Continuous Assessment Test (CAT) – I FEB 2024

Programme	:	Master of Computer Applications	Semester	:	Winter 23-24
Course Code & Course Title	:	PMCA607L & Big Data Analytics	Class Number	:	CH2023240501412
Faculty	:	Dr.K.Madheswari	Slot	:	E2+TE2
Duration	:	90 minutes	Max. Mark		50

General Instructions:
- Write only your registration number on the question paper in the box provided and do not write other information.
- Use statistical tables supplied from the exam cell as necessary
- Use graph sheets supplied from the exam cell as necessary
- Only non-programmable calculator without storage is permitted

Answer all questions

Q. No	Sub Sec.	Description	Marks
1	a	Imagine engaging in various activities such as making voice calls, video calls, sending emojis, GIF images on WhatsApp, sharing GPS information on WhatsApp and Facebook, and receiving SMS notifications about bank transactions. Additionally, you're verifying marks and attendance on VTOP. Classify the types of data you have sent and received in these scenarios, and justify your answers.	5
	b	Illustrate the process of splitting a 350MB input file into blocks within HDFS. Explain the splitting strategy and file storage using clear diagrams.	5
2	a	Consider a scenario where a large financial institution relies on a distributed computing framework like Hadoop Distributed File System (HDFS) to store and process vast amounts of financial data. Discuss the potential consequences of node failure within this system and outline the mechanisms implemented to mitigate and recover from such failures. Utilize relevant examples and technical explanations to support your analysis.	10
3		Describe in depth the various phases of the data analytics life cycle, utilizing a case study on diamond carat price prediction.	10
4	a	Explain the fundamental elements comprising the architecture of Hadoop 1.0.	5
	b	Outline the process of submitting and executing jobs in a Hadoop cluster. Assume you intend to execute a Java file (sum.java) tasked with summing 10,000 numbers within a timeframe of 20 seconds. Detail the steps involved in achieving this within a Hadoop cluster.	5
5	a	Describe in detail the characteristics of big data.	4
	b	Discuss the obstacles and utilization of big data in various industries, supported by relevant instances.	6

***********All the best ************