| Reg. No. | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| Question Paper Code | 13482 |
|---|---|

**B.E. / B.Tech. - DEGREE EXAMINATIONS, APR / MAY 2025**

Fifth Semester

**Information Technology**

**20ITPC502 - BIG DATA ESSENTIALS**

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

**PART - A (MCQ) (10 × 1 = 10 Marks)**

Answer ALL Questions

| | | Marks | K – Level | CO |
|---|---|---|---|---|
| 1. | Which of the following is a use case of Big Data in marketing?<br>(a) Crop yield prediction (b) Weather monitoring<br>(c) Customer behaviour analysis (d) Space exploration | 1 | K1 | CO1 |
| 2. | Which of the following is NOT a characteristic of Big Data?<br>(a) Volume (b) Variety (c) Velocity (d) Uniformity | 1 | K1 | CO1 |
| 3. | What is the primary motivation behind the development of Hadoop?<br>(a) Real-time data processing (b) High availability of storage<br>(c) Distributed computing on large datasets (d) In-memory data processing | 1 | K1 | CO2 |
| 4. | What is a key characteristic of Hadoop Streaming?<br>(a) Real-time processing (b) Use of MapReduce framework<br>(c) Batch processing only (d) Integration with SQL databases | 1 | K1 | CO2 |
| 5. | Which phase of a MapReduce job involves transferring intermediate data over the network from mappers to reducers?<br>(a) Map phase (b) Reduce phase (c) Shuffle phase (d) Sort phase | 1 | K1 | CO3 |
| 6. | What happens when a task fails in a MapReduce job?<br>(a) The job is terminated immediately (b) The task is restarted on the same node<br>(c) The task is retried on a different node (d) The output is discarded | 1 | K1 | CO3 |
| 7. | What is Apache Pig primarily used for<br>(a) Storing large datasets (b) Processing large datasets<br>(c) Transmitting large datasets (d) Querying large datasets | 1 | K1 | CO4 |
| 8. | How many execution modes does Pig support?<br>(a) One (b) Two (c) Three (d) Four | 1 | K1 | CO4 |
| 9. | Which Spark component allows for real-time stream processing?<br>(a) Spark SQL (b) Spark Streaming (c) MLib (d) GraphX | 1 | K1 | CO5 |
| 10. | What does CUDA stand for?<br>(a) Compute Unified Device Architecture (b) Central Unified Data Architecture<br>(c) Cluster Unit Data Architecture (d) Compute Unit Data Algorithm | 1 | K1 | CO6 |

**PART - B (12 × 2 = 24 Marks)**

Answer ALL Questions

| | | Marks | K – Level | CO |
|---|---|---|---|---|
| 11. | List the risks of Big Data. | 2 | K1 | CO1 |
| 12. | State one best practice for managing big data analytics programs. | 2 | K1 | CO1 |
| 13. | Draw the structure of sequence file format. | 2 | K1 | CO2 |
| 14. | Define Data Digest in Hadoop. | 2 | K1 | CO2 |
| 15. | State the General form of Map and Reduce functions. | 2 | K1 | CO3 |
| 16. | Identify two types of failures that can occur in a MapReduce job and describe how they can be handled. | 2 | K2 | CO3 |
| 17. | Compare Hive with Traditional databases. | 2 | K2 | CO4 |

| | | | | | |
|---|---|---|---|---|---|
| 18. | List few HDFS commands that is used in Pig Grunt. | *2* | *K1* | *CO4* |
| 19. | How is data distributed across nodes in a Spark cluster? | *2* | *K2* | *CO5* |
| 20. | Describe Spark's lazy evaluation mechanism. | *2* | *K2* | *CO5* |
| 21. | How does GPU computing differ from CPU computing? | *2* | *K2* | *CO6* |
| 22. | What is the purpose of thread divergence in CUDA? | *2* | *K2* | *CO6* |

**PART - C (6 × 11 = 66 Marks)**

Answer ALL Questions

| | | | | | | |
|---|---|---|---|---|---|---|
| 23. | a) | Discuss in briefly about the concepts of Big Data Technologies. | *11* | *K2* | *CO1* |
| | | **OR** | | | |
| | b) | Illustrate Big Data Analytics process and its types. List out the benefits and challenges. | *11* | *K2* | *CO1* |
| 24. | a) | Explain in briefly about Hadoop Distributed File System. | *11* | *K2* | *CO2* |
| | | **OR** | | | |
| | b) | Discuss in detail about the Components of Hadoop and Hadoop Archives. | *11* | *K2* | *CO2* |
| 25. | a) | Explain about developing a MapReduce application and how job runs on MapReduce. | *11* | *K2* | *CO3* |
| | | **OR** | | | |
| | b) | Illustrate in detail about Job Scheduling, task execution in YARN. | *11* | *K2* | *CO3* |
| 26. | a) | Compare and Contrast HIVE, PIG, HBASE, also list their benefits in detail. | *11* | *K2* | *CO4* |
| | | **OR** | | | |
| | b) | Discuss briefly about HBASE Concepts with necessary diagrams. | *11* | *K2* | *CO4* |
| 27. | a) | Illustrate the execution process of a Spark job, including job stages and tasks. | *11* | *K2* | *CO5* |
| | | **OR** | | | |
| | b) | Discuss how Spark optimizes the performance of distributed data processing using DAGs. | *11* | *K2* | *CO5* |
| 28. | a) | Discuss the differences between CPU and GPU architectures in terms of processing power and parallelism. | *11* | *K2* | *CO6* |
| | | **OR** | | | |
| | b) | Describe the challenges and solutions in handling memory access patterns in CUDA. | *11* | *K2* | *CO6* |