Try again once you are ready

Grade received 55%

Latest Submission Grade 55%

To pass 80% or higher

Try again

1. How does Apache Spark solve read/write problems encountered by other tools? 1 / 1 point

- By keeping much of the required data in-memory
- By only using certain processors in the distributed group.
- By using special proprietary APIs.
- By leveraging redundancy.

Expand

Correct
Correct, Keeping data in-memory avoids disk I/O, which speeds up processes.

2. The three components of Spark architecture are: 0 / 1 point

- Data storage, cluster management framework, and APIs.
- Hadoop, APIs, and Spark Core.
- Cluster management framework, Spark Core, and task schedulers.
- Hadoop, data storage, and executors.

Expand

Incorrect
Incorrect, Please refer to the Scale out/Data Parallelism in Apache Spark video.

3. Select the characteristics of datasets. 1 / 1 point

- Strongly-typed; use APIs in Java, Scala, Python and R; built on top of DataFrames; added in earlier Spark versions
- Strongly-typed; use APIs in Java, Scala, Python and R; built on top of DataFrames; are the latest data abstraction added to Spark.
- Strongly-typed; use APIs in Java, Scala, Python and R; built on top of RDDs; are the latest data abstraction added to Spark.
- Strongly-typed; use unified Java and Scala APIs; built on top of DataFrames; are the latest data abstraction added to Spark.

Expand

Correct
Yes! This answer is correct.

4. Which of these is one of the four phases of Catalyst query optimization? 0 / 1 point

- Code analysis
- Physical optimization
- Logical planning
- Analysis

Expand

Incorrect
Incorrect, Please refer to the Catalyst and Tungsten video.

5. How does IBM Spectrum Conductor help avoid downtime when running Spark? 0 / 1 point

- Shares cluster resources
- Cluster resources divided dynamically
- Automatic troubleshooting
- Deploy multiple versions

Expand

Incorrect
Incorrect, Please review the Using Apache Spark on IBM Cloud video.

6. What is the name of the Spark unified interface? 0 / 1 point

- spark-submit
- SUI
- YARN
- spark-default

Expand

Incorrect
Incorrect, Please refer to the How to Run a Spark Application video.

7. Which command specifies the number of executor cores for a Spark standalone cluster for the application? 1 / 1 point

- Use the command `--total-executor-cores` followed by the number of cores.
- Use the command `--app-total-executor-cores` followed by the number of cores.
- Use the command `--app-executor-cores` followed by the number of cores.
- Use the command `--app-total-executor-cores` followed by the number of cores

Expand

Correct
This answer is correct!

8. If a task fails due to an error, Spark _____.. 1 / 1 point

- continues with related executor tasks
- can attempt to rerun the task for a set number of retries.
- attempts to locate a missing dependency
- terminates the application and reports an error to the driver.

Expand

Correct
Correct, The cause of an application failure can usually be found in the driver event log.

9. Which of the following was mentioned in the course as a common application of Big Data? 1 / 1 point

- Running automotive assembly lines
- Recommendation engines on websites like Amazon and Google
- Optimizing streaming video services
- Writing new video games

Expand

Correct
Correct, Feedback about why this answer is correct

10. What is the fifth V of Big Data? 1 / 1 point

- Vigor
- Vulnerability
- Value
- Validity

Expand

Correct
Correct, The other four Vs - velocity, volume, variety, and veracity - create value for business.

11. What is Data Scaling? 0 / 1 point

- Data scaling is only applicable within cloud environments.
- Data scaling divides workloads to run in parallel.
- Data scaling is a technique to manage, store, and process the overflow of data.
- Data scaling is the process of transforming data values for end use.

Expand

Incorrect
This answer is incorrect. Please review the Parallel Processing and Scalability video.

12. What is the current projected yearly growth rate for data? 0 / 1 point

- 75 percent
- 25 percent
- 40 percent
- 90 percent

Expand

Incorrect
This answer is incorrect. Review the Beyond the Hype video.

13. Which of the following Hadoop core components prepares the RAM and CPU for Hadoop to run data in batch, stream, interactive, and graph processing?

- HDFS
- MapReduce
- YARN
- Hadoop Common

Expand

Correct
Correct, Yarn is short for "yet another resource negotiator" and it's one of the most important components because it prepares the RAM and CPU for Hadoop to run data in these types of processing.

14. Increasing Executors and cores _____. 1 / 1 point

- necessitates dividing jobs into tasks
- requires a shuffle
- increases cluster parallelism
- transforms data partitions

Expand

Correct
Correct, Tasks run in separate threads until all cores are used.

15. What is the Spark property configuration that follows a precedence order, with the highest being configuration set programmatically, then spark-submit configuration and lastly configuration set in the spark-defaults.conf file? 0 / 1 point

- Setting how many cores are used
- Application version
- Properties related to the application
- Application name

Expand

Incorrect
Incorrect, Please review the Setting Apache Spark Configuration video.

16. Why might you want to host Kubernetes on a local machine? 0 / 1 point

- For better security
- As a test development environment
- To keep costs low
- To limit the scope of information used

Expand

Incorrect
Incorrect, Please refer to the Running Spark on Kubernetes video.

17. Select the answer that identifies the licensing types available for open source software. 0 / 1 point

- Public domain, Copyright, Permissive, General Public License
- Public domain, Copyleft, Permissive, General Public License
- Public domain, Copyright, Permissive, Lesser General Public License
- Public domain, Copyleft, Permissive, Lesser General Public License

Expand

Incorrect
Incorrect, Please review the Open Source and Big Data video.

18. What is a key advantage of MapReduce that was mentioned in the course? 1 / 1 point

- It's specialized for the social media industry.
- It can run independently from Hadoop.
- It reduces the data footprint.
- It allows a high level of parallel jobs across nodes.

Expand

Correct
Correct, This saves time and gives flexibility.

19. Which of the following characteristics are part of Hive rather than a traditional relational database? 1 / 1 point

- Used to maintain a Database and uses SQL.
- Can handle petabytes of data.

  Correct
  Correct Hive can handle petabytes of data.
- Designed on the methodology of write once, read many

  Correct
  Correct Hive is designed on the methodology of write once, read many
- Can handle terabytes of data.

Expand

Correct
Great, you got all the right answers.

20. Which of the following is NOT included in the Spark workflow? 1 / 1 point

- Jobs in progress running as tasks in the executors
- Jobs created by the SparkContext in the driver program.
- Jobs transferring results back to the driver or writing to disk
- Jobs held over as incomplete from a previous stage

Expand

Correct
Correct, A new stage begins after all the tasks in the previous stage are complete.