QP CODE: 22101186

22101186

Reg No : .....

Name : .....

B.Sc DEGREE (CBCS) REGULAR / REAPPEARANCE EXAMINATIONS, APRIL 2022

Sixth Semester

B.Sc Computer Science Model III

CORE - CC6CRT07 - BIG DATA : ANALYTICS

2017 Admission Onwards

84D8DD6E

Time: 3 Hours Max. Marks : 80

Part A

*Answer any **ten** questions.*

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

1. Define variance.
2. Define probability and probability axioms.
3. What do you mean by prediction error?
4. What are standing queries?
5. Define bloom filtering.
6. Mention an algorithm which is used for estimating distinct elements.
7. What is Hadoop?
8. What happens in the partiiton phase?
9. Explain the contents of masters and slaves file in Hadoop configuration.
10. What is the use of using StrictHostKeyChecking option in SSH configuration in Hadoop.
11. Define Hive shell.
12. What is a webtable?

(10×2=20)

Part B

*Answer any **six** questions.*

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

13. What do you mean by intelligent data analysis?

14. Describe real time analytics.

15. How write operation is performed on HDFS.

16. Explain the Architecture of YARN.

17. Briefly explain the use of counters available in MapReduce. List the different type of counters.

18. Explain how the components of HDFS organize their persistent data on disk?

19. Recall the steps involved in commisioning and decommisioning nodes in a Hadoop cluster.

20. Describe the zookeeper architecturel

21. Describe IBM Infosphere streamsl

(6×5=30)

Part C

*Answer any **two** questions.*

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

22. Define a decaying window. Explain how to find the most popular element in the stream?

23. What is HDFS? Explain the architecture and working of HDFS.

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

25. Explain various Pig latin data processsing operators with suitable examples

(2×15=30)