Q1- How to use the IAM credentials provided by the AWS classroom account for the initial AWS CLI configuration?

A1– Assume that you have the following credentials:

```
[default]
aws_access_key_id=ASIAZDQOJORVE5HP746Y
aws_secret_access_key=khEb4p0JGfSvvlNsZ2yy8Du0BGSPmGaPUc8Ru3ac
aws_session_token=FQoGZXIvYXdzEA4aDE4K7OLITypUTxrTLCKFAteWMVbQn8HHM0iAxD8yQZO
cTtUjK/CF1pvbQuiuMom0RE4Df+fJLk/qR7b/s5XrOUvY3QhkCs/4kQQCPP4o/iziuunZEIK9sfVb
pLUMBW+khsYH41cKBhT2VZsWL82qnzz2FWgdFWK7ZbMjzg5eMgGYGE59rpflbb0Fd3ToGGh4Qrerv
CAUc89n8fSM49Lo1xVSaLNlmkHUv+XghaKFwJwsY9LA7/RovazNkTwwWz2GVbXaW3zib9NJG0Ea/W
Km8UhKrYBtBlHeFSr0TFuoUQugOF3kCDJkQIoujdmTaoHAtFVe795wKuGYyujGh7yFQC96Dj29Dwq
ud2AdMlrNfGN1kD8nOyjpuv3eBQ==
```

Follow the steps below for MAC/Linux:

1- Run `aws configure`

```
PFSL131$ sudo aws configure
Password:
AWS Access Key ID [****************QVUJ]: ASIAZDQOJORVE5HP746Y
AWS Secret Access Key [****************Xn6A]: khEb4p0JGfSvvlNsZ2yy8Du0BGSPmGaPUc8Ru3ac
Default region name [None]:
Default output format [json]:
```

2- Copy and paste the credentials in ~/.aws/credentials

3- Set local variables `AWS_ACCESS_KEY_ID, AWS_SECRET_ACESS_KEY, and AWS_SESSION_TOKEN`

```
PFSL131$ export AWS_ACCESS_KEY_ID=ASIAZDQOJORVE5HP746Y
PFSL131$ export AWS_SECRET_ACESS_KEY=khEb4p0JGfSvvlNsZ2yy8Du0BGSPmGaPUc8Ru3ac
PFSL131$ export
AWS_SESSION_TOKEN=FQoGZXIvYXdzEA4aDE4K7OLITypUTxrTLCKFAteWMVbQn8HHM0iAxD8yQZOc
TtUjK/CF1pvbQuiuMom0RE4Df+fJLk/qR7b/s5XrOUvY3QhkCs/4kQQCPP4o/iziuunZEIK9sfVbpLUMBW+khsYH41c
KBhT2VZsWL82qnzz2FWgdFWK7ZbMjzg5eMgGYGE59rpflbb0Fd3ToGGh4QrervCAUc89n8fSM49Lo1xVSaLNlmkHUv+
XghaKFwJwsY9LA7/RovazNkTwwWz2GVbXaW3zib9NJG0Ea/WKm8UhKrYBtBlHeFSr0TFuoUQugOF3kCDJkQIoujdmTa
oHAtFVe795wKuGYyujGh7yFQC96Dj29Dwqud2AdMlrNfGN1kD8nOyjpuv3eBQ==
```

4- `PFSL131$ source ~/.bash_profile`

5- Test the connection

```
PFSL131$  aws s3 ls
2018-10-31 15:33:29 bamn-123
2018-11-03 17:02:53 bamn-300
2018-11-03 22:22:34 bamn-400
```

Q2- IAM policy for the AWS Educate Classroom account.

A2- In the AWS Educate classroom account you are not allowed to create your own IAM policy, but you are free to use any of the AWS Managed Policy. Please note that, for the coursework project, you are not necessarily required to create your own (new) IAM policy. In fact, you might only be required to create new IAM roles (which should not be mistaken with IAM policies). And your AWS Classroom account has no limitations to create new roles using the IAM console/dashboard (`AWS Console->Services->IAM->from the left side dashboard: Roles->Create role`).

"However, you must note that you can create a new role within the IAM console and referencing it within the service you are using (ARN or AWS resource name of the new role can be used to refer a role), rather than creating a new role within a service directly.

For example, it is not possible to create a lambda execution role directly from the lambda console; instead create the lambda execution role from the IAM console and reference it from the lambda console when using it. Similarly, for any external tools/services (non-AWS services/tools), if it is required, you need to configure them to use existing roles (that you have already created by IAM console) rather than letting the external tool to create new roles directly." [from the Vocareum Amazon Educate Q&A web site]

Q3- Are we allowed to use tools like "kops/Kubeadm", or should we use the raw AWS API to build the Kubernetes cluster manually?

A3- Yes, you are allowed to use these tools as long as you are covering all the requirements of the project described in the coursework description. Please note that your code must be precisely reproduceable as described by the coursework description. This means that your final implementation should be able to automatically (not manually) perform all steps including any kind of configurations.

Q4- Are there any restrictions for the custom application implementation in Step 3?

A4- Yes, there are some limitations. For any part of the project you can freely use Java, Python, Linux/Shell scripting, or a combination of them. You are not allowed to use any other languages.

Q5- Can you provide more information on the example inputs/outputs on the Assessment web page?

A5- The answer is based on the **sample-c.txt**.

All potential "kubernetes" instances include:

Kubernetes→1-**valid** (the analysis is not case-sensitive)

Kubernetes→2-**valid** (the analysis is not case-sensitive)

KuberneteS→3-**valid** (the analysis is not case-sensitive)

Kuber5netes→invalid (as described in the coursework, you need to ignore all words containing non-letter characters such as [0-9], @, %, &, *, $, £, \, /, ^, `, #, >, <, ~, =, +, etc)

Kubernetes→4-**valid** (the analysis is not case-sensitive)

Kubernetes7→invalid (similarly, we ignore all words containing non-letter characters like 7)

KuberneTes→5-**valid**

Kubernetes@Kubernetes→invalid similarly, we ignore all words containing non-letter characters like @)

```
Popular Words
+----+------------+---------+
|Rank|Word        |Frequency|
+----+------------+---------+
|1   |applications|6        |
|2   |kubernetes  |5        |
|3   |system      |4        |
+----+------------+---------+
```

Similar to above:
system→ 1-**valid**
SystEM→ 2-**valid**
System. → 3-**valid** (as mentioned in the coursework, we identify words using punctuation marks. Here, “.” is a punctuation/tokenizer. For this project, the list of valid punctuations/tokenizers are described as **","**, **"."**, **";"**, **":"**, **"?"**, **"!"**, **"""**, **"("**, **")"**, **"["**, **"]"**, **"{"**, **"}"**, **"-"**, and **"_"**)
system, → 4-**valid** “,” is among valid punctuations.
sYstem> → invalid (the character “>” is not described in the punctuation list, rather it is a non-letter character. Any word containing non-letter character must be ignored in your analysis)

Q6- How are you tokenising the input in the examples?

A6- For this project, the list of valid punctuations/tokenizers are described as **","**, **"."**, **";"**, **":"**, **"?"**, **"!"**, **"""**, **"("**, **")"**, **"["**, **"]"**, **"{"**, **"}"**, **"-"**, and **"_"**)

Q7- In step 5 what are the grading criteria?

A7- You will accessed based on the *functionality, design* and *performance* of your work.
Functionality: your code matches the requirements given in Step 5.
Design: we will look into the design/architecture of your schedulers and the algorithms/performance models used to find the required allocation.
Performance: we will look into the time it takes for the scheduler to find the required allocation.

Note that the performance criteria is the least important one amongst the three given. We are primarily interested on the *functionality* and *design*.

Q8- Are there any rules using my IAM user account?

A8- You are given new IAM user accounts (one per group) created within a regular root AWS account. Please use these accounts with the following very important rules in mind:

1. Although you will be able to see other users and groups and their used resources, we expect that you will be working solely with the resources (such as EC2 instances) that you have created. You are not allowed for any kind of efforts aiming to

influence/modify/monitor/destroy other users resources (e.g., stop/terminate/login to other groups' instances). To avoid any confusions when created resources use tags to name and identity your resources. Please use your group names in the tags.

2. You are not allowed to add/delete other users.
3. You are not allowed to add/modify/delete IAM policies. If you need any additional policies please email me asap and I will add these for you.
4. You are allowed to use up to 10 t2.small and 1 c4.large instances in the EU(London) region as per coursework specifications. Feel free to grab additional instances from other regions.
5. You do not have access to the Billing Information.
6. Currently, there are no set limits on the budget allowed to use. However, please use resources sensibly as you were to use your AWS Educate Classroom account. Closely pay attention to the storage you allocate. We highly recommend to use storage volumes with small sizes and work with small input files as per your coursework specifications. We will be closely monitoring the budget usage and will be sending notification emails in case of heavily over-spending.

Q9- Do we need to support an arbitrary number of masters and workers?

A9- Yes, you do need to support an arbitrary number of masters and workers but for the coursework results you only need to run 1 master and 10 worker nodes. Due to budget considerations please test your code against only small numbers of worker nodes (<10) and 1 master. Your code will not be tested against large numbers.

Q10- Is there a list of configuration parameters we need to support for option 1 in the CLI?

A10- Yes, you need to support the following parameters with default values shown in parenthesis:

1. Number of worker nodes (10)
2. Number of master nodes (1)
3. The instance type of workers (t2.small) (please refer to **Q14** for additional clarifications)
4. The instance type of the master (c4.large)
5. AWS Region (EU (London))

Q11- Is there a specific address we can use to download/access the input files for the applications?

A11- You can use any of the following two options to access the input files for your applications.

1. URL: For example:
   https://s3.eu-west-2.amazonaws.com/cam-cloud-computing-data-source/data-200MB.txt
   https://s3.eu-west-2.amazonaws.com/cam-cloud-computing-data-source/data-400MB.txt
   https://s3.eu-west-2.amazonaws.com/cam-cloud-computing-data-source/data-500MB.txt
2. S3 Addresses: You can use the following addresses to find the files stored in S3:
   s3://s3.eu-west-2.amazonaws.com/cam-cloud-computing-data-source/data-200MB.txt
   s3://s3.eu-west-2.amazonaws.com/cam-cloud-computing-data-source/data-400MB.txt
   s3://s3.eu-west-2.amazonaws.com/cam-cloud-computing-data-source/data-500MB.txt

The data input files are also stored in Google Drive. The related coursework web page links point out to these locations:
200MB: https://drive.google.com/file/d/1-L2rTLYTTLCJRUC4fBxq-V7xAogaAbdw
400MB: https://drive.google.com/file/d/1xTKR6eKiis7a8ywpA4LH4obdc1QilAhq
500MB: https://drive.google.com/file/d/1rEPbjpsrMYPxIUcZ_s5RfT0FvIyg3CYH

Please note that your implementation should be able to handle arbitrary URLs and S3 addresses. If your code has any specific limitations regarding getting the input files, please discuss these in your report.

Q12- Which CRSid should I use to name the database given that we are two students per group?

A12- You can use any of the following options:

1. <CRSid for student1>_CloudComputingCoursework
2. <CRSid for student2>_CloudComputingCoursework
3. <CRSid for student1>_<CRSid for student2>_CloudComputingCoursework
4. <group id>_CloudComputingCoursework

Q13- To access the web-dashboard, what type of credentials are required to be extracted by the 3.1 and 3.2 options of the CLI interface?

A13- Both admin password and admin token might be needed to access the web dashboard. There are APIs which can be used to extract these values dynamically (you shouldn't copy/paste).

Please note that to open the web-dashboard on the localhost:80001, you may only need the token. But, for doing this on the apiserver (the master node), both passwords are required. This information are required for the assessment purpose.

Q14- If I use the PySpark approach then I might run out of memory with t2.small instances. What should I do?

A14- In this case you can use t2.medium. However, please note that we your cluster should always be homogeneous, meaning that in this case all 10 instances should be t2.medium.
In this case the default values for the worker nodes in option 1 of the CLI should be t2.medium.

Q15- For the database you say that we "*do not need to re-create the database from scratch every time you run your application*". Can you please clarify this?

A15- The script in Step 2, when running option "5: Run Spark WordLetterCount App URL", should create the database once if the database is not already created. So effectively you create the database once. If you have already implemented this as part of the Spark WordLetterCount application itself then you could leave it as it. Again please clarify this in your report.