QPC: RA20BTECH893 AY 20 Reg. No | | | | | | | | | | |

GIET UNIVERSITY, GUNUPUR – 765022

B. Tech (Eight Semester – Regular) Examinations, April– 2024

BOEEL8020 /BOEEE8021 – Big Data Analytics

(EE & EEE)

Time: 3 hrs Maximum: 70 Marks

The figures in the right hand margin indicate marks.

PART – A: (Multiple Choice Questions) **(1 x 10 =10 Marks)**

Q.1. Answer *ALL* questions [CO#] [PO#]

a. In Big Data environments, Velocity refers – CO1 PO1

i. Data can arrive at fast speed
ii. Enormous datasets can accumulate within very short periods of time
iii. Velocity of data translates into the amount of time it takes for the data to be processed
iv. All of the mentioned above

b. Data that does not conform to a data model or data schema is known as _____. CO1 PO2

i. Structured data
ii. Unstructured data
iii. Semi-structured data
iv. All of the mentioned above

c. Amongst which of the following can be considered as the main source of unstructured data. CO1 PO1

i. Twitter
ii. Facebook
iii. Webpages
iv. All of the mentioned above

d. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the JobTracker. CO2 PO2

i. Map Reduce
ii. Mapper
iii. Task Tracker
iv. Job Tracker

e. What type of data analysis does intelligent data analysis usually involve? CO2 PO1

i. Qualitative analysis
ii. Quantitative analysis
iii. Both qualitative and quantitative analysis
iv. Neither qualitative nor quantitative analysis

f. What is the most common type of analytics? CO2 PO1

i. Descriptive
ii. Predictive
iii. Structured
iv. Unstructured

g. Which of the following is an example of data analysis? CO3 PO2

i. Generating a report from a database
ii. Creating a graph to show trends
iii. Writing a summary of the data
iv. All of the above

h. Data reporting typically requires CO3 PO1

i. Advanced software programs
ii. Extensive data cleaning
iii. Manual manipulation
iv. All of the above

i. Sampling distribution is a probability distribution of a ---------- CO4 PO1

i. sample statistic
ii. population parameter
iii. population statistic
iv. sample parameter

j. The standard deviation of the sampling distribution of a statistic is equal to CO4 PO1

i. the standard deviation of the population
ii. the standard deviation of the sample
iii. the variance of the population
iv. the variance of the sample

PART – B: (Short Answer Questions) **(2 x 10=20 Marks)**

Q.2. Answer ***ALL*** questions	[CO#]	[PO#]
a. What is data?	CO1	PO2
b. What are the basics of a big data platform?	CO1	PO2
c. What is the purpose of Cloudera Manager?	CO1	PO1
d. How do conventional computing and intelligent computing differ in terms of their capabilities?	CO2	PO2
e. What analytical processes and tools can be used to identify customer needs and preferences?	CO2	PO1
f. What are the advantages of using Bootstrap for web development?	CO3	PO1
g. Is there any correlation between the observed data and the predicted data?	CO3	PO2
h. Are there any other factors that could have contributed to the prediction error?	CO	PO1
i. What are the key features of the Map Reduce model?	CO4	PO1
j. What is the difference between descriptive and inferential statistics?	CO4	PO2

PART – C: (Long Answer Questions) **(10 x 4=40 Marks)**

Answer ***ALL*** questions	Marks	[CO#]	[PO#]
3. a. What are the key challenges associated with collecting, storing, and processing large datasets in order to gain meaningful insights from big data analytics?	5	CO1	PO3
b. What are the key features of a big data platform?	5	CO1	PO2
(OR)			
c. What is the purpose of Apache Hadoop and how does it help organizations in the modern data-driven world?	5	CO1	PO2
d. How do structured data help organizations gain insights into customer behavior and preferences?	5	CO1	PO1
4. a. What are the major challenges inherent in conventional systems of computing and data storage?	5	CO2	PO
b. What are the benefits of using intelligent data analysis to create more accurate predictive models for decision making?	5	CO2	PO1
(OR)			
c. What types of problems are best suited to using the IDA algorithm?	5	CO2	PO3
d. What techniques can be used to convert data from one format to another?	5	CO2	PO2
5. a. How does analysis help to inform decision-making?	5	CO3	PO1
b. What challenges are associated with implementing and using modern analytics tools?	5	CO3	PO2
(OR)			
c. What types of data can be analyzed using modern analytic tools?	5	CO3	PO2
d. How do cloud-based big data tools integrate with existing IT infrastructure?	5	CO3	PO3
6. a. How is a sampling distribution used to make inferences about a population?	5	CO4	PO1
b. What is the difference between a Type I and Type II error?	5	CO4	PO3
(OR)			
c. How is an inferential statistic used to draw conclusions from data?	5	CO4	PO2
d. How can Bootstrap and Cross-Validation be used to evaluate the accuracy of a model?	5	CO4	PO2

--- End of Paper ---