

# Question Paper

Exam Date & Time: 16-Aug-2022 (10:00 AM - 01:00 PM)



## MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences, Manipal  
Second Semester Master of Engineering - ME (Embedded Systems) Degree Examination (Makeup) - August 2022

### Big Data and Data Visualization Elective -2 [BDA 5232]

Marks: 100

Duration: 180 mins.

Tuesday, August 16, 2022

Answer all the questions.

- 1) Discuss the components of Big Data Architecture with the Diagram (TLO 1.1) (10 Marks) (10)
- 2) Explain how P-value and confidence intervals play a role in drawing the inference out of the data set (TLO 2.1) (10 Marks) (10)
- 3) Explain five types of web scraping techniques. (TLO: 7.1) (2x5 Marks) (10)
- 4) `html_doc = ""< html>< head>< title>Student Details< /title>< /head>` (10)

#### Student Details

contact information

`Ramesh< /a.`

`id="link2">9876543210` and

`id="link3">Ramesh.m@xyz .com;`

and they lived at the bottom of a well.

""

Using beautiful soup, write the script to extract following details

To display the title of the HTML file

To display the Name, mobile number and email

To display the email if the mail ID is valid (TLO: 7.1) (10 Marks)

- 5) Using scrapy tool, write a spyder script to extract quotes, author and tags from the URL: `http://quotes.toscrape.com` and store it in a csv file. (TLO: 7.1) (10 Marks) (10)
- 6) Write a regular expression for matching URL under following consideration: (10)

Must start with http or https or ftp followed by `://`

Must match a valid domain name

Could contain a port specification (`http://www.sitepoint.com:80`)

Could contain digit, letter, dots, hyphens, forward slashes, multiple times (TLO: 7.1) (2.5+2.5+2.5+2.5 Marks)

- 7) 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)
- 8) Differentiate special indexing operator `loc` and `iloc` with example. (TLO 7.1) (10 Marks) (10)
- 9) Explain the difference of nosql with sql databases with respect to the properties. (TLO 3.2) (10 Marks) (10)
- 10) Explain the 4 different characteristics of the Big Data (TLO 1.1) (10 Marks) (10)

-----End-----