Status	Finished
Started	Friday, 3 October 2025, 11:52
Completed	Friday, 3 October 2025, 11:53
Duration	1 min 11 secs
Marks	5.00/5.00
Grade	**10.00** out of 10.00 (**100**%)

Question 1

Correct

Mark 1.00 out of 1.00

A dataflow is a directed acyclic graph (DAG) where nodes represent operations and edges represent the data flowing from one operator to the other.

- ◉ a. True ⊘
- ○ b. False

Your answer is correct.

The correct answer is:
True

Question 2

Correct

Mark 1.00 out of 1.00

A dataflow engine receives as input a [logical] ⊘ plan and creates a [physical] ⊘ plan. It includes both [logical] ⊘ optimizations, such as [operator re-ordering] ⊘ , and [physical] ⊘ optimizations such as [join algorithm selection] ⊘ . It runs on a [cluster] ⊘ in [parallel] ⊘ while handling [failures] ⊘ .

Your answer is correct.

The correct answer is:
A dataflow engine receives as input a [logical] plan and creates a [physical] plan. It includes both [logical] optimizations, such as [operator re-ordering], and [physical] optimizations such as [join algorithm selection]. It runs on a [cluster] in [parallel] while handling [failures].

Question **3**

Correct

Mark 1.00 out of 1.00

The following code in Spark will return a list of (10, 20, 30, 40).

data = ArrayList(1, 2, 3, 4)

values = sc.parallellize(data)
            .map(number -> number*10)

○ True

◉ False ⊘

The above program does not contain any action operator so it won't be executed. Thus, it won't return anything.

The correct answer is 'False'.

Question **4**

Correct

Mark 1.00 out of 1.00

Spark is faster than Hadoop because:

- ☐ a.   It provides a functional programming interface.

- ☑ b.   It keeps data in memory (if possible) instead of saving intermediate results on disk. ⊘

- ☐ c.   It uses hash partitioning to shuffle the data.

- ☑ d.   It can re-use data computed by certain operations. ⊘

- ☐ e.   It uses modern hardware and accelarators.

Your answer is correct.

The correct answers are:
It keeps data in memory (if possible) instead of saving intermediate results on disk.,

It can re-use data computed by certain operations.

Question 5

Correct

Mark 1.00 out of 1.00

Match the following:

A Spark job is	a set of tasks executed as a result of an action operation.	✓
A task in Spark is	a unit of work over one partition sent to one executor.	✓
A stage in Spark is	a set of tasks in a job executed in parallel without any network communication.	✓

Your answer is correct.

The correct answer is:

A Spark job is → a set of tasks executed as a result of an action operation.,

A task in Spark is → a unit of work over one partition sent to one executor.,

A stage in Spark is → a set of tasks in a job executed in parallel without any network communication.