



Knowledge check

6 minutes

Choose the best response for each of the following questions, then select **Check your answers**.

1. Which Hadoop component is used to manage resources on a Hadoop system? *

☒ YARN

✓ **Correct.** This is the component that is responsible for managing resources on a Hadoop cluster.

☐ Spark

☐ MapReduce

2. What type of file system does Apache Hadoop store data in? *

☐ It has no storage. Data is stored in memory.

☐ RDD

☒ HDFS

✓ **Correct.** HDFS is an acronym for Hadoop Distributed File System, which is the data store for a Hadoop system.

3. Which of the following options is a parallel processing framework? *

☐ Apache Hive

☒ Apache Spark

✓ **Correct.** Apache Spark supports in-memory processing, which helps boost the performance of big-data analytic applications.

☐ Apache Kafka

4. In a Hadoop cluster on HDInsight, which of the following nodes is responsible for processing data? *

☐ Head node

☒ Worker node

✓ Correct. Worker nodes are responsible for processing data in a cluster.

5. Which of the following scenarios for big data processing enable an organization to prepare big data for further analysis? *

☐ Data warehousing

☐ Data science

☒ Batch processing

✓ Correct. Organizations use batch-processing jobs to prepare big data for further analysis.

Next unit: Summary

Continue >
