**Quiz**

**Question 1: What is a job?**

An activity spawned in the response to a Spark action.

**Question 2: What is a task?**

A unit of work performed by the executor.

**Question 3: What is a job stage?**

A pipelineable part of the computation.

**Question 4: How does your application find out the executors to work with?**

The SparkContext object allocates the executors by communicating with the cluster manager.

**Question 5: Mark all the statements that are true.**

You can ask Spark to make several copies of your persistent dataset.

Data can be cached both on the disk and in the memory.

Spark can be hinted to keep particular datasets in the memory.

**Question 6: Imagine that you need to deliver three floating-point parameters for a machine learning algorithm used in your tasks. What is the best way to do it?**

Capture them into the closure to be sent during the task scheduling.

**Question 7: Imagine that you need to somehow print corrupted records from the log file to the screen. How can you do that?**

Use an action to collect filtered records in the driver.

**Question 8: How broadcast variables are distributed among the executors?**

The executors distribute the content with a peer-to-peer, torrent-like protocol, and the driver seeds the content.

**Question 9: What will happen if you use a non-associative, non-commutative operator in the accumulator variables?**

Operation semantics are ill-defined in this case.

**Question 10: Mark all the operators that are both associative and commutative.**

min(x, y) = if x > y then y else x end

sum(x, y) = x + y

max(x, y) = if x > y then x else y end

prod(x, y) = x \* y

**Question 11: Does Spark guarantee that accumulator updates originating from actions are applied only once?**

Yes

**Question 12: Does Spark guarantee that accumulator updates originating from transformations are applied at least once?**

No