**In the context of Data Science Big Data, generate 13 relatively straightforward random multiple-choice questions on MapReduce that would suit a university student in the first year of a Data Science Masters degree.**

**Put the answers at the end.**

Sure, here are 13 multiple-choice questions about MapReduce:

1. What is the primary purpose of MapReduce?

A. Data visualization

B. Data warehousing

C. Real-time data processing

D. Distributed data processing

2. In the context of MapReduce, what are the two main functions?

A. Filter and Reduce

B. Map and Combine

C. Map and Reduce

D. Partition and Reduce

3. What is the role of the Map function in MapReduce?

A. Aggregating data

B. Transforming and filtering data

C. Storing data

D. Visualizing data

4. What is the role of the Reduce function in MapReduce?

A. Transforming and filtering data

B. Aggregating data

C. Storing data

D. Visualizing data

5. In MapReduce, which function is responsible for partitioning the intermediate key-value pairs?

A. Mapper

B. Partitioner

C. Reducer

D. Combiner

6. What is the purpose of the Combiner in a MapReduce job?

A. Data visualization

B. Data partitioning

C. Local aggregation of intermediate data

D. Data storage

7. What is the default number of Reduce tasks in a Hadoop MapReduce job?

A. 0

B. 1

C. 2

D. 3

8. What is the input data format for a MapReduce job?

A. CSV

B. JSON

C. Key-Value pairs

D. XML

9. Which of the following is NOT a phase in the MapReduce data processing model?

A. Shuffle

B. Partition

C. Sort

D. Distribute

10. In the context of MapReduce, what is the output of the Map function?

A. A list of key-value pairs

B. A single key-value pair

C. A single key with multiple values

D. A list of keys

11. What does the InputFormat class do in a MapReduce job?

A. Defines the input data format

B. Specifies the number of Reduce tasks

C. Splits the input data into chunks

D. Both A and C

12. Which of the following is a disadvantage of using MapReduce?

A. Scalability

B. Fault tolerance

C. Complexity

D. Data distribution

13. In a MapReduce job, where does the intermediate data get stored?

A. In the HDFS

B. In the Map task's local file system

C. In the Reduce task's local file system

D. In the memory of the Map task

Answers:

1. D. Distributed data processing

2. C. Map and Reduce

3. B. Transforming and filtering data

4. B. Aggregating data

5. B. Partitioner

6. C. Local aggregation of intermediate data

7. B. 1

8. C. Key-Value pairs

9. D. Distribute

10. A. A list of key-value pairs

11. D. Both A and C

12. C. Complexity

13. B. In the Map task's local file system

Question 1: What is the primary purpose of the MapReduce programming model?

A. Data visualization

B. Network programming

C. Distributed data processing

D. Web application development

Question 2: Which of the following is NOT a key component of the MapReduce framework?

A. Mapper

B. Reducer

C. Combiner

D. Interpreter

Question 3: What does the Map function in MapReduce do?

A. Aggregates intermediate values

B. Processes input data and generates key-value pairs

C. Combines the output of the Reduce function

D. Distributes data across multiple nodes

Question 4: What does the Reduce function in MapReduce do?

A. Splits input data into chunks

B. Processes and filters input data

C. Aggregates intermediate values for the same key

D. Distributes data across multiple nodes

Question 5: Which of the following tasks is MapReduce NOT suitable for?

A. Text processing

B. Real-time data processing

C. Log analysis

D. Graph processing

Question 6: In the MapReduce framework, what is the purpose of the Shuffle phase?

A. Distributing the input data

B. Sorting and transferring intermediate data from Mappers to Reducers

C. Combining intermediate values

D. Aggregating final results

Question 7: In MapReduce, what is the role of the Combiner?

A. Pre-aggregating intermediate data locally on the Mapper node

B. Distributing input data across multiple nodes

C. Sorting and transferring intermediate data between nodes

D. Combining the output of the Reduce function

Question 8: What is the output of the MapReduce process?

A. A single aggregated value

B. A sorted list of key-value pairs

C. An unsorted list of key-value pairs

D. A set of clusters

Question 9: Which of the following programming languages can be used to write MapReduce jobs?

A. Java

B. Python

C. C++

D. All of the above

Question 10: What is the Hadoop Distributed File System (HDFS) used for in the context of MapReduce?

A. Storing input data and output results

B. Executing the Map and Reduce functions

C. Balancing the load between nodes

D. Providing a real-time query interface

Question 11: Which of the following is NOT a key feature of the MapReduce framework?

A. Scalability

B. Fault tolerance

C. Data replication

D. Real-time processing

Question 12: What is the primary advantage of using the MapReduce framework for data processing tasks?

A. It provides a simple and easy-to-use programming model

B. It can process data in real-time

C. It can analyze data in a single-node environment

D. It can automatically visualize the processed data

Question 13: In a MapReduce job, which of the following is true about the relationship between Mappers and Reducers?

A. Mappers and Reducers can run concurrently on the same data

B. Mappers always run before Reducers

C. Reducers always run before Mappers

D. Mappers and Reducers are not related to each other

Answers:

1. C

2. D

3. B

4. C

5. B

6. B

7. A

8. B

9. D

10. A

11. D

12. A

13. B