

Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



Department of Research & Development
Mid - Term Examinations - SEPTEMBER 2024

Odd Semester: Ph.D. Course Work	Date: 27 /09/2024
Course Code: CSE902	Time: 10:00am – 11:30am
Course Name: Big Data Analytics tools and Techniques	Max Marks: 50
Department: SOCSE	Weightage: 25%

Instructions:

- (i) Read all questions carefully and answer accordingly.
- (ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 5 marks.		4Qx5M=20M
1	Discuss the two data flow anatomies in detail	5 Marks
2	What are the characteristics of big data	5 Marks
3	Explain HDFS concepts	5 Marks
4	Describe Serialization, Deserialization and brief RPC serialization format	5 Marks

Part B

Answer ALL Questions. Each question carries 15 marks.		2QX15M=30M
5	Apache Hadoop is an open-source framework intended to make interaction with big data easier. Hadoop has made its place in industries and companies that need to work on large data sets which are sensitive and needs efficient handling. Hadoop is a framework that enables the processing of large data sets which reside in the form of clusters. Being a framework, Hadoop is made up of several modules that are supported by a large ecosystem of technologies. Identify the components in the ecosystem and brief them in detail	15 Marks
6	MapReduce is a programming model for writing applications that can process Big Data in parallel on multiple nodes. MapReduce provides analytical capabilities for analyzing huge volumes of complex data. Build a MapReduce model to count the words in this text. [The birds are singing with joy. It is sitting on the trees. They look really beautiful]. Brief each phase in detail.	15 Marks