

Question Paper

Exam Date & Time: 29-Jun-2022 (02:00 PM - 05:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Second Semester Master of Engineering - ME (Embedded Systems / Cloud Computing) Degree Examination - June 2022
Big Data and Data Visualization Elective -2 [BDA 5232]

Marks: 100

Duration: 180 mins.

Wednesday, June 29, 2022

Answer all the questions.

- 1) Illustrate the working of Distributed File System with the neat diagram (TLO 1.1) (10 Marks) (10)
- 2) Explain the 4 different types of data can be modelled with nosql (TLO 3.2) (10 Marks) (10)
- 3) Explain five types of web scraping techniques. (TLO: 7.1) (2x5 Marks) (10)
- 4) Explain various steps involved in Web scraping framework and illustrate it using Beautiful Soup. (10)
(URI: www.quotesoscrape.com) (TLO: 7.1) (10 Marks)
- 5) Illustrate scrapy framework. Write a spyder script to extract information about an product having attributes: name, cost, manufacturer, rating and store it in a Jason file. URI: www.shopping-TV.com. (10)
Note: website is not allowed to scrape using bots or script, mention the required steps to overcome this problem. (TLO: 7.1) (10 Marks)
- 6) Write a program to create 3x4 array, delete 3rd column and insert new column in the 3rd column. (10)
Print the intermediate results. (TLO: 7.1) (10 Marks)
- 7) Write a regular expression for matching passwords under following consideration: (10)

6 to 12 characters in length
Must have at least one uppercase letter
Must have at least one lower case letter
Must have at least one digit
Should contain other characters (TLO 7.1) (2x5 Marks)
- 8) Assume that "power.csv" contains information about power production. Fields in csv are "Type of fuel", "Country", "Place", "Mega Watt", and "Cost per Unit". Write the script using pandas, to perform following tasks. (10)

List different types of fuels used.
Which country generates more power using 'Coal' as fuel used.
Which country generates maximum power (including all types of fuels)
List the place name that produces maximum power. (TLO 7.1) (2.5x4 Marks)
- 9) Explain how P-value and confidence intervals play a role in drawing the inference out of the data set (10)
(TLO 2.1) (10 Marks)
- 10) Write a Pandas program to replace the missing values with the most frequent values present in each column of a given dataframe. Test data consists of following attributes: ord_no, purch_amt, sale_amt, ord_date, customer_id, salesman_id. (TLO 7.1) (10 Marks) (10)

-----End-----