



WordCount

amazon
EMR



Word count example

To run word count in AWS you have two different ways; either use the already exist WordCount program, or to write your own file.

We will be using our own word count program

Prerequisites :

- a. S3 bucket
- b. EC2 key pair

The left side of the slide features three overlapping, semi-transparent geometric shapes in shades of blue and grey. These shapes are angular and layered, creating a modern, abstract background element.

Steps:

1. Login to your account in AWS
2. If you did not create Key-Pair and S3 (see below i explained how to do).
3. From Amazon Web Services, go to EMR

Create Key Pair

Step1- Create key-Pair -> Go to the EC2

The screenshot shows the AWS Management Console interface. At the top, the AWS logo is on the left, and navigation links for 'Services', 'Resource Groups', and a star icon are in the center. On the right, there's a user profile '2017msbda008@curaj.ac.in @ ...', location 'Mumbai', and a 'Support' link. A dark navigation bar is at the very top. On the left side, a 'History' sidebar lists 'Console Home', 'EMR', 'EC2', 'Billing', and 'Support'. The main area features a search bar with the placeholder text 'Find a service by name or feature (for example, EC2, S3 or VM, storage)'. To the right of the search bar is a 'Group' dropdown set to 'A-Z'. Below the search bar, services are organized into categories: 'Compute' (with EC2 highlighted by a red circle), 'Developer Tools', 'Analytics', 'Customer Engagement', 'Storage', 'Management Tools', 'Security, Identity & Compliance', 'Business Productivity', 'Desktop & App Streaming', 'Database', and 'Internet Of Things'. Each category has a list of services, some with external links. At the bottom center, there is a 'close' button with an upward arrow icon.

aws Services Resource Groups

2017msbda008@curaj.ac.in @ ... Mumbai Support

History

Console Home

EMR

EC2

Billing

Support

Find a service by name or feature (for example, EC2, S3 or VM, storage).

Group A-Z

Compute

- EC2
- LightSail
- Elastic Container Service
- EKS
- Lambda
- Batch
- Elastic Beanstalk

Developer Tools

- CodeStar
- CodeCommit
- CodeBuild
- CodeDeploy
- CodePipeline
- Cloud9
- X-Ray

Analytics

- Athena
- EMR
- CloudSearch
- Elasticsearch Service
- Kinesis
- QuickSight
- Data Pipeline
- AWS Glue

Customer Engagement

- Amazon Connect
- Pinpoint
- Simple Email Service

Storage

- S3
- EFS
- Glacier
- Storage Gateway

Management Tools

- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Systems Manager
- Trusted Advisor

Security, Identity & Compliance

- IAM
- Cognito
- Secrets Manager
- GuardDuty
- Inspector
- Amazon Macie
- AWS Single Sign-On

Business Productivity

- Alexa for Business
- Amazon Chime
- WorkDocs
- WorkMail

Desktop & App Streaming

- WorkSpaces
- AppStream 2.0

Database

- RDS
- DynamoDB

Internet Of Things

- IoT Core

close

EC2 Management Console

New Tab

+

← → ↻ 🏠

https://ap-south-1.console.aws.amazon.com/ec2/v2/home?region=ap-south-1#Home:

☆

📧

📱

🌐

🔒

👤

⋮

Services ▾

Resource Groups ▾

★

🔔

2017msbda008@curaj.ac.in @ ... ▾

Mumbai ▾

Support ▾

EC2 Dashboard

Events

Tags

Reports

Limits

INSTANCES

Instances

Launch Templates

Spot Requests

Reserved Instances

Dedicated Hosts

IMAGES

AMIs

Bundle Tasks

ELASTIC BLOCK STORE

Volumes

Snapshots

Lifecycle Manager

NETWORK & SECURITY

Resources

You are using the following Amazon EC2 resources in the Asia Pacific (Mumbai) region:

9 Running Instances

0 Elastic IPs

0 Dedicated Hosts

0 Snapshots

20 Volumes

0 Load Balancers

5 Key Pairs

9 Security Groups

0 Placement Groups

Learn more about the latest in AWS Compute from AWS re:Invent 2017 by viewing the [EC2 Videos](#).

✕

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 Asia Pacific (Mumbai) region

Service Health

Service Status:

✓ Asia Pacific (Mumbai):

Availability Zone Status:

Scheduled Events

Asia Pacific (Mumbai):

No events

Account Attributes

Supported Platforms

VPC

Default VPC

vpc-bc65f7d4

Resource ID length management

Console experiments

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:

Barracuda CloudGen Firewall for AWS -

Create S3 bucket..

The screenshot shows the AWS Management Console interface. At the top, the navigation bar includes the AWS logo, 'Services' with a dropdown arrow, 'Resource Groups' with a dropdown arrow, and a user profile icon. The right side of the navigation bar shows the user email '2017msbda008@curaj.ac.in', location 'Mumbai', and a 'Support' link. On the left, a 'History' sidebar lists 'EC2', 'S3', 'Console Home', 'EMR', 'Billing', and 'Support'. The main content area features a search bar with the placeholder text 'Find a service by name or feature (for example, EC2, S3 or VM, storage.)' and a 'Group' dropdown set to 'A-Z'. Below the search bar, services are organized into categories: 'Compute' (EC2, Lightsail, Elastic Container Service, EKS, Lambda, Batch, Elastic Beanstalk), 'Developer Tools' (CodeStar, CodeCommit, CodeBuild, CodeDeploy, CodePipeline, Cloud9, X-Ray), 'Analytics' (Athena, EMR, CloudSearch, Elasticsearch Service, Kinesis, QuickSight, Data Pipeline, AWS Glue), 'Customer Engagement' (Amazon Connect, Pinpoint, Simple Email Service), 'Business Productivity' (Alexa for Business, Amazon Chime, WorkDocs, WorkMail), 'Desktop & App Streaming' (WorkSpaces, AppStream 2.0), 'Security, Identity & Compliance' (IAM, Cognito, Secrets Manager, GuardDuty, Inspector), 'Management Tools' (CloudWatch, AWS Auto Scaling, CloudFormation, CloudTrail, Config, OpsWorks, Service Catalog), and 'Storage' (EFS, Glacier, Storage Gateway). The 'S3' service is highlighted with a red circle. At the bottom of the console, a 'close' button is visible. The footer of the console shows the URL 'https://s3.console.aws.amazon.com/s3/home?region=ap-south-1', copyright information '© 2008 - 2018, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.', and links to 'Privacy Policy' and 'Terms of Use'.

aws Services Resource Groups

2017msbda008@curaj.ac.in @ Mumbai Support

History

EC2

S3

Console Home

EMR

Billing

Support

Find a service by name or feature (for example, EC2, S3 or VM, storage.)

Group A-Z

Compute

- EC2
- Lightsail
- Elastic Container Service
- EKS
- Lambda
- Batch
- Elastic Beanstalk

Developer Tools

- CodeStar
- CodeCommit
- CodeBuild
- CodeDeploy
- CodePipeline
- Cloud9
- X-Ray

Analytics

- Athena
- EMR
- CloudSearch
- Elasticsearch Service
- Kinesis
- QuickSight
- Data Pipeline
- AWS Glue

Customer Engagement

- Amazon Connect
- Pinpoint
- Simple Email Service

Business Productivity

- Alexa for Business
- Amazon Chime
- WorkDocs
- WorkMail

Desktop & App Streaming

- WorkSpaces
- AppStream 2.0

Security, Identity & Compliance

- IAM
- Cognito
- Secrets Manager
- GuardDuty
- Inspector

Management Tools

- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog

Storage

- S3**
- EFS
- Glacier
- Storage Gateway

Database

close

https://s3.console.aws.amazon.com/s3/home?region=ap-south-1

© 2008 - 2018, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Samiksha_008.pem

Show all



Type here to search

Desktop



23:06
20-09-2018

[Services](#) ▾[Resource Groups](#) ▾

2017msbda008@curaj.ac.in @ ... ▾

[Global](#) ▾[Support](#) ▾[Click here to learn how to store and access objects in S3 via NFS](#) [Click here »](#)[Documentation](#)

Amazon S3

[Discover the new console](#)[Quick tips](#)[+ Create bucket](#)[Delete bucket](#)[Empty bucket](#)

13 Buckets

2

Public

3 Regions



| Bucket name | Access | Region | Date created |
|----------------------------------|--------------|-----------------------|-----------------------------------|
| 2017msbda001 | Public | Asia Pacific (Mumbai) | Sep 20, 2018 11:10:12 AM GMT+0530 |
| 2017msbda008 | Not public * | Asia Pacific (Mumbai) | Sep 20, 2018 12:38:09 AM GMT+0530 |
| 2017msbda010 | Public | Asia Pacific (Mumbai) | Sep 20, 2018 10:31:12 PM GMT+0530 |
| aws-logs-233892270056-ap-south-1 | Not public * | Asia Pacific (Mumbai) | Aug 16, 2018 12:27:06 PM GMT+0530 |
| aws-logs-233892270056-us-east-2 | Not public * | US East (Ohio) | Sep 19, 2018 10:45:54 AM GMT+0530 |
| curajtmp | Not public * | Asia Pacific (Mumbai) | Jun 22, 2018 11:48:07 AM GMT+0530 |



Samiksha_008.pem

[Show all](#)

Now, create a folder WordCount and upload your Mapper and Reducer program and your Input file in this folder..

aws

Services

Resource Groups

2017msbda008@curaj.ac.in @ ...

Global

Support

Amazon S3 > 2017msbda008

Overview

Properties

Permissions

Management

Q Type a prefix and press Enter to search. Press ESC to clear.

Upload

Create folder

Actions

Versions

Hide

Show

Asia Pacific (Mumbai)

Viewing 1 to 3

| <input type="checkbox"/> | Name | Last modified | Size | Storage class |
|--------------------------|----------------------|-----------------------------------|------|---------------|
| <input type="checkbox"/> | WordCount | -- | -- | -- |
| <input type="checkbox"/> | J-3SG9AK9K60SNC | -- | -- | -- |
| <input type="checkbox"/> | WordCount_\$folder\$ | Sep 20, 2018 12:15:50 PM GMT+0530 | 0 B | Standard |

Viewing 1 to 3

Samiksha_008.pem

Show all

Here, i uploaded all the files..

Upload

Create folder

Actions ▾

Versions

Hide

Show

Asia Pacific (Mumbai)

| | | | | |
|-------------------------------------|----------------------------|-----------------------------------|----------|----------|
| <input type="checkbox"/> | hadoop-core-1.2.1.jar | Sep 20, 2018 12:41:44 AM GMT+0530 | 4.0 MB | Standard |
| <input type="checkbox"/> | hadoop-streaming-2.7.3.jar | Sep 20, 2018 11:30:21 AM GMT+0530 | 126.2 KB | Standard |
| <input type="checkbox"/> | hbase.id | Sep 20, 2018 11:24:17 AM GMT+0530 | 42.0 B | Standard |
| <input type="checkbox"/> | hbase.version | Sep 20, 2018 11:24:16 AM GMT+0530 | 7.0 B | Standard |
| <input checked="" type="checkbox"/> | mapper.py | Sep 20, 2018 12:20:54 PM GMT+0530 | 553.0 B | Standard |
| <input checked="" type="checkbox"/> | reducer.py | Sep 20, 2018 12:20:55 PM GMT+0530 | 1.0 KB | Standard |
| <input checked="" type="checkbox"/> | sample_text.txt | Sep 20, 2018 12:41:43 AM GMT+0530 | 499.3 KB | Standard |
| <input type="checkbox"/> | wordSplitter.py | Sep 20, 2018 11:45:11 AM GMT+0530 | 285.0 B | Standard |
| <input type="checkbox"/> | wordcount.jar | Sep 20, 2018 1:06:05 AM GMT+0530 | 3.4 KB | Standard |

Operations 0 In progress 1 Success 0 Error

Samiksha_008.pem

Show all



Go to EMR

The screenshot shows the AWS Management Console interface. At the top, the navigation bar includes the AWS logo, a 'Services' dropdown menu, a 'Resource Groups' dropdown, a notification bell, the user email '2017msbda008@curaj.ac.in', a 'Global' region selector, and a 'Support' link. On the left side, there is a 'History' sidebar with links to 'S3', 'EC2', 'Console Home', 'EMR', 'Billing', and 'Support'. The main content area features a search bar with the placeholder text 'Find a service by name or feature (for example, EC2, S3 or VM, storage)'. To the right of the search bar are two buttons: 'Group' and 'A-Z'. The services are organized into a grid of categories, each with an icon and a title. The 'Analytics' category, which includes 'Athena', 'EMR', 'CloudSearch', 'Elasticsearch Service', 'Kinesis', 'QuickSight', 'Data Pipeline', and 'AWS Glue', is circled in red. Other categories include 'Compute' (EC2, Lightsail, Elastic Container Service, EKS, Lambda, Batch, Elastic Beanstalk), 'Developer Tools' (CodeStar, CodeCommit, CodeBuild, CodeDeploy, CodePipeline, Cloud9, X-Ray), 'Customer Engagement' (Amazon Connect, Pinpoint, Simple Email Service), 'Business Productivity' (Alexa for Business, Amazon Chime, WorkDocs, WorkMail), 'Storage' (S3, EFS, Glacier, Storage Gateway), 'Management Tools' (CloudWatch, AWS Auto Scaling, CloudFormation, CloudTrail, Config, OpsWorks, Service Catalog), 'Security, Identity & Compliance' (IAM, Cognito, Secrets Manager, GuardDuty, Inspector), and 'Desktop & App Streaming' (WorkSpaces, AppStream 2.0). The 'Database' category is partially visible at the bottom left.

aws Services Resource Groups

2017msbda008@curaj.ac.in @ ... Global Support

History

S3

EC2

Console Home

EMR

Billing

Support

Find a service by name or feature (for example, EC2, S3 or VM, storage).

Group A-Z

Compute

- EC2
- Lightsail
- Elastic Container Service
- EKS
- Lambda
- Batch
- Elastic Beanstalk

Developer Tools

- CodeStar
- CodeCommit
- CodeBuild
- CodeDeploy
- CodePipeline
- Cloud9
- X-Ray

Analytics

- Athena
- EMR**
- CloudSearch
- Elasticsearch Service
- Kinesis
- QuickSight
- Data Pipeline
- AWS Glue

Customer Engagement

- Amazon Connect
- Pinpoint
- Simple Email Service

Business Productivity

- Alexa for Business
- Amazon Chime
- WorkDocs
- WorkMail

Storage

- S3
- EFS
- Glacier
- Storage Gateway

Management Tools

- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog

Security, Identity & Compliance

- IAM
- Cognito
- Secrets Manager
- GuardDuty
- Inspector

Desktop & App Streaming

- WorkSpaces
- AppStream 2.0

Database

Create cluster..

EMR – AWS Console

Inbox (1,076) - samiksha20167

after termination cluster restart

Step 5: Terminate the Cluster

restart terminated emr cluster

← → ↺ ⌂

https://ap-south-1.console.aws.amazon.com/elasticmapreduce/home?region=ap-south-1#cluster-list:

☆

46

aws

Services ▾

Resource Groups ▾

★

🔔

2017msbda008@curaj.ac.in @ ▾

Mumbai ▾

Support ▾

Amazon EMR

Clusters

Security configurations

VPC subnets

Events

Help

What's new

📘 You can use the AWS Glue Data Catalog as your external Hive metastore for [Apache Spark](#), [Apache Hive](#), and [Presto](#) workloads on Amazon EMR release label 5.10.0 and later. To get started, simply select the AWS Glue Data Catalog for table metadata when creating your cluster.

Create cluster

View details

Clone

Terminate

Filter: All clusters ▾ Filter clusters ...

8 clusters (all loaded) 🔄

| | | Name | ID | Status | Creation time (UTC+5:30) ▾ | Elapsed time | Normalized instance |
|--------------------------|---|--------------|-----------------|----------------------------|-----------------------------|----------------------|---------------------|
| <input type="checkbox"/> | ▶ | 2017msbda010 | j-2CBQZWIZIJ53Z | Terminated User request | 2018-09-20 12:12 (UTC+5:30) | 10 hours, 59 minutes | 132 |
| <input type="checkbox"/> | ▶ | Samiksha_008 | j-3SG9AK9K60SNC | Terminated User request | 2018-09-20 12:03 (UTC+5:30) | 35 minutes | 12 |
| <input type="checkbox"/> | ▶ | My cluster | j-MAMAMN55GRDM | Terminated User request | 2018-09-20 11:54 (UTC+5:30) | 45 minutes | 12 |
| <input type="checkbox"/> | ▶ | curajcluster | j-3JZT0U5E06LA6 | Waiting Cluster ready | 2018-09-20 11:51 (UTC+5:30) | 11 hours, 28 minutes | 132 |
| <input type="checkbox"/> | ▶ | My cluster | j-176HW9BA5Y1GS | Terminated User request | 2018-09-20 11:15 (UTC+5:30) | 1 hour, 23 minutes | 24 |
| <input type="checkbox"/> | ▶ | My cluster | j-TG1X81F0U1JC | Terminated User request | 2018-08-30 11:57 (UTC+5:30) | 2 weeks | 3832 |
| <input type="checkbox"/> | ▶ | My cluster | j-8WCYK0KPGJQ3 | Terminated User request | 2018-08-30 11:51 (UTC+5:30) | 4 minutes | 0 |

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

General Configuration

Cluster name 2017msbda008

☒ Logging ⓘ

S3 folder s3://2017msbda008/WordCount/

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Release emr-5.17.0 ⓘ

Applications

- ☒ Core Hadoop: Hadoop 2.8.4 with Ganglia 3.7.2, Hive 2.3.3, Hue 4.2.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.8.4
- ☐ HBase: HBase 1.4.6 with Ganglia 3.7.2, Hadoop 2.8.4, Hive 2.3.3, Hue 4.2.0, Phoenix 4.14.0, and ZooKeeper 3.4.12
- ☐ Presto: Presto 0.206 with Hadoop 2.8.4 HDFS and Hive 2.3.3 Metastore
- ☐ Spark: Spark 2.3.1 on Hadoop 2.8.4 YARN with Ganglia 3.7.2 and Zeppelin 0.7.3

- ☐ Presto: Presto 0.206 with Hadoop 2.8.4 HDFS and Hive 2.3.3 Metastore
- ☐ Spark: Spark 2.3.1 on Hadoop 2.8.4 YARN with Ganglia 3.7.2 and Zeppelin 0.7.3
- ☐ Use AWS Glue Data Catalog for table metadata

Hardware configuration

Instance type **m4.large** The selected instance type adds a default 32 GiB GP2 EBS volume per instance. [Learn more](#)

Number of instances **5** (1 master and 4 core nodes)

Security and access

EC2 key pair Samiksha_008 [Learn how to create an EC2 key pair.](#)

Permissions
Choose an option
Proceed without an EC2 key pair
2017msbda010
AdityaDL
EC2Key_Pair_CURAJ
be automatically created

EMR role Samiksha_008

EC2 instance profile cluster
curajtmp1_KP



Services ▾

Resource Groups ▾



2017msbda008@curaj.ac.in @ ... ▾

Mumbai ▾

Support ▾

Create Cluster - Quick Options

[Go to advanced options](#)

General Configuration

Cluster name

☒ Logging ⓘ

S3 folder 📁

Launch mode ☒ Cluster ⓘ ☐ Step execution ⓘ

Software configuration

Release ⓘ

- Applications
- ☒ Core Hadoop: Hadoop 2.8.4 with Ganglia 3.7.2, Hive 2.3.3, Hue 4.2.0, Mahout 0.13.0, Pig 0.17.0, and Tez 0.8.4
 - ☐ HBase: HBase 1.4.6 with Ganglia 3.7.2, Hadoop 2.8.4, Hive 2.3.3, Hue 4.2.0, Phoenix 4.14.0, and ZooKeeper 3.4.12
 - ☐ Presto: Presto 0.206 with Hadoop 2.8.4 HDFS and Hive 2.3.3 Metastore
 - ☐ Spark: Spark 2.3.1 on Hadoop 2.8.4 YARN with

Create Cluster - Advanced Options [Go to quick options](#)

Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Software Configuration

Release

- | | | |
|--|---|--|
| <input checked="" type="checkbox"/> Hadoop 2.8.4 | <input type="checkbox"/> Zeppelin 0.7.3 | <input type="checkbox"/> Livy 0.5.0 |
| <input checked="" type="checkbox"/> JupyterHub 0.8.1 | <input type="checkbox"/> Tez 0.8.4 | <input type="checkbox"/> Flink 1.5.2 |
| <input type="checkbox"/> Ganglia 3.7.2 | <input checked="" type="checkbox"/> HBase 1.4.6 | <input checked="" type="checkbox"/> Pig 0.17.0 |
| <input checked="" type="checkbox"/> Hive 2.3.3 | <input type="checkbox"/> Presto 0.206 | <input type="checkbox"/> ZooKeeper 3.4.12 |
| <input type="checkbox"/> MXNet 1.2.0 | <input checked="" type="checkbox"/> Sqoop 1.4.7 | <input type="checkbox"/> Mahout 0.13.0 |
| <input type="checkbox"/> Hue 4.2.0 | <input type="checkbox"/> Phoenix 4.14.0 | <input type="checkbox"/> Oozie 5.0.0 |
| <input checked="" type="checkbox"/> Spark 2.3.1 | <input type="checkbox"/> HCatalog 2.3.3 | <input checked="" type="checkbox"/> TensorFlow 1.9.0 |

AWS Glue Data Catalog settings (optional)

- ☐ Use for Hive table metadata
- ☐ Use for Spark table metadata

HBase storage settings

Storage Mode ☒ HDFS

☐ Use for Hive table metadata ⓘ

☐ Use for Spark table metadata ⓘ

HBase storage settings


Storage Mode ☒ HDFS ☐ S3

Edit software settings ⓘ

☒ Enter configuration ☐ Load JSON from S3

```
classification=config-file-name,properties=[myKey1=myValue1,myKey2=myValue2]
```

Add steps (optional) ⓘ

Step type **Select a step** 


☐ Auto-terminate **Streaming program**

Hive program

Pig program

Spark application

Custom JAR



Cancel

Next

EMR – AWS Console

Inbox (1,076) - samiksha

after termination cluste

Step 5: Terminate the C

restart terminated emr

S3 Management Conso

+

–

✕

← → ↺ 🏠 ⓘ https://ap-south-1.console.aws.amazon.com/elasticmapreduce/home?region=ap-south-1#create-cluster: ☆ 📧 46 📄 ⚙️ 👤 ⋮

aws

Services ▾

Resource Groups ▾

★

🔔 2017msbda008@curaj.ac.in @ ... ▾

Mumbai ▾

Support ▾

☐ Use for Hive table metadata ⓘ

☐ Use for Spark table metadata ⓘ

HBase storage settings ⓘ

Storage Mode

☒ HDFS

☐ S3

Edit software settings ⓘ

☒ Enter configuration

☐ Load JSON from S3

`classification=config-file-name,properties=[myKey1=myValue1,myKey2=myValue2]`

Add steps (optional) ⓘ

Step type

Streaming program ▾

Configure

☐ Auto-terminate cluster after the last step is completed

Cancel

Next

🗨 Feedback

🌐 English (US)

© 2008 - 2018, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

📄 Samiksha_008.pem ^

🏠 🔍 Type here to search

🖨 🖱 📁 🌐 📄

Desktop ? 👤 ^ 🔊 🖥 ENG

23:24 20-09-2018 🗨

EMR – AWS Console

Inbox (1,076) - samiksha

after termination cluste

Step 5: Terminate the C

restart terminated emr

S3 Management Conso

←

→

↺

🏠

https://ap-south-1.console.aws.amazon.com/elasticmapreduce/home?region=ap-south-1#create-cluster:

☆

Services ▾

Resource Groups ▾

📌

2017msbda008@curaj.ac.in @ ... ▾

Mumbai ▾

Support ▾

Add step

×

Step type

Streaming program

Name*

WordCount

Mapper*

s3://2017msbda008/WordCount/mapper.py

S3 location of the map function or the name of the Hadoop streaming command to run.

Reducer*

s3://2017msbda008/WordCount/reducer.py

S3 location of the reduce function or the name of the Hadoop streaming command to run.

Input S3 location*

s3://2017msbda008/WordCount/sample_text.txt

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

Output S3 location*

s3://2017msbda008/WordCount/result

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

Arguments

Action on failure

Continue ▾

What to do if the step fails.

Cancel

Add

Next

Feedback

English (US)

© 2008 - 2018, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.

Privacy Policy

Terms of Use

Samiksha_008.pem

↑

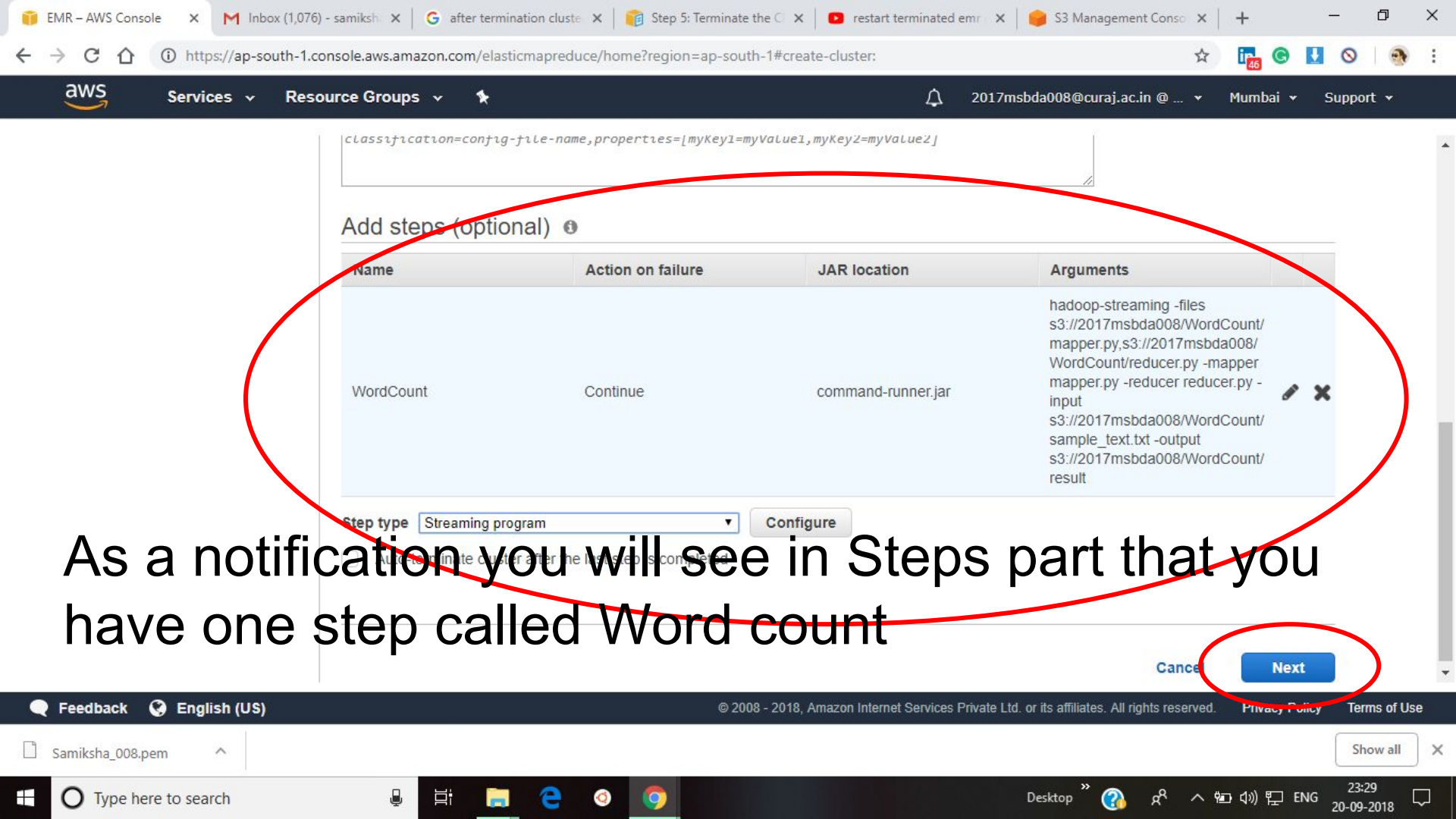
Type here to search

Desktop

ENG

23:26

20-09-2018



Specify target capacity and how Amazon EMR fulfills it for each node type. Mix instance types and purchasing options. [Learn more](#)

Network: [Create a VPC](#)

EC2 Subnet:

Root device EBS volume size: GiB

Choose the instance type, number of instances, and a purchasing option. You can choose to use On-Demand Instances, Spot Instances, or both. The instance type and purchasing option apply to all EC2 instances in each instance group, and you can only specify these options for an instance group when you create it. [Learn more about instance purchasing options](#)

| Node type | Instance type | Instance count | Purchasing option | Auto Scaling |
|----------------------|--|--|--|--------------------------|
| Master Master - 1 | m4.large 4 vCore, 8 GiB memory, EBS only storage EBS Storage: 32 GiB | 1 Instances | <input checked="" type="radio"/> On-demand <input type="radio"/> Spot Use on-demand as max price | Not available for Master |
| Core Core - 2 | m4.large 4 vCore, 8 GiB memory, EBS only storage EBS Storage: 32 GiB | <input type="text" value="2"/> Instances | <input checked="" type="radio"/> On-demand <input type="radio"/> Spot Use on-demand as max price | Not enabled |
| Task Task - 3 | m4.large 4 vCore, 8 GiB memory, EBS only storage EBS Storage: 32 GiB | <input type="text" value="0"/> Instances | <input checked="" type="radio"/> On-demand <input type="radio"/> Spot Use on-demand as max price | Not enabled |

[+ Add task instance group](#)

Cancel Previous **Next**

Create Cluster - Advanced Options

[Go to quick options](#)[Step 1: Software and Steps](#)[Step 2: Hardware](#)**Step 3: General Cluster Settings**[Step 4: Security](#)

General Options

Cluster name ☒ Logging ⓘS3 folder ☒ Debugging ⓘ☒ Termination protection ⓘ

Tags ⓘ

| Key | Value (optional) | |
|--|---|---|
| <input type="text" value="Enrollment_number"/> | <input type="text" value="2017msbda008"/> | ✕ |
| <input type="text" value="Add a key to create a tag"/> | <input type="text"/> | |

Additional Options

☐ EMRFS consistent view ⓘCustom AMI ID ⓘ

Step 3: General Cluster Settings

Step 4: Security

☒ Logging ⓘS3 folder 📁☒ Debugging ⓘ☒ Termination protection ⓘ

Tags ⓘ

| Key | Value (optional) | |
|--|---|---|
| <input type="text" value="Enrollment_number"/> | <input type="text" value="2017msbda008"/> | ✕ |
| <input type="text" value="Add a key to create a tag"/> | <input type="text"/> | |

Additional Options

☐ EMRFS consistent view ⓘCustom AMI ID ⓘ

▶ Bootstrap Actions

Cancel

Previous

Next

In Security and Access part choose your key pair

aws Services ▾ Resource Groups ▾ ☆

2017msbda008@curaj.ac.in @ ... ▾ Mumbai ▾ Support ▾

Create Cluster - Advanced Options [Go to quick options](#)

Step 1: Software and Steps

Step 2: Hardware

Step 3: General Cluster Settings

Step 4: Security

Security Options

EC2 key pair **Samiksha_008** ⓘ

☒ Cluster visible to all IAM users

Permissions **EC2Key_Pair_CURAJ**

☒ Default ☐ Use default IAM role

Use default IAM role **curajtmp1_KP** ⓘ created

Let us help you with managed policies for automatic policy updates.

EMR role **EMR_DefaultRole** ⓘ

EC2 instance profile **EMR_EC2_DefaultRole** ⓘ

Auto Scaling role **EMR_AutoScaling_DefaultRole** ⓘ

▶ Authentication and encryption

▶ EC2 security groups

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

Click Create Cluster.

aws

Services

Resource Groups

2017msbda008@curaj.ac.in @ ... Mumbai Support

Amazon EMR

Clusters

Security configurations

VPC subnets

Events

Help

What's new

Clone Terminate AWS CLI export

Cluster: 2017msbda008 Starting

Summary Application history Monitoring Hardware Events Steps Configurations Bootstrap actions

Add step Clone step Cancel step

Steps

Filter: All steps Filter steps ... 2 steps (all loaded)

| | ID | Name | Status | Start time (UTC+5:30) | Elapsed time | Log files | Actions |
|--|----------------|------------------------|---------|-----------------------|--------------|-----------|-----------|
| | s-3M1GBOJPF63R | WordCount | Pending | | | View logs | View jobs |
| | s-QXUJV00AR3BG | Setup hadoop debugging | Pending | | | View logs | View jobs |

View all interactive jobs | View all jobs

The page of creating cluster will appear and the state of your cluster is Starting. The cluster now in creating process and after the creation finished the state of your cluster will be changed to Running.

Go to S3 console and choose the bucket ,you will find the result folder that contains the output, it's something like in figure below.

Amazon S3 > 2017msbda008 / WordCount / result

Overview

Type a prefix and press Enter to search. Press ESC to clear.

Upload Create folder Actions Versions Hide Show Asia Pacific (Mumbai)

Viewing 1 to 4

| <input type="checkbox"/> | Name ↑ | Last modified ↑ | Size ↑ | Storage class ↑ |
|--------------------------|------------|-----------------------------------|--------|-----------------|
| <input type="checkbox"/> | _SUCCESS | Sep 20, 2018 11:45:32 PM GMT+0530 | 0 B | Standard |
| <input type="checkbox"/> | part-00000 | Sep 20, 2018 11:45:17 PM GMT+0530 | 1.7 KB | Standard |
| <input type="checkbox"/> | part-00001 | Sep 20, 2018 11:45:25 PM GMT+0530 | 1.7 KB | Standard |
| <input type="checkbox"/> | part-00002 | Sep 20, 2018 11:45:31 PM GMT+0530 | 1.6 KB | Standard |

Viewing 1 to 4

Samiksha_008.pem

Show all