

ASSIGNMENT 1

SUBMISSION SPECIFICATIONS

Python instructions

1. Please add `#!/usr/bin/python3` at the beginning of your python files.
2. Please make sure you are testing your scripts on hadoop before submitting, and not just locally.
3. Ensure you do not mix up tabs and spaces, run your code with Python3 to ensure no syntactical errors.
4. You will be submitting **four files, two** for each task. The following naming convention is to be strictly followed:
 - a. Task 1 - TeamID_mapper.py, TeamID_reducer.py (eg:- BD_0002_0098_0218_1367_mapper.py)
 - b. Task 2 - TeamID_mapper.py, TeamID_reducer.py (eg:- BD_0002_0098_0218_1367_reducer.py)

Java instructions

1. You will be submitting **two files, one** for each task. The following naming convention is to be strictly followed:
 - a. Class Name - **TeamID** (eg:- BD_0002_0098_0218_1367)
 - b. File Name - **TeamID.java** (eg:- BD_0002_0098_0218_1367.java)

Note: There will be no difference in the naming convention for both the tasks.

2. Your java files will be exported to a jar file, which will be run as follows:

```
bin/hadoop jar <Full Path to Jar File>.jar /inputA1 /outputA1
```

We request you to not create a package as the command to run the same will be different, which is not supported by the evaluation portal.

Converting Windows/DOS Text Files to Linux/UNIX Format

If you have written your solutions on a Windows system, we recommend you run either command before submitting before submitting.

```
dos2unix filename
```

or

```
fromdos filename
```

(Source: <https://forum.pesu.io/t/assignment-1-issues/13282/6>)

QUICK WALKTHROUGH

[Submission Portal : <http://34.235.91.135>]

Step 1: Login using your TeamID and default password.

Big Data - UE18CS322

BD_5_6_7_8

LOGIN

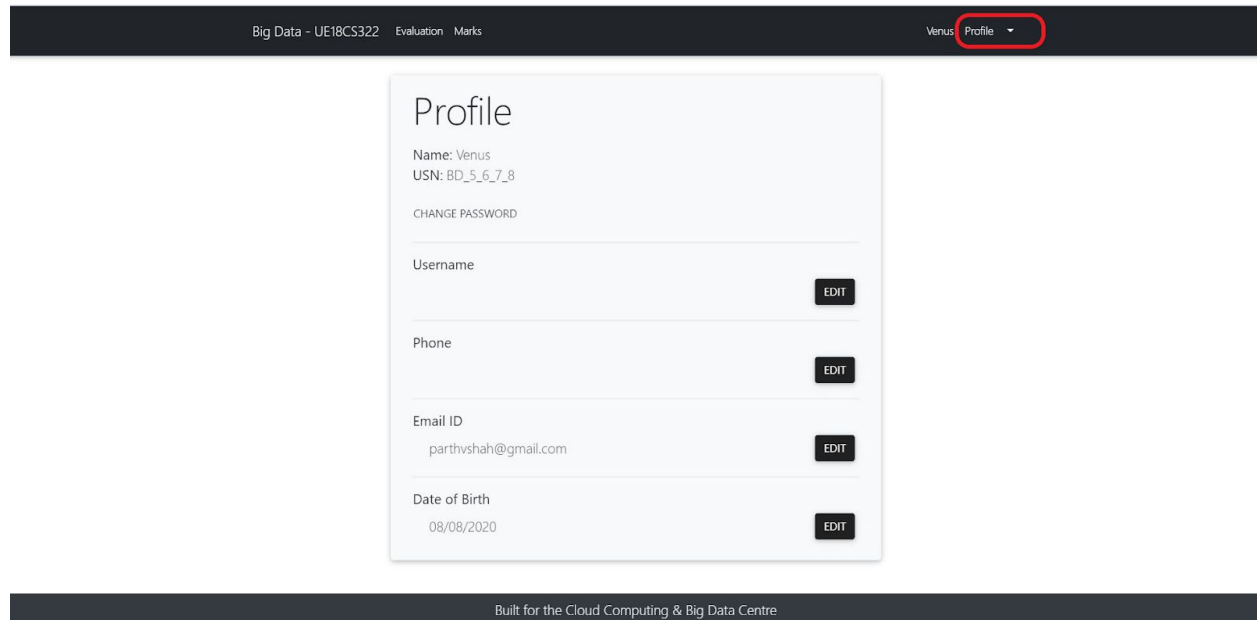
Announcements

- 20-08-2020 - The first test of the portal is successful! Assignments are autoevaluated.
- 26-08-2020 - Assignment 1 is now live. More details at <https://forum.pesu.io/t/assignment-1-specification/13140>. Due date is 09-09-2020.
- 29-08-2020 - Beta testing is enabled! Register using the form that can be found in the forum and login here.

Built for the Cloud Computing & Big Data Centre

© 2020 [The Alcoding Club](#), CS&E, PES University, Bengaluru, India.

Step 2: Change your password by clicking on Profile.



The screenshot shows a web application interface. At the top, a dark navigation bar contains the text "Big Data - UE18CS322 Evaluation Marks" on the left and "Venus Profile" on the right, with "Profile" highlighted by a red circle. Below the navigation bar is a white "Profile" card. The card displays the user's name "Venus" and USN "BD_5_6_7_8". Below this is a "CHANGE PASSWORD" link. The card also features four input fields with "EDIT" buttons: "Username", "Phone", "Email ID" (containing "parthvshah@gmail.com"), and "Date of Birth" (containing "08/08/2020"). At the bottom of the page, a dark footer bar contains the text "Built for the Cloud Computing & Big Data Centre".

Step 3 : To submit your assignment, click on Evaluation.

Step 3.1: Enter the Assignment ID by cross-checking with the official list of assignments (refer screenshot below)

Step 3.2: Upload the required files. Naming convention should be followed strictly.

Step 3.3: Click on the “ARM” button. The “ARM” button checks when your last submission was. If 15 minutes have elapsed since, the “SUBMIT” button will be activated and you can click on it to submit your files to the backend for evaluation. If not, you will see an error message and you must try again later.

Note: If you click on “SUBMIT” after “ARM”ing but you had made a mistake (eg:- wrong assignment ID, incorrect file **type**), an error will show. You will still be able to “ARM” your submission, and you can click on “SUBMIT” after correcting the error.

Big Data - UE18CS322

Evaluation Marks

Venus Profile

Assignments

Assignment ID	Question	Max Marks
ATT1	Find count of word specified.	2

Evaluation

Team ID: BD_5_6_7_8

Assignment ID: ATT1

Naming convention must be strictly followed.

Files:

Choose Files

No file chosen

ARM

SUBMIT

Built for the Cloud Computing & Big Data Centre

Step 4: To view your marks, click on the Marks tab, enter your assignment ID and view marks for each of your submissions for that assignment.

Note : Each test case is checked, and marks are awarded accordingly as can be seen in the examples below.

Assignments

Assignment ID	Question	Max Marks
A1T1	Find count of word specified.	2

Marks

Team ID: BD_5_6_7_8

Assignment ID: A1T1

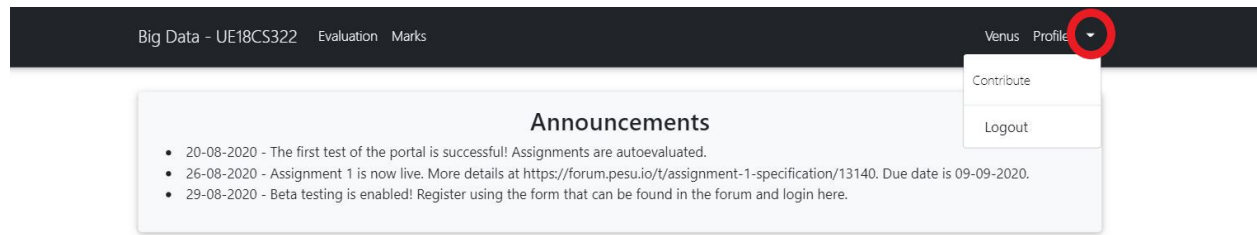
SUBMIT

Assignment ID	Team ID	Submission ID	Submitted At	Message	Marks Obtained
A1T1	BD_5_6_7_8	47	Sun, Aug 30, 2020 4:17 PM	A1T1 - TC1: Correct Answer\nA1T1 - TC2: Correct Answer\n	2/2
A1T1	BD_5_6_7_8	46	Sun, Aug 30, 2020 3:55 PM	A1T1 - TC1: Correct Answer\nA1T1 - TC2: Wrong Answer\n	1/2

Note: Attached below is an example of the submission results you may receive - either a wrong output or a compilation error. If your submission hasn't been evaluated yet, it will appear as Not Evaluated. This means it is in the pipeline, and you will receive your marks shortly. Do remember each task takes a significant amount of time to run, and with the amount of submissions, there will be wait times involved with receiving your marks.

A1T1	BD_5_6_7_8	42	Sun, Aug 30, 2020 3:33 PM	A1T1 - TC1: Wrong Answer\nA1T1 - TC2: Wrong Answer\n	0/2
A1T1	BD_5_6_7_8	41	Sun, Aug 30, 2020 3:32 PM	ID not found	Not evaluated
A1T1	BD_5_6_7_8	40	Sun, Aug 30, 2020 3:29 PM	Compile Error Exception in thread "main" java.lang.ClassNotFoundException: 40-A1T1 at java.net.URLClassLoader.findClass(URLClassLoader.java:382) at java.lang.ClassLoader.loadClass(ClassLoader.java:418) at java.lang.ClassLoader.loadClass(ClassLoader.java:351) at java.lang.Class.forName0(Native Method) at java.lang.Class.forName(Class.java:348) at org.apache.hadoop.util.RunJar.run(RunJar.java:316) at org.apache.hadoop.util.RunJar.main(RunJar.java:236)	0/2

Step 6 [OPTIONAL] : You may choose to contribute to the Alcoding Club, PES University.



IMPORTANT

- Please keep in mind that your scripts will be run through a plagiarism checker. Copying of code is strictly prohibited. While you are allowed to discuss varied approaches to your solutions, be aware that any kind of cheating will not be tolerated, and strict action will be taken against all teams who disregard this warning.
- The portal contains checks on file types, as well as sizes. If there is an attempt to sabotage the portal, your team will get 0 as everything is being logged.
- Discussions regarding faculty decisions about the same will not be entertained.