

1.

What does IaaS provide?



Hardware Only

Correct Response

See [this video](#) to review.



Computing Environment



Software On-Demand

2.

What does PaaS provide?



Hardware Only



Computing Environment

Correct Response

See [this video](#) to review.



Software On-Demand

3.

What does SaaS provide?



Software On-Demand

Correct Response

See [this video](#) to review.



Computing Environment



Hardware Only

4.

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



NameNode for metadata and DataNode for block storage.

Correct Response

See [this video](#) to review.



FASTA for genome sequence and Rasters for geospatial data.



NameNode for block storage and Data Node for metadata.

5.

What is the job of the NameNode?



Coordinate operations and assigns tasks to Data Nodes

Correct Response

See [this video](#) to review.



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 -> Reduce -> Shuffle and Sort



Map -> Shuffle and Sort -> Reduce

Correct Response

See [this video](#) to review.

7.

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



Less software choices to choose from.



Quick prototyping, deploying, and validating of projects.

Correct Response

See [this video](#) to review.



Quick prototyping, deploying, and guaranteed bug free.



Guaranteed hardware support.

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.

Correct Response

See [this video](#) to review.



Zookeeper, analyze social graphs.



Giraph, for SQL-like queries.

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.

Correct Response

See [this video](#) to review.



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 Alogrithms



Random Data Access



Task Level Parallelism



Data Level Parallelism

Correct Response

See [this video](#) to review.

11.

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



Handle Fault Tolerance



Facilitate a Shared Environment



Latency Sensitive Taks

Correct Response

See [this video](#) to review.



Provide Value for Data



Optimized for a Variety of Data Types



Enable Scalability

12.

What is the purpose of YARN?



Allows various applications to run on the same Hadoop cluster.

Correct Response

See [this video](#) to review.



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 Container



Node Manager and Container



Node Manager and Applications Master



Resource Manager and Node Manager

Correct Response

See [this video](#) to review.



Applications Master and Container