

STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI –600 086
(For candidates admitted from the academic year 2023 – 2024 & thereafter)

M. Sc. DEGREE EXAMINATION, APRIL 2026
BIOINFORMATICS
FOURTH SEMESTER

COURSE : CORE
PAPER : BIG DATA ANALYSIS
SUBJECT CODE : 23BI/PC/BD44
TIME : 3 HOURS

MAX. MARKS: 100

Q. No.	SECTION A (10 x 1=10 marks)	CO	KL
	All questions to be answered		
1	Which type of data includes text, images, and videos? a. structured data b. unstructured data c. printed data d. manual data	CO1	K1
2	EHR are considered as a big data repository because they: a. Store large amounts of patient health information b. Manufacturing medical equipment c. Produce laboratory chemicals d. Print medical textbooks	CO1	K1
3	Which data structure operates on the principle Last In, First Out (LIFO)? a. Queue b. Stack c. Linked List d. Tree	CO1	K1
4	Machine generated data differs from people-generated data because machine-generated data is mainly: a. written manually by humans b. collected only through surveys c. Stored in books d. Produced automatically by devices and sensors	CO1	K1
5	In medical imaging, the process of aligning images from different time points or modalities is known as: a. Reconstruction b. Registration c. Annotation d. Thresholding	CO1	K1
6	The storage component of Hadoop is a. MapReduce b. HDFS c. YARN d. Hive	CO2	K2
7	HDFS is part of : a. windows b. Linux c. Apache Hadoop d. Oracle	CO2	K2
8	MapReduce works in how many phases? a. One b. Two c. Three d. Four	CO2	K2
9	Apache Spark is faster than: a. Python b. JAVA c. Hadoop MapReduce d. SQL	CO2	K2
10	Which technology helps analyze large healthcare datasets? a. Data Science b. Painting software c. Video editing tools d. Music apps	CO2	K2
Q. No.	SECTION B (10 x 2= 20 marks)	CO	KL
	Answers in about 50 words		
11	Explain the six V's of Big Data with examples	CO3	K3
12	Differentiate between structured and unstructured data	CO3	K3
13	Describe the role of data repositories in Big Data analytics.	CO3	K3
14	Explain the steps involved in extracting value from Big Data.	CO3	K3
15	Discuss people-generated data with suitable examples.	CO3	K3
16	Analyze the role of Hadoop in healthcare data analysis	CO4	K4
17	Describe the working principle of YARN.	CO4	K4

18	Examine the significance of Spark in Big Data processing.	CO4	K4
19	Discuss applications of Big Data in biological research.	CO4	K4
20	Discuss the importance of probabilistic modeling in healthcare.	CO4	K4
Q. No.	SECTION C (4 x 10= 40 marks) Answer in about 600 words	CO	KL
21	a) Explain distributed computing and computer clusters in detail. (OR) b) Describe cloud computing architecture with reference to AWS or Google Cloud.	CO4	K4
22	a) Discuss job scheduling and resource management in HPC clusters. (OR) b) Explain Linux essentials required for cluster computing.	CO4	K4
23	a) Describe CHIP-Seq analysis workflow including peak identification and motif finding. (OR) b) Explain ATAC-seq and Bisulfite sequencing in Big Biological Data analysis.	CO5	K5
24	a) Discuss Snakemake and Nextflow workflows for bioinformatics pipelines. (OR) b) Explain Docker and Singularity containers with suitable use cases.	CO5	K5
Q. No.	SECTION D (2 x 15= 30 marks) Answer any TWO questions in about 1200 words	CO	KL
25	Critically analyze the role of Big Data analytics in modern healthcare systems.	CO5	K6
26	Design a Big Data architecture for biological data analysis using Hadoop and Spark.	CO5	K6
27	Evaluate the challenges and advantages of integrating multi-omics Big Data.	CO5	K6
28	Critically discuss workflow management, containerization, and version control in Big Data projects.	CO5	K6
