**QP CODE: 20100701**

20100701

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

**Name** : .....

## BSc DEGREE (CBCS) EXAMINATION, MARCH 2020

### Sixth Semester

B.Sc Computer Science Model III

### Core Course - CC6CRT07 - BIG DATA : ANALYTICS

2017 Admission Onwards

B71041C4

Time: 3 Hours Marks: 80

**Part A**

*Answer any **ten** questions.*

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

1. Define the important v's in big data.
2. Define variance.
3. Define random experiment.
4. Define Stream.
5. What are the applications for counting distinct elements.
6. Define real time analytics.
7. What is a block in HDFS? What is the default blocksize?
8. What are InputSplit?
9. List the methods adpopted to synchronize configuration files.
10. Define context. What is the use composite context?
11. Differentiate schema on read on write design.
12. List out the basic operations in zookeeper.

(10×2=20)

## Part B

*Answer any **six** questions.*

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

13. Explain the types of big data.
14. Explain the concept of decaying windows.
15. What are the components of Hadoop?
16. Explain Failure Modes in Classic MapReduce.
17. Write about the Shuffle and Sort Phase in MapReduce.
18. Explain the two-level network architecture of a Hadoop cluster with an emphasis on rack awareness.
19. Illustrate the routine administration procedures in the maintenance of Hadoop.
20. Write examples for sorting and aggregation in Hive data.
21. List out the applications of IBM BigInsights and infosphere streams.

(6×5=30)

## Part C

*Answer any **two** questions.*

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

22. Explain filtering streams.
23. What are InputFormats? Explain the InputFormats and OutputFormats of Hadoop.
24. What are the various methods and tools that can be used in monitoring the performance of a Hadoop cluster.
25. Illustrate the architecture of Pig and explain its working with Hadoop.

(2×15=30)