Roll No._____

Serial No. : 4639

Unique Paper Code	: 2346005001
Name of Paper	: Big Data Analytics - II
Name of Course Course	: Skill Enhancement
Semester	: **IV**
Duration	: 1 Hour
Maximum Marks	: 30

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. The paper has two sections. All questions in "**Section A**" are compulsory.
3. Attempt any four questions from "**Section B**".
4. Parts of a question must be answered together.

Section A

1. (a) State whether the following statements are true or false: **(5)**

(i) Scala is a popular language used to write programs in Apache Spark.

(ii) All Sqoop jobs are map-only jobs and do not require the reduce phase.

(iii) Data storage is the primary focus of machine learning in the context of big data.

(iv) The main benefit of GPU architecture over traditional CPUs is increased memory size.

(v) In the **CUDA** memory model, shared memory is used among all threads of a thread block.

(b) Fill in the blanks. **(5)**

(i) The primary purpose of Apache Flume is to reliably and efficiently ingest data from various sources into the _____.

(ii) _____ is the computing model that facilitates parallel execution on graphics hardware.

(iii) The tool used for cross-language service communication and data encoding in Hadoop is called _____.

(iv) In Hadoop, the _____ facility allows the user to work with many small files efficiently.

(v) The type of learning under which K-Means clustering is categorized is known as _____.

Section B

(Attempt any four)

2. Name the various types of memory provided by the CUDA memory model. **(5)**

3. Differentiate between RDD transformations and RDD actions in Apache Spark. Give one example of each. **(5)**

4. Explain the key advantages of Apache Spark compared to traditional Hadoop MapReduce. Also, mention two application scenarios where Apache Spark is considered more suitable. **(5)**

5. Explain the algorithm used to implement K-means clustering with pseudocode. **(5)**

6. Distinguish between supervised and unsupervised learning. Give one example of each. **(5)**