



Registration Number:

Date & session:

ST. JOSEPH'S UNIVERSITY, BENGALURU -27
M.Sc. (BIG DATA ANALYTICS) – II SEMESTER
SEMESTER EXAMINATION: APRIL 2023
(Examination conducted in May 2023)

BDA2421 – ENABLING TECHNOLOGIES FOR DATA SCIENCE I
(For current batch students only)

Time: 2 Hours

Max Marks: 50

This paper contains ONE printed page and THREE parts

PART-A

Answer ALL the questions

5X2=10

1. How is the secondary name node different from the primary?
2. What is the concept of heartbeat?
3. Explain the need for high durability and high availability. How is it measured?
4. How does Apache Sqoop work?
5. Name the datatypes in Pig.

PART B

Answer any FIVE questions

5X4=20

6. Explain the concept of replication in Hadoop ecosystem.
7. Show the flow of map reduces with a diagram.
8. Differentiate Hadoop 1x, 2x and 3x architectures.
9. Explain security in Hadoop.
10. Differentiate between an external and internal table in Hive.
11. Through examples, discuss how Pig is different from Hive.

PART C

Answer any TWO questions

2X10=20

12. Explain the advantages of Hadoop with relevant examples
13. Explain Hive architecture with a diagram.
14. Explain erasure coding with a diagram.