



BIG DATA TOOLS (1/1 point)

Why is using traditional analysis tools for big data a poor choice?

- ☐ Storage is becoming less expensive.
- ☐ The time to read from a 1 TB drive is 3 hours.
- ☒ Big data does not fit on a single machine. ✓
- ☐ CPUs are getting faster and faster every year.

CHECK

SHOW ANSWER

MODERN BIG DATA HARDWARE (1/1 point)

Which of the following properties does modern hardware for big data have:

- ☐ Uses premium hardware
- ☒ Uses consumer grade hardware ✓
- ☐ Uses complex hardware
- ☒ Uses complex software ✓
- ☒ Easy to add capacity ✓



Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Modern big data hardware is based on less expensive, consumer grade hardware which makes it easy to grow capacity. Software is used to handle any problems, instead of hardware.

CHECK

HIDE ANSWER

USING DIVIDE AND CONQUER (1/1 point)

What are some of the challenges of using divide and conquer:

☒ Moving data is very expensive ✓

☐ Using a single machine is faster than multiple machines

☒ Having many machines means having to deal with many failures ✓

☒ Having many machines means having to deal with slow machines ✓

☐ Using hash tables for very large documents works well



Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

When using divide and conquer, you have to consider the network and data locality because moving data between machines is expensive. Even with a low per-machine failure rate, using many machines means that several will fail per day. As machines age, they may fail in ways that cause slow performance (e.g., a failing disk drive that retries each read or write operation multiple times before successfully completing).

CHECK

HIDE ANSWER

MAP REDUCE (1/1 point)

Which of the following problems does a MapReduce implementation handle?

☒ Recovering from machine failures ✓

☒ Shuffling data between the Map and Reduce functions ✓

☒ Running the Map and Reduce functions on many machines ✓

☒ Recovering from slow machines ✓



Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Map Reduce handles the execution of Map and Reduce functions on many machines including the shuffling of data between Map and Reduce functions. It also automatically recovers from both machine failures and slow machines.

CHECK

HIDE ANSWER

MAP REDUCE AND SPARK DIFFERENCES (1/1 point)

Spark is often faster than a traditional MapReduce implementation because:

☐ It sends less data over the network

☒ Results do not need to be written to disk ✓

☐ It detects machine failures more quickly

☐ It replicates the output of each task to recover from failures quickly



Note: Make sure you select all of the correct options—there may be more than one!

EXPLANATION

Spark keeps results in memory so they do not need to be written to disk.

CHECK

HIDE ANSWER

[About](#) [Blog](#) [News](#) [FAQs](#) [Contact](#) [Jobs](#) [Donate](#) [Sitemap](#)[Terms of Service & Honor Code](#) [Privacy Policy](#) [Accessibility Policy](#)

© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY
OPENedX

