Foundations

Part-I

Q1 Answer the following questions:
(2 x 10)

a) What are the libraries used for data visualization in Python?

b) What is the difference between a series and a dataframe in Pandas?

c) What is the difference between lists and tuples in Python?

d) What are the differences between indexing and slicing?

e) What is a compound datatype?

f) What the difference is between is and '==' in Python?

g) What's the difference between / and // in Python?

h) What is data cleansing?

i) Define HDFS.

j) What is Big Data platform?

Part-II

Q2 Only Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)   (6 x 8)

a) What built-in data types are used in Python?

b) How is a negative index used in Python?

c) What is tuple unpacking? Why is it important?

d) How would you find duplicate values in a dataset for a variable in Python? Explain with example.

e) How do you convert integers to strings? Discuss with one example.

f) What's the difference between mutable and immutable objects in Python?

g) What is Big Data? Why we need to analyze Big Data?

h) List down the entities of YARN.

i) State the purpose of Hadoop Pipes.

j) Write the characteristics of Big Data.

k) What is cluster analysis? Explain.

l) What is a Decision Tree?

Part-III

Only Long Answer Type Questions (Answer Any Two out of Four)

Q3 How are data analysis libraries used in Python? What are some of the most common libraries? Write in brief.
(16)

Q4 Write a program in Python: Given a string, return the first recurring character in it, or "None" if there is no recurring character.

Example:

input = "interviewquery"

output = "i"
(16)

Q5 Why Hadoop is called a Big Data technology? Explain how it supports Big Data? Illustrate on how Cloud and Big Data related to each other.

(16)

Q6 Explain in detail about storage considerations in big data.
(16)