Intro to Hadoop

13 questions

1 point

1.
What does IaaS provide?

- Hardware Only
- Software On-Demand
- Computing Environment

1 point

2.
What does PaaS provide?

- Hardware Only
- Software On-Demand
- Computing Environment

1 point

3.
What does SaaS provide?

- Hardware Only
- Software On-Demand
- Computing Environment

---

1 point

4.
What are the two key components of HDFS and what are they used for?

- NameNode for block storage and Data Node for metadata.
- NameNode for metadata and DataNode for block storage.
- FASTA for genome sequence and Rasters for geospatial data.

---

1 point

5.
What is the job of the NameNode?

- Coordinate operations and assigns tasks to Data Nodes
- Listens from DataNode for block creation, deletion, and replication.

For gene sequencing calculations.

---

1 point

6.
What are the three steps to Map Reduce?

- Shuffle and Sort -> Reduce -> Map
- Shuffle and Sort -> Map -> Reduce
- Map -> Reduce -> Shuffle and Sort
- Map -> Shuffle and Sort -> Reduce

---

1 point

7.
What is a benefit of using pre-built Hadoop images?

- Quick prototyping, deploying, and validating of projects.
- Guaranteed hardware support.
- Quick prototyping, deploying, and guaranteed bug free.
- Less software choices to choose from.

---

1 point

8.
What is an example of open-source tools built for Hadoop and what does it do?

- Pig, for real-time and in-memory processing of big data.
- Giraph, for processing large-scale graphs.
- Zookeeper, analyze social graphs.
- Giraph, for SQL-like queries.

---

1 point

9.
What is the difference between low level interfaces and high level interfaces?

- Low level deals with storage and scheduling while high level deals with interactivity.
- Low level deals with interactivity while high level deals with storage and scheduling.

---

1 point

10.
Which of the following are problem sto look out for when you want to integrate your project with Hadoop?

- [ ] Random Data Access
- [ ] Advanced Alogrithms
- [ ] Infrastructure Replacement
- [ ] Task Level Parallelism

- [ ] Data Level Parallelism

---

1 point

11.
As covered in the slides, which of the following are the major goals of Hadoop?

- [ ] Latency Sensitive Tasks
- [ ] Handle Fault Tolerance
- [ ] Facilitate a Shared Environment
- [ ] Enable Scalability
- [ ] Provide Value for Data
- [ ] Optimized for a Variety of Data Types

---

1 point

12.
What is the purpose of YARN?

- ○ Allows various applications to run on the same Hadoop cluster.
- ○ Enables large scale data across clusters.
- ○ Implementation of Map Reduce.
