

Branch – **SOFTWARE SYSTEMS (FIVE YEARS INTEGRATED)**

MAJOR ELECTIVE COURSE – II BIG DATA ANALYTICS

Time: Three Hours

Maximum: 75 Marks

SECTION-A (10 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** marks

(10 × 1 = 10)

Module No.	Question No.	Question	K Level	CO
1	1	What are all the four characteristics of Big Data? (i) Volume, Variance, Velocity, Value (ii) Volume, Variety, Velocity, Value (iii) Volume, Data, Redundancy, Confidentiality (iv) Volume, Scalability, Energy, Value	K1	CO1
	2	How Apache Hadoop define _____ as "datasets which could not be captured, managed and processed by general computers within an acceptable scope". (i) Data Analysis (ii) Big Data (iii) Social Networks (iv) Database	K2	CO1
2	3	Choose: means the delivery and use mode of IT infrastructure that is acquiring necessary resources through the Internet on-demand or in an expandable way. (i) Grid Computing (ii) Big Data (iii) Cloud Computing (iv) Capacity	K1	CO2
	4	Show which of the following is the advantage of Hadoop? (i) Distributed Processing (ii) Centralized Management (iii) Strong Flexibility (iv) IoT enabled	K2	CO2
3	5	What is the first step of big data, specifically, it is large-scale, highly diverse and complex datasets generated through longitudinal and distributed data sources. (i) Data acquisition (ii) Data generation (iii) Data Distribution (iv) All the above	K1	CO3
	6	Show which is a widely used network data packet capture function library (i) Log files (ii) Cloud Computing files (iii) Mobile technology (iv) Libpcap	K2	CO3
4	7	State the term unbounded data refers to _____ (i) Data is often represented as series of events (ii) technique used to segment data streams over time (iii) Data flows continuously without a defined start or end (iv) All the above	K1	CO4
	8	Infer which layer responsible for collecting data streams from the sources and making them available for processing by the stream processing engine (i) Ingestion layer (ii) Data Sources (iii) Data Routing (iv) Data Transformation	K2	CO4
5	9	Which among the following is the next generation power grid constituted by traditional energy networks integrated with computation, communications and control for optimized generation, supply and consumption of electric energy (i) Grid Computing (ii) Smart Grid (iii) Collective Intelligence (iv) Social Networking	K1	CO5

Cont...

5	10	Identify which aims to automatically retrieve, extract and evaluate information from web documents and services so as to discover useful knowledge. (i) Web content (ii) Web Analysis (iii) Supervised learning (iv) Web structured mining	K2	CO5
---	----	--	----	-----

SECTION - B (35 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** Marks

(5 × 7 = 35)

Module No.	Question No.	Question	K Level	CO
1	11.a.	Illustrate the development of Big Data.	K2	CO1
		(OR)		
	11.b.	Discuss the Big Data era and its characteristics.		
2	12.a.	Demonstrate the relationship between Cloud computing and big data.	K3	CO2
		(OR)		
	12.b.	Show the usage of Data Center.		
3	13.a.	Examine Enterprise data in Big Data generation.	K3	CO3
		(OR)		
	13.b.	Explore Internet Data and Bio-medical Data.		
4	14.a.	Explain Stream Computing.	K4	CO4
		(OR)		
	14.b.	Analyze about the sampling data in a stream.		
5	15.a.	Explain Big Data Application evolution.	K5	CO5
		(OR)		
	15.b.	Defect Map Reduce for machine learning.		

SECTION -C (30 Marks)

Answer **ANY THREE** questions

ALL questions carry **EQUAL** Marks

(3 × 10 = 30)

Module No.	Question No.	Question	K Level	CO
1	16	Briefly explain Intelligent Data Analysis and nature of data.	K4	CO1
2	17	Analyse Hadoop preliminaries and NoSQL.	K4	CO2
3	18	Explain Data Collection process in a step by step manner.	K4	CO3
4	19	Examine Real Time Analytics Platform (RTAP) applications.	K4	CO4
5	20	Compare and contrast Web Data Analysis and Multimedia Data Analysis.	K4	CO5

Z-Z-Z

END