

Registration Number:

Date & Session:



ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU -27
M.Sc (BIG DATA ANALYTICS) – III SEMESTER
SEMESTER EXAMINATION: OCTOBER 2022
(Examination conducted in December 2022)
BDA3221: ENABLING TECHNOLOGIES FOR DATASCIENCE

Time: 2 ½ Hours

Max Marks: 70

This paper contains THREE printed pages and THREE parts

PART- A

Answer All Questions

10X1=10

1. Given a data frame df, select the code that returns its number of rows:

- a. df.take('all')
- b. df.collect()
- c. df.count()
- d. df.numRows()

2. Which of the following language is not supported by Spark?

- a. Java
- b. Pascal
- c. Scala
- d. Python

3. Which of the following is a transformation?

- a. foreach()
- b. flatMap()
- c. save()
- d. count()

4. Which of the following is an actions

- a. count()
- b. printSchema()
- c. cache()
- d. sort()

BDADA3321_A_O_22

5. Spark is best suited for _____ data.
- a. Real-time
 - b. Virtual
 - c. Structured
 - d. All of the above
6. RDD is fault-tolerant and immutable
- a. True
 - b. False
 - c. Both
 - d. None
7. Spark is engineered from the bottom-up for performance, running _____ faster than Hadoop by exploiting in memory computing and other optimizations.
- a. 100x
 - b. 150x
 - c. 200x
 - d. None of the mentioned
8. _____ is a distributed machine learning framework on top of Spark.
- a. MLlib
 - b . Spark Streaming
 - c . GraphX
 - e. RDDs
9. Fault Tolerance in RDD is achieved using
- a: Immutable nature of RDD
 - b . DAG (Directed Acyclic Graph)
 - c. Lazy-evaluation
 - d .None of the above

10. The shortcomings of Hadoop Map Reduce was overcome by Spark RDD by
- a. Lazy-evaluation
 - b .DAG
 - c .In-memory processing
 - d .All of the above

PART B

Answer Any Six Questions

6x5=30

- 11. Explain briefly about big data characteristics.
- 12. What are the important components of the Spark ecosystem?
- 13. Explain the types of operations supported by RDDs.
- 14. Explain lazy evaluation in RDD.
- 15. Explain how partition works in RDD.
- 16. What is the difference between RDD and Data frame?
- 17. Explain spark security in details.
- 18. What are the different levels of persistence in Spark?

PART C

Answer Any Three Questions

3x10=30

- 19. Explain the significance of RDD and its operations with examples.
- 20. Explain spark architecture in details with how read and write works.
- 21. Explain pyspark data frame and features of pyspark sql in detail.
- 22. Explain lambda architecture and spark streaming architecture.