Congratulations! You passed!

TO PASS 80% or higher

Keep Learning

GRADE
89.50%

Module 1 Quiz

LATEST SUBMISSION GRADE
89.5%

1. Which of the following are true when it comes to the business value of big data? (Select all that apply.) **1 / 1 point**

- [x] Businesses are increasingly making data-driven decisions

  ✓ **Correct**
  More and more, businesses are seeing the value of driving decision-making using data.

- [ ] Automated technologies mean that data scientists and data analysts are no longer needed

- [x] The size of the data businesses collect is growing

  ✓ **Correct**
  As the technology improves, businesses are collecting more and more data.

2. Spark uses... **1 / 1 point**

(Select all that apply.)

- [x] A distributed cluster of networked computers made of a driver node and many executor nodes

  ✓ **Correct**

- [ ] A distributed cluster of networked computers made of many driver nodes and many executor nodes
- [ ] One very large computer that is able to run computation against large databases
- [ ] Your database technology (e.g., Postgres or SQL Server) to run Spark queries
- [x] A driver node to distribute work across a number of executor nodes

✓ **Correct**

3. How does Spark execute code backed by DataFrames? (Select all that apply.) **1 / 1 point**

- [x] It separates the "logical plan" of what you want to accomplish from the "physical plan" of how to do it so it can optimize the query

✓ **Correct**

Spark generates code on the fly to provide the most optimal way of serving your query.

- [ ] It executes code determined in advance
- [x] It optimizes your query by figuring out the best "how" to execute what you want

✓ **Correct**

Since Spark knows what you want to accomplish, it's able to figure out the best way to do it.

- [ ] It iterates over all of the source data to exhaustively evaluate queries

4. What are the properties of Spark DataFrames? (Select all that apply.) **1 / 1 point**

- [x] Distributed: Computed across multiple nodes

Correct

Each node computes on its own data.

- [ ] Tables: Operates as any table in SQL environments
- [x] Dataset: Collection of partitioned data

Correct

The collection of data is partitioned so it can be distributed across the cluster.

- [x] Resilient: Fault-tolerant

Correct

If you lose a worker, only recompute work that worker was responsible for.

5. What is the difference between Spark and database technologies? (Select all that apply.) **1 / 1 point**

- [x] Spark is a highly optimized compute engine and is not a database

Correct

Spark is a robust unified analytics engine and does not act like a database.

- [ ] Spark does not interact with databases but uses its proprietary DataFrame technology instead
- [ ] Spark operates for both data storage and computation
- [ ] Spark in an alternative to traditional databases
- [x] Spark is a computation engine and is not for data storage

Correct

Spark is a computation engine, whereas database technology is meant for data storage.

6. What is Amdahl's law of scalability? (Select all that apply.) **1 / 1 point**

- [x] Amdahl's law states that the speedup of a task is a function of how much of that task can be parallelized

Correct

- [ ] A formula that gives the theoretical speedup as a function of the size of a partition (or subset) of data
- [ ] A formula that gives the expected speed of a single processor performing a computation
- [ ] A formula that gives the number of processors (or other unit of parallelism) needed to complete a task
- [x] A formula that gives the theoretical speedup as a function of the percentage of a computation that can be parallelized

Correct

7. Spark offers a unified approach to analytics. What does this include? (Select all that apply.) **0.2 / 1 point**

- [ ] Spark code can be written in the following languages: SQL, Scala, Java, Python, and R
- [ ] Spark unifies applications such as SQL queries, streaming, and machine learning
- [ ] Spark unifies databases with optimized computation allowing for faster computation against the data it stores

- [ ] Spark allows analysts, data scientists, and data engineers to all use the same core technology
- [ ] Spark is able to connect to data where it lives in any number of sources, unifying the components of a data application

You didn’t select all the correct answers

8. What is a Databricks notebook? **1 / 1 point**

- ( ) A single Spark query
- (•) A collaborative, interactive workspace that allows you to execute Spark queries at scale
- ( ) A Spark instance that executes queries
- ( ) A cluster that executes Spark code

✓ **Correct**

A notebook is an interactive way of interacting with Spark code.

9. How can you get data into Databricks? (Select all that apply.) **0.75 / 1 point**

- [x] By "mounting" data backed by cloud storage

✓ **Correct**

Mounting data makes it appear in Spark as though the data were sitting on the Cluster itself.

- [x] By uploading it through the user interface

✓ **Correct**

Uploading data through the user interface works well for small datasets.

- [x] By registering the data as a table

Correct

Data in Spark can be registered as its own table.

- [x] By connecting to Dropbox or Google Drive

This should not be selected

Please revisit the lesson: **Import Data**.

10. What are the qualities of big data? (Select all that apply.) **1 / 1 point**

- [x] Volume: the amount of data

Correct

The amount of data is growing exponentially.

- [x] Variety: the diversity of data

Correct

More and more different kinds of data are being processed by data applications.

- [x] Veracity: the reliability of data

Correct

Data is not always reliable as it is sometimes user generated, poorly processed, or with other problems.

- [ ] Valorous: the positives impact of data
- [x] Velocity: the speed of data

Correct

The speed at which data arrives in architectures is growing exponentially.