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

**Put the answers at the end.**

\*\*Question 1:\*\* What is Hadoop primarily used for in the field of Big Data?

A) Web development

B) Machine Learning

C) Distributed computing and storage

D) Mobile App development

\*\*Question 2:\*\* Which of the following is NOT a core component of the Hadoop ecosystem?

A) HDFS

B) MapReduce

C) Hive

D) YARN

\*\*Question 3:\*\* What does HDFS stand for in the Hadoop ecosystem?

A) Hadoop Data File System

B) Hadoop Distributed File System

C) Hadoop Data Framework System

D) Hadoop Distributed Framework System

\*\*Question 4:\*\* Which of the following is a NoSQL database in the Hadoop ecosystem?

A) Hive

B) HBase

C) Pig

D) Spark

\*\*Question 5:\*\* What is the primary function of Hadoop YARN?

A) Data storage

B) Data processing

C) Job scheduling

D) Data visualization

\*\*Question 6:\*\* Which of the following tools is used for data ingestion in Hadoop?

A) Flume

B) Sqoop

C) Both A and B

D) None of the above

\*\*Question 7:\*\* What is MapReduce primarily used for in Hadoop?

A) Data storage

B) Data transformation

C) Job scheduling

D) Data processing

\*\*Question 8:\*\* Which of the following tools is used for data warehousing in the Hadoop ecosystem?

A) Flume

B) HBase

C) Hive

D) Spark

\*\*Question 9:\*\* What is the primary language used in Hive?

A) Python

B) SQL

C) Java

D) R

\*\*Question 10:\*\* What does the term "block" refer to in HDFS?

A) A cluster of data nodes

B) A unit of storage

C) A type of error

D) A programming structure

\*\*Question 11:\*\* Which component of Hadoop is responsible for storing and replicating data blocks?

A) YARN

B) HBase

C) HDFS

D) Hive

\*\*Question 12:\*\* What is the default size of a data block in HDFS?

A) 32 MB

B) 64 MB

C) 128 MB

D) 256 MB

\*\*Question 13:\*\* What is the purpose of Apache Zookeeper in the Hadoop ecosystem?

A) Data warehousing

B) Data ingestion

C) Coordinating distributed processing

D) Data visualization

\*\*Answers:\*\*

1) C) Distributed computing and storage

2) C) Hive

3) B) Hadoop Distributed File System

4) B) HBase

5) C) Job scheduling

6) C) Both A and B

7) D) Data processing

8) C) Hive

9) B) SQL

10) B) A unit of storage

11) C) HDFS

12) C) 128 MB

13) C) Coordinating distributed processing

Sure, here are 13 multiple-choice questions about the Apache Hadoop ecosystem:

1. What is Hadoop primarily used for?

A. Web Development

B. Data Visualization

C. Distributed Processing of Large Data Sets

D. Software Development

2. Which programming language is Hadoop written in?

A. Python

B. Java

C. C++

D. Scala

3. What are the core components of Hadoop?

A. HBase and ZooKeeper

B. HDFS and MapReduce

C. Pig and Hive

D. Spark and Storm

4. What does HDFS in the Hadoop ecosystem stand for?

A. Hadoop Data File System

B. Hadoop Distributed File System

C. Hadoop Data Framework System

D. Hadoop Distributed Framework System

5. Which component of the Hadoop ecosystem is used for data warehousing?

A. MapReduce

B. HDFS

C. Hive

D. HBase

6. What is Apache Pig primarily used for in the Hadoop ecosystem?

A. Data warehousing

B. Distributed storage

C. Data processing

D. Data visualization

7. What does YARN stand for in the context of Hadoop?

A. Yet Another Resource Node

B. Yet Another Resource Notation

C. Yet Another Resource Negotiator

D. Yet Another Runtime Node

8. What is the role of ZooKeeper in the Hadoop ecosystem?

A. Data Processing

B. Data Visualization

C. Data Warehousing

D. Coordination of Distributed Systems

9. What is the purpose of Apache Storm in the Hadoop ecosystem?

A. Batch processing

B. Real-time data processing

C. Data storage

D. Data visualization

10. What is the primary language for scripting in Apache Pig?

A. Python

B. Pig Latin

C. Java

D. Scala

11. In the Hadoop ecosystem, what is HBase used for?

A. Real-time data processing

B. Data storage and retrieval in column-oriented databases

C. Data visualization

D. Data warehousing

12. What is the role of Apache Flume in the Hadoop ecosystem?

A. Data aggregation and movement to HDFS

B. Data visualization

C. Data warehousing

D. Real-time data processing

13. What type of data processing is MapReduce used for in Hadoop?

A. Real-time data processing

B. Batch processing

C. Data visualization

D. Data storage

Answers:

1. C. Distributed Processing of Large Data Sets

2. B. Java

3. B. HDFS and MapReduce

4. B. Hadoop Distributed File System

5. C. Hive

6. C. Data processing

7. C. Yet Another Resource Negotiator

8. D. Coordination of Distributed Systems

9. B. Real-time data processing

10. B. Pig Latin

11. B. Data storage and retrieval in column-oriented databases

12. A. Data aggregation and movement to HDFS

13. B. Batch processing