24001291

**QP CODE: 24001291**

**Reg No :** .....

**Name :** .....

## B.Sc DEGREE (CBCS) REGULAR / REAPPEARANCE EXAMINATIONS, MARCH 2024

### Sixth Semester

B.Sc Computer Science Model III

### CORE COURSE - CC6CRT07 - BIG DATA : ANALYTICS

2017 Admission Onwards

5530B1B8

Time: 3 Hours Max. Marks : 80

**Part A**

*Answer any **ten** questions.*

*Each question carries **2** marks.*

1. Define the term 'Big Data'.
2. What are the categoriess of big data?
3. What is random variable?
4. Describe any one stream filtering technique.
5. What are the time complexity and space complexity of FM algorithm?
6. What is the significance of second order moments?
7. What are InputSplit?
8. What are problems with FIFO scheduler in MapReduce?
9. Develop a sample specification of a machine to run a Hadoop datanode.
10. Define metrics . Give an example of its usage in Hadoop.
11. What are ephemeral znodes?
12. What is visual data analysis?

(10×2=20)

**Part B**

*Answer any **six** questions.*

*Each question carries **5** marks.*

13. Explain data analysis process.

14. Explain the characteristics of data streams.

15. Explain the differrence between InputSplit and DataBlock. What should be the appropriate size of InpuSplit and why?

16. Write about the Hadoop Java API.

17. Write about the Shuffle and Sort Phase in MapReduce.

18. Describe the various tools used in the administration of Hadoop.

19. Organize the steps in setting a Hadoop cluster on Amazon EC2.

20. Discuss the the advantages of Pig.

21. Write a note on Hive data types.

(6×5=30)

**Part C**

*Answer any **two** questions.*

*Each question carries **15** marks.*

22. Explain stream data model Architecture using diagram.

23. Define MapReduce, Job and Task? Explain the working of a Classic MapReduce.

24. Discuss the importance of security in Hadoop. Discuss how security is enforced using Kerberos.

25. Explain about querying data in Hive.

(2×15=30)