**In the context of Data Science Big Data, generate 28 relatively straightforward random multiple-choice questions on Hadoop and MapReduce that would suit a first year Data Science university student.**

**Put the answers at the end.**

**16 x 2**

**Question 1: What does Hadoop primarily deal with?**

**A. Image processing**

**B. Natural language processing**

**C. Big data processing**

**D. Voice recognition**

**Question 2: Which of the following is a component of the Hadoop ecosystem?**

**A. Oracle**

**B. Hive**

**C. MySQL**

**D. MongoDB**

**Question 3: Hadoop Distributed File System (HDFS) is designed for:**

**A. High-speed data access**

**B. High-latency data access**

**C. Fault tolerance**

**D. Data encryption**

**Question 4: What is the primary programming language used for Hadoop?**

**A. Java**

**B. Python**

**C. R**

**D. JavaScript**

**Question 5: What is the purpose of a NameNode in HDFS?**

**A. Store actual data**

**B. Manage the file system metadata**

**C. Perform MapReduce operations**

**D. Store intermediate data**

**Question 6: Which of the following is NOT a key feature of Hadoop?**

**A. Scalability**

**B. Data encryption**

**C. Fault tolerance**

**D. Horizontal scalability**

**Question 7: What is MapReduce in Hadoop?**

**A. A storage system**

**B. A data processing framework**

**C. A machine learning algorithm**

**D. A data visualization tool**

**Question 8: What is the function of the map phase in MapReduce?**

**A. Combine intermediate values**

**B. Store intermediate values**

**C. Process input data and create key-value pairs**

**D. Distribute data across nodes**

**Question 9: What is the function of the reduce phase in MapReduce?**

**A. Combine intermediate values**

**B. Store intermediate values**

**C. Process input data and create key-value pairs**

**D. Distribute data across nodes**

**Question 10: Which of the following is a benefit of using Hadoop for big data processing?**

**A. Real-time processing**

**B. Cost-effectiveness**

**C. Data encryption**

**D. Support for structured data only**

**Question 11: What type of data can Hadoop process?**

**A. Structured data only**

**B. Unstructured data only**

**C. Both structured and unstructured data**

**D. Semi-structured data only**

**Question 12: What is the primary purpose of the Hadoop YARN component?**

**A. Data storage**

**B. Resource management**

**C. Data processing**

**D. Data encryption**

**Question 13: What does the "R" in YARN stand for?**

**A. Resource**

**B. Reducer**

**C. Replication**

**D. Reliability**

**Question 14: Which of the following is an alternative to Hadoop MapReduce for data processing?**

**A. Pig**

**B. Oozie**

**C. Zookeeper**

**D. Spark**

**Question 15: What is the default replication factor in HDFS?**

**A. 1**

**B. 2**

**C. 3**

**D. 4**

**Question 16: In Hadoop, what is the purpose of DataNodes?**

**A. Store actual data**

**B. Manage the file system metadata**

**C. Perform MapReduce operations**

**D. Store intermediate data**

**In the context of Data Analytics and Big Data, what does HDFS stand for?**

**Question 1: What does Hadoop mainly deal with?**

**a) Real-time processing**

**b) Batch processing**

**c) In-memory processing**

**d) Interactive processing**

**Question 2: Which of the following is not a Hadoop ecosystem component?**

**a) Hive**

**b) Pig**

**c) Impala**

**d) TensorFlow**

**Question 3: What does HDFS stand for?**

**a) Hadoop Data File System**

**b) Hadoop Distributed File System**

**c) Hadoop Dynamic File System**

**d) Hadoop Domain File System**

**Question 4: Which of the following is responsible for data storage in Hadoop?**

**a) MapReduce**

**b) HDFS**

**c) YARN**

**d) Spark**

**Question 5: What is the default replication factor in HDFS?**

**a) 1**

**b) 2**

**c) 3**

**d) 4**

**Question 6: Which of the following is not a valid HDFS operation?**

**a) mkdir**

**b) put**

**c) delete**

**d) sort**

**Question 7: Which of the following Hadoop components is responsible for resource management?**

**a) YARN**

**b) Hive**

**c) MapReduce**

**d) HBase**

**Question 8: What does MapReduce do in the Hadoop framework?**

**a) Data storage**

**b) Data processing**

**c) Resource management**

**d) Data querying**

**Question 9: What is the output of the Map phase in MapReduce?**

**a) Key-value pairs**

**b) Lists**

**c) Dictionaries**

**d) Arrays**

**Question 10: What is the input to the Reduce phase in MapReduce?**

**a) Key-value pairs**

**b) Lists**

**c) Dictionaries**

**d) Arrays**

**Question 11: Which programming language is primarily used to write MapReduce jobs?**

**a) Python**

**b) Java**

**c) Scala**

**d) R**

**Question 12: Which of the following is not a phase in the MapReduce process?**

**a) Shuffle**

**b) Merge**

**c) Sort**

**d) Scan**

**Question 13: What is the role of the combiner in MapReduce?**

**a) Combine input data**

**b) Combine output data**

**c) Aggregate intermediate map output**

**d) Aggregate final reduce output**

**Question 14: How many default partitions are created by the Hadoop MapReduce framework?**

**a) 1**

**b) 2**

**c) 3**

**d) 4**

**Question 15: Which of the following is a columnar storage file format in Hadoop?**

**a) Avro**

**b) Parquet**

**c) SequenceFile**

**d) ORC**

**Question 16: Which of the following tools allows SQL-like querying in Hadoop?**

**a) Pig**

**b) Hive**

**c) Sqoop**

**d) Flume**

**Answers:**

**1: C, 2: B, 3: C, 4: A, 5: B, 6: B, 7: B, 8: C, 9: A, 10: B, 11: C, 12: B, 13: A, 14: D, 15: C, 16: A**

**Answers:**

**1. b**

**2. d**

**3. b**

**4. b**

**5. c**

**6. d**

**7. a**

**8. b**

**9. a**

**10. a**

**11. b**

**12. d**

**13. c**

**14. b**

**15. b**

**16. b**