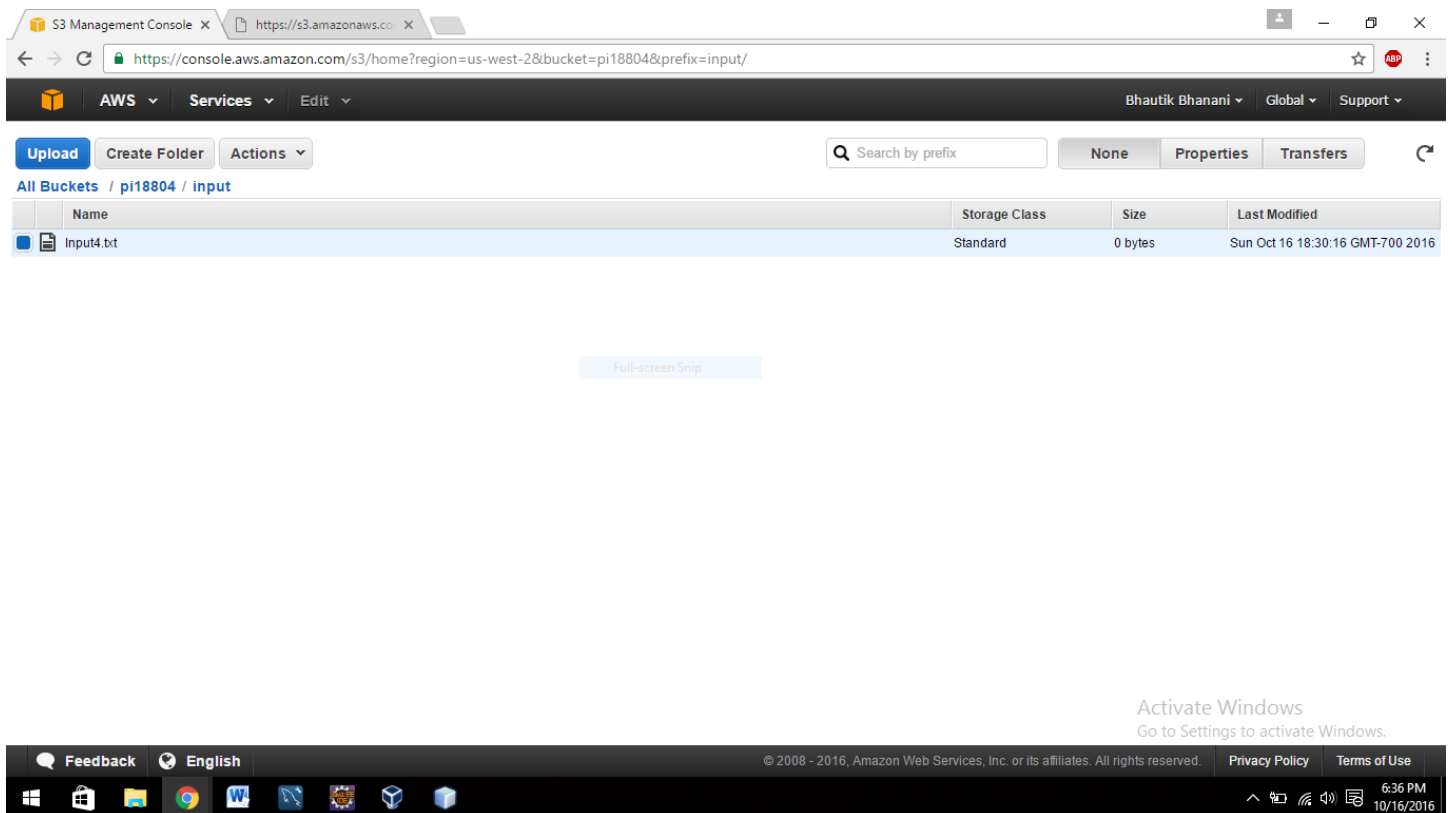
ID	Name	Status	Start time (UTC-7)	Elapsed time	Log files
s-1LO4RMR85ZNT	Setup hadoop debugging	Completed	2016-10-16 16:59 (UTC-7)	2 seconds	View logs

Activate Windows
Go to Settings to activate Windows.

Feedback English © 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use 6:00 PM 10/16/2016



Step 10: When your created step shows status Completed, then go to input folder under **S3**.



Open Input4.txt

S3 Management Console x https://s3.amazonaws.co x

← → ↻ https://s3.amazonaws.com/pi18804/input/Input4.txt?X-Amz-Date=20161017T013557Z&X-Amz-Expires=300&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Signature=3cfd4424430d ☆

(2,2) (1,1) (2,0) (1,2) (1,1) (2,1) (2,2) (2,2) (2,1) (0,1) (0,1) (1,2) (1,1) (1,2) (0,2) (0,1) (0,2) (1,2) (1,1) (2,2) (0,1) (2,1) (0,2) (0,1) (1,0) (1,2) (0,0) (2,0) (2,0) (2,2) (1,0)
(2,2) (1,0) (1,2) (1,1) (0,1) (0,0) (0,0) (0,2) (0,1) (1,0) (2,2) (2,0) (0,1) (0,0) (2,2) (2,1) (1,2) (1,0) (2,0) (1,0) (2,1) (2,2) (2,0) (1,1) (0,1) (1,2) (0,0) (2,1) (1,2) (1,1) (1,2)
(0,2) (1,2) (0,1) (1,1) (0,1) (2,0) (2,0) (0,0) (1,2) (1,0) (2,1) (2,1) (2,2) (2,2) (1,1) (0,0) (0,2) (1,0) (0,0) (2,2) (2,1) (1,2) (2,1) (2,2) (0,0) (2,1) (2,2) (2,1) (2,0) (1,2) (1,1)
(0,0) (0,1) (0,1) (1,1) (0,1) (2,1) (1,2) (1,2) (1,1) (2,0) (1,2) (2,2) (0,0) (1,1) (0,2) (0,1) (0,1) (2,2) (1,1) (1,0) (2,2) (0,2) (2,1) (1,0) (1,1) (1,1) (1,0) (0,0) (0,1) (0,2) (1,1)
(2,0) (1,2) (0,1) (1,0) (2,0) (1,1) (2,2) (2,0) (2,2) (0,2) (0,0) (2,2) (0,2) (1,0) (0,1) (1,1) (2,0) (0,0) (1,0) (2,1) (2,0) (0,2) (2,1) (2,0) (0,0) (1,1) (1,2) (0,0) (0,0) (0,2) (0,1)
(1,0) (0,1) (0,0) (2,0) (2,1) (2,0) (2,2) (2,1) (0,2) (2,1) (2,1) (1,2) (2,1) (1,0) (2,1) (0,0) (0,0) (2,1) (2,2) (2,2) (0,1) (2,1) (2,0) (2,1) (1,0) (2,2) (2,0) (2,0) (0,0) (1,0)
(2,1) (0,1) (2,1) (2,2) (0,1) (2,2) (1,1) (1,0) (0,1) (2,2) (1,2) (0,0) (1,1) (0,0)

Activate Windows
Go to Settings to activate Windows.

6:36 PM
10/16/2016

Step 12: Now check output folder.

S3 Management Console x

← → ↻ https://console.aws.amazon.com/s3/home?region=us-west-2&bucket=pi18804&prefix=output/ ☆

AWS Services Edit Bhautik Bhanani Global Support

Upload Create Folder Actions

Search by prefix None Properties Transfers

All Buckets / pi18804 / output

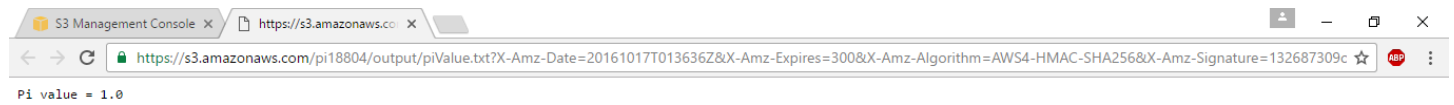
	Name	Storage Class	Size	Last Modified
<input type="checkbox"/>	_SUCCESS	Standard	0 bytes	Sun Oct 16 18:33:45 GMT-700 2016
<input type="checkbox"/>	part-r-00000	Standard	22 bytes	Sun Oct 16 18:33:43 GMT-700 2016
<input type="checkbox"/>	piValue.txt	Standard	14 bytes	Sun Oct 16 18:33:44 GMT-700 2016

Activate Windows
Go to Settings to activate Windows.

Feedback English © 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

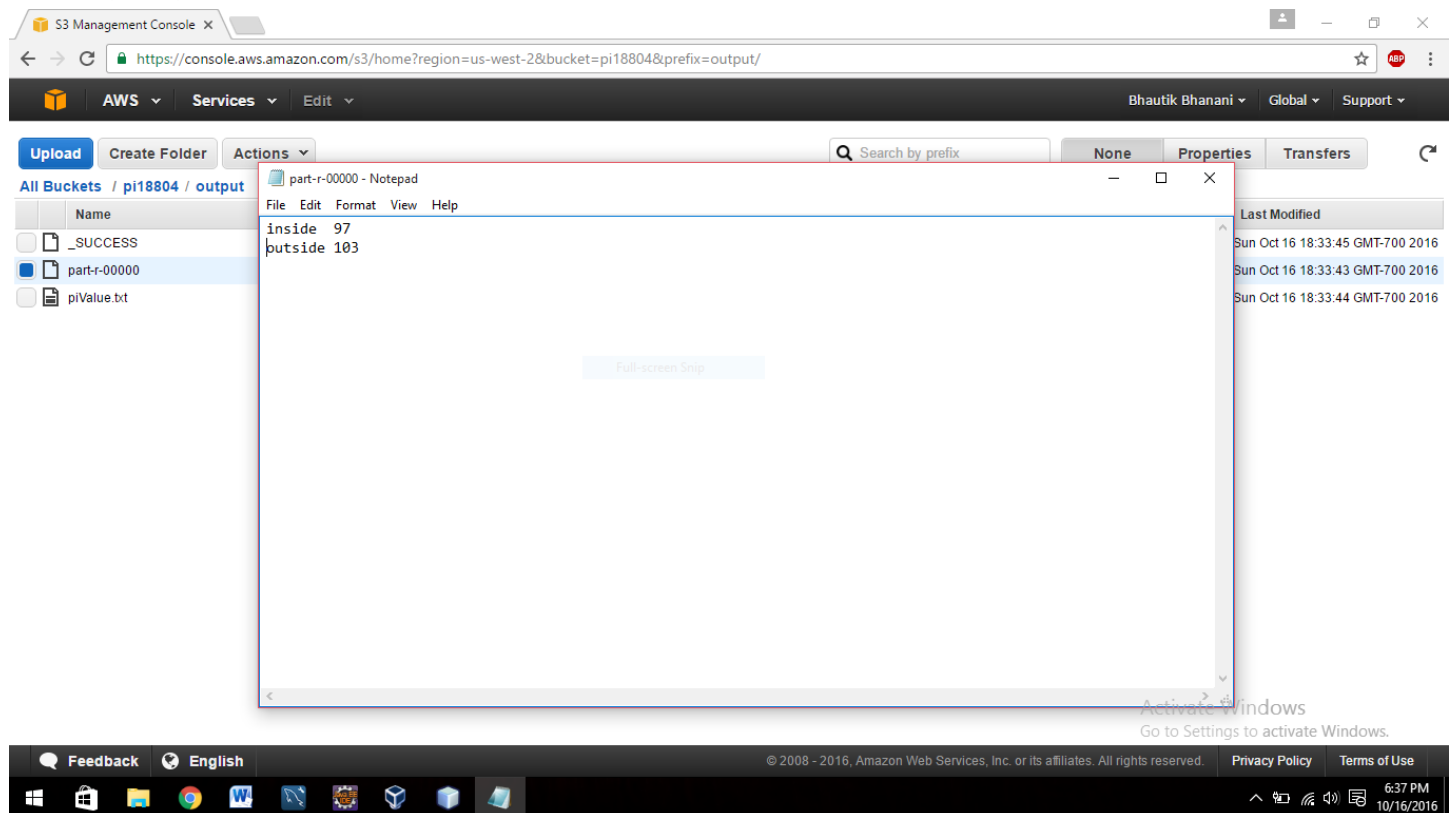
6:36 PM
10/16/2016

Open piValue.txt



Activate Windows
Go to Settings to activate Windows.

Now open part-r-0000 file



Step 13: After this, terminate your cluster. If it will be running then AWS will charge more for that.

AWS Elastic MapReduce

← → ↻ https://us-west-2.console.aws.amazon.com/elasticmapreduce/home?region=us-west-2#cluster-details:j-4B6IWMR4JRPD ☆ 48P

AWS

Services

Edit

Bhautik Bhanani Oregon Support

Amazon EMR

Cluster list

Security configurations

VPC subnets

Help

Add step

Resize

Clone

Terminate

AWS CLI export

Cluster: PiProject

Terminating

Terminated by user request

↻

Connections:

--

Master public DNS:

ec2-54-70-81-206.us-west-2.compute.amazonaws.com

SSH

Tags:

--

Summary

Configuration Details

Network and Hardware

ID: j-4B6IWMR4JRPD

Release label: emr-5.0.0

Availability us-west-2c zone:

Creation date: 2016-10-16 16:51 (UTC-7)

Hadoop Amazon 2.7.2 distribution:

Subnet ID: subnet-68f66230

Elapsed time: 1 hour, 46 minutes

Applications: Ganglia 3.7.2, Hive 2.1.0, Hue 3.10.0, Mahout 0.12.2, Pig 0.16.0, Tez 0.8.4

Master: Running 1 m3.xlarge

Auto- No

Core: Running 2 m3.xlarge

terminate:

Log URI: s3://pi18804/log/

Task: --

Termination protection: Off

EMRFS Disabled

Security and Access

Key name: piProjectKeyPair

EC2 instance profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Visible to all All

Change

Activate Windows

Go to Settings to activate Windows.

Feedback

English

© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

6:37 PM 10/16/2016