



THE CATHOLIC UNIVERSITY OF EASTERN AFRICA

A. M. E. C. E. A

P.O. Box 62157

00200 Nairobi - KENYA

Telephone: 891601-6

Ext 1022/23/25

MAIN EXAMINATION

JANUARY – APRIL 2022

FACULTY OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

REGULAR PROGRAMME

CMT 432: SCALA & APACHE SPARK PROGRAMMING

Date: APRIL 2022

Duration: 2 Hours

- 1. INSTRUCTIONS: Answer Question ONE and any other TWO Questions.**
- 2. Calculators MAY be used for this Examination.**

Q1

- a) What are the two main programming paradigms supported by Scala?
(2 marks)
- b) Distinguish between the following
 - i. Java String and Scala String (1 mark)
 - ii. Java main and Scala main methods (1mark)
 - iii. Java statement and Scala statement (1mark)
 - iv. Scala Anonymous function and Currying functions (2marks)
 - v. Scala var and val keywords. (1mark)
 - vi. The f and raw string interpolation as used in scala (1mark)
 - vii. Nested function and local variable (2marks)

- c) In an organization an employee must pay tax which is based on the employee total earnings less the legal deductions. The PAYEE is based on the following brackets

Bracket	Tax
$x \leq 15000$	0%
$15000 < x \leq 25000$	10%
$25000 < x \leq 50000$	20%
$50000 < x \leq 70000$	30%
$70000 < x$	50%

Use Scala **match** structure to create a function `payee()` that receives an employee taxable income and return the tax payable. **(6marks)**

- d) Define the following terms as they are used in Apache Spark.
- i. Resilient Distributed Datasets (RDDs) **(2 marks)**
 - ii. A DataFrame **(2 marks)**
 - iii. A Dataset **(2 marks)**
- e) Briefly discuss any four RDD transformations **(4 marks)**
- f) What is the key main difference between Spark and Hadoop MapReduce **(1 marks)**
- g) List any three application areas of Spark. **(2 marks)**

Q2

- a) Explain in details four conditions that makes a Scala function impure. **(4 marks)**
- b) What do we mean when we say that the **for expression** and **foreach** method in Scala are used for their **side effects**. **(2 marks)**
- c) Write the Scala equivalent for the scenarios below in the shortest possible way.

i. `val (age: Int, fname: String, lname:String) = Pair(40, "Joyce", "Ogogo")`

(2marks)

- ii. A function that can return the sum of three integer inputs (2marks)
- d) Write a Scala program that uses default argument function that receive as argument randomly generated, CAT1 marks out 10, CAT2 out of 20, EXAM out of 70 and return their sum. (6 marks)
- e) With examples give the difference between a class in Scala and in Java . (4 marks)

Q3

- a) Define the following terms as they apply to Scala
 - i. Lazy collections and give an example (2 marks)
 - ii. Immutable collection and give an example (2 marks)
 - iii. Extractor (2 marks)
 - iv. Option (2marks)
- b) Give the method that can be used to add the elements of a given list **newages** in front of the list **currages**. (2 marks)
- c) Distinguish between Sets and List in Scala (2 marks)
- d) Write a Scala method that receives a two Sets integers and return another Set which is an intersection of the two sets. (8 marks)

Q4

- a) You are given a string object `val mystr="This question is cool for all those who have read and made sure they understand String class methods in Scala"`.
 - i. Which string method will you use to return the number of characters in the string? (1 mark)
 - ii. Which string method will you use to search for a token from the left end of the string? (1 mark)
 - iii. Which string method will you use to search for a token from the right end of the string? (1 mark)

- iv. Which string method will you use to help you know the number of tokens in the string? **(1 mark)**
- v. Write a scala program that will print out the number of white space in the string and replace the white spaces with the character “*”? **(6marks)**
- b) Write a scala method that will receive the string above and then return the largest token in the string. Illustrate the working of this method in a program. **(10 marks)**

Q5

- a) Discuss in detail Spark architectural components. **(10 marks)**
- b) Consider a the file myfile in the current directory that contains the text. **”Apache Spark specifies different Api’s such as Java, R, Scala, Python that can be used to write programs to analyze data stored within the cluster environment or in the local file system. The core programming data structure for Spark is the RDDs”.**
- i. Create an RDD **txtfile** using file **myfile** above for content **(2 marks)**
- ii. Give the Spark code that will use the RDD you create above to Create a flat map for splitting each line into words, then read each word as a key with a value ‘1’ using map function then reduce the keys by adding values of similar keys. **(6 marks)**
- iii. Give the code that can now save the result in the count above to the file **resultFile** in the current directory. **(2marks)**

END