Code No: **R194205E** **R19** **Set No. 1**

**IV B.Tech II Semester Regular Examinations, April – 2023**

# BIG DATA ANALYTICS

**(Elective V for CSE & IT, Open Elective for Other Branches)**

**Time: 3 hours** **Max. Marks: 75**

***Answer any FIVE Questions***
***ONE Question from Each unit***
***All Questions Carry Equal Marks***

*****

## UNIT I

| | | | |
|---|---|---|---|
| 1 | a) | List and explain different analytic processes. | [7] |
| | b) | Write down the characteristics of Big Data Applications. | [8] |
| | | (OR) | |
| 2 | | Discuss the following in detail. | |
| | | a) Conventional challenges in big data | [7] |
| | | b) Nature of Data | [8] |

## UNIT II

| | | | |
|---|---|---|---|
| 3 | a) | Explain Filtering a stream in detail | [7] |
| | b) | Explain stream data model and architecture. | [8] |
| | | (OR) | |
| 4 | a) | Discuss Real Time Analytics platform application for Stock Market predictions. | [7] |
| | b) | Illustrate Stream Processing Model | [8] |

## UNIT III

| | | | |
|---|---|---|---|
| 5 | a) | Draw the architecture of HDFS and explain its components. | [7] |
| | b) | How Hadoop streaming is suited with text processing explain. | [8] |
| | | (OR) | |
| 6 | a) | Discuss the various types of map reduce & its formats. | [7] |
| | b) | Explain various phases of Map Reduce job with an example. | [8] |

## UNIT IV

| | | | |
|---|---|---|---|
| 7 | a) | Explain the key components of PIG architecture. | [7] |
| | b) | Write short notes on: i) HBase ii) zookeeper | [8] |
| | | (OR) | |
| 8 | a) | How will you query the data in HIVE? | [7] |
| | b) | Explain two execution types or modes in PIG. | [8] |

## UNIT V

| | | | |
|---|---|---|---|
| 9 | a) | Distinguish between Regression and Classification. | [7] |
| | b) | Explain the importance of predictive analytics for improving Business. | [8] |
| | | (OR) | |
| 10 | a) | How do businesses use Regression Analysis? | [7] |
| | b) | Explain in detail about Multiple Linear Regression technique. | [8] |

Code No: **R194205E** **R19** **Set No. 2**

**IV B.Tech II Semester Regular Examinations, April – 2023**

# BIG DATA ANALYTICS

**(Elective V for CSE & IT, Open Elective for Other Branches)**

**Time: 3 hours** **Max. Marks: 75**

***Answer any FIVE Questions***
***ONE Question from Each unit***
***All Questions Carry Equal Marks***

*****

## UNIT I

| | | | |
|---|---|---|---|
| 1 | a) | Discuss in detail about Intelligent Data Analysis. | [7] |
| | b) | Discuss various sources of Big Data and the significance of Big Data Analytics. | [8] |
| | | (OR) | |
| 2 | a) | What is Bigdata? Describe the main features of a big data in detail. | [7] |
| | b) | What are the various forms of Big Data? Explain the suitable Data Exploration techniques. | [8] |

## UNIT II

| | | | |
|---|---|---|---|
| 3 | a) | What is meant by stream Processing? Explain how to count distinct elements in a stream. | [7] |
| | b) | Explain the concept of Mining data streams and applying filters. | [8] |
| | | (OR) | |
| 4 | | Explain the following: (i) Decaying Windows | [7] |
| | | (ii) RTAP Applications | [8] |

## UNIT III

| | | | |
|---|---|---|---|
| 5 | a) | Explain in detail about HDFS. | [7] |
| | b) | Write in detail the concept of developing the Map Reduce Application. | [8] |
| | | (OR) | |
| 6 | a) | Mention the configuration of Hadoop Cluster using Fully Distributed Mode in detail. | [7] |
| | b) | Name different configuration files in Hadoop. | [8] |

## UNIT IV

| | | | |
|---|---|---|---|
| 7 | a) | Give a detail note on HBASE. | [7] |
| | b) | Illustrate the Architecture of PIG. | [8] |
| | | (OR) | |
| 8 | a) | How to create and Manage the database and tables using Hive. | [7] |
| | b) | Write a brief notes on distributed modes of running PIG Scripts. | [8] |

## UNIT V

| | | | |
|---|---|---|---|
| 9 | a) | Explain about Predictive Analysis. | [7] |
| | b) | Illustrate Simple Linear Regression. | [8] |
| | | (OR) | |
| 10 | a) | Write the importance of Regression in Data Science and Data Analytics. | [7] |
| | b) | How to interpret coefficients of Multiple Linear Regression? Explain. | [8] |

Code No: **R194205E** **R19** **Set No. 3**

**IV B.Tech II Semester Regular Examinations, April – 2023**

# BIG DATA ANALYTICS

**(Elective V for CSE & IT, Open Elective for Other Branches)**

**Time: 3 hours** **Max. Marks: 75**

***Answer any FIVE Questions***
***ONE Question from Each unit***
***All Questions Carry Equal Marks***

*****

## UNIT I

| | | | |
|---|---|---|---|
| 1 | a) | Explain in detail about Nature of Data and its applications. | [7] |
| | b) | List out the challenges of Conventional Systems. | [8] |
| | | (OR) | |
| 2 | | Define the following:<br>a) Intelligent Data Analytics<br>b) Analysis Vs Reporting. | [7]<br>[8] |

## UNIT II

| | | | |
|---|---|---|---|
| 3 | a) | Explain with a neat diagram about Stream data model and its Architecture. | [7] |
| | b) | Explain the real time application of stream computing. | [8] |
| | | (OR) | |
| 4 | a) | Illustrate Stream Computing. | [7] |
| | b) | Explain the usecases of Real Time Sentiment Analysis. | [8] |

## UNIT III

| | | | |
|---|---|---|---|
| 5 | a) | List and explain the important features of Hadoop. | [7] |
| | b) | Explain the differences between old and new Java MapReduce API. | [8] |
| | | (OR) | |
| 6 | a) | Explain the architecture of Building blocks of Hadoop. | [7] |
| | b) | Describe the anatomy of Map Reduce program and discuss various types of failures in running a Map Reduce job. | [8] |

## UNIT IV

| | | | |
|---|---|---|---|
| 7 | a) | Explain various types of Data processing operators in pig. | [7] |
| | b) | Describe two modes for running scripts in Pig | [8] |
| | | (OR) | |
| 8 | a) | Explain the creating, dropping and altering databases using Apache Hive. | [7] |
| | b) | Discuss in detail about windowing in HiveQL. | [8] |

## UNIT V

| | | | |
|---|---|---|---|
| 9 | a) | What are the interpretations of the coefficients in the multiple linear regression models? | [7] |
| | b) | How do you interpret statistically significant coefficients? | [8] |
| | | (OR) | |
| 10 | a) | Discuss in detail about visualization in big data analytics. | [7] |
| | b) | Illustrate Multiple Linear Regression. | [8] |

Code No: **R194205E** **R19** **Set No. 4**

**IV B.Tech II Semester Regular Examinations, April – 2023**

# BIG DATA ANALYTICS

**(Elective V for CSE & IT, Open Elective for Other Branches)**

**Time: 3 hours** **Max. Marks: 75**

***Answer any FIVE Questions***
***ONE Question from Each unit***
***All Questions Carry Equal Marks***

*****

## UNIT I

| | | | |
|---|---|---|---|
| 1 | a) | Distinguish between conventional computing and intelligent computing. | [7] |
| | b) | List out and Explain the Features of Big Data Platforms/frameworks. | [8] |
| | | (OR) | |
| 2 | a) | Discuss in detail about Structured Unstructured Data and Semi-structured Data. | [7] |
| | b) | Define Big Data. Describe its characteristics. | [8] |

## UNIT II

| | | | |
|---|---|---|---|
| 3 | a) | Explain the Data streaming concept in detail. | [7] |
| | b) | Write a short note on Decaying Window Algorithm. | [8] |
| | | (OR) | |
| 4 | a) | Explain the different applications of data streams in detail. | [7] |
| | b) | What is Real Time Analytics? Discuss their technologies in detail | [8] |

## UNIT III

| | | | |
|---|---|---|---|
| 5 | a) | Explain the anatomy of write operation in HDFS. | [7] |
| | b) | Explain the map reduce data flow with single reduce and multiple reduce. | [8] |
| | | (OR) | |
| 6 | a) | Write a Java program to implement the word count program using Map Reduce paradigm. | [7] |
| | b) | Explain the role of combiner and partitioner phases in Map Reduce job. | [8] |

## UNIT IV

| | | | |
|---|---|---|---|
| 7 | a) | What is HiveQL? Explain its features. | [7] |
| | b) | What is Zookeeper explain its features with applications. | [8] |
| | | (OR) | |
| 8 | a) | Give a brief note on Querying Data in Hive. | [7] |
| | b) | What is Apache PIG? Give its features, running modes and applications. | [8] |

## UNIT V

| | | | |
|---|---|---|---|
| 9 | a) | How to interpret p values and coefficients in regression analysis? | [7] |
| | b) | Explain Cross-Validation in Multiple linear regression. | [8] |
| | | (OR) | |
| 10 | a) | Is multiple linear regression predictive analytics? justify. | [7] |
| | b) | Give a brief note on Model Selection and Stepwise Regression. | [8] |