1.

What does IaaS provide?

Computing Environment

Hardware Only

Software On-Demand

2.

What does PaaS provide?

Hardware Only

Software On-Demand

Computing Environment

3.

What does SaaS provide?

Hardware Only

Computing Environment

Software On-Demand

4.

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

FASTA for genome sequence and Rasters for geospatial data.

NameNode for block storage and Data Node for metadata.

NameNode for metadata and DataNode for block storage.

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.

6.

What are the three steps to Map Reduce?

Shuffle and Sort -> Map -> Reduce

Shuffle and Sort -> Reduce -> Map

Map -> Shuffle and Sort -> Reduce

Map -> Reduce -> Shuffle and Sort

Correct

1 / 1 points

7.

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

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

8.

What are some examples of open-source tools built for Hadoop and what does it do?

- [ ] Zookeeper, analyze social graphs.
- [ ] Pig, for real-time and in-memory processing of big data.
- [ ] Giraph, for SQL-like queries.
- [x] Zookeeper, management system for animal named related components.

9.

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

- [x] 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.

10.

What is **NOT** a problem to look out for when you want to integrate your project with Hadoop?

Infrastructure Replacement

Advanced Algorithms

Random Data Access

Task Level Parallelism

Data Level Parallelism

11.

What is **NOT** a major goal of Hadoop as covered in the slides?

Enable Scalability

Optimized for a Variety of Data Types

Handle Fault Tolerance

Provide Value for Data

Facilitate a Shared Environment

Latency Sensitive Tasks

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.

13.

What are the two main components for a data computation framework that were described in the slides?

Resource Manager and Node Manager

Resource Manager and Container

Node Manager and Applications Master

Applications Master and Container

Node Manager and Container