1T00163 - M.C.A. (SEM-III) (Choice Based) (Two Year Course) / 70721 - Big Data Analytics & Visualization   QP CODE : 10016823   DATE: 25/07/2023

(3 Hours)   **[Total Marks : 80]**

N.B. : 1) Question No.1 is **compulsory.**
2) Attempt any **three** from the remaining **five** questions.

Write a short note on following (any Four)

1.	(a)	5 V's of Big Data	**(5)**
	(b)	HDFS	**(5)**
	(c)	No SQL	**(5)**
	(d)	Hive	**(5)**
	(e)	Pig	**(5)**
	(f)	Apache Kafka	**(5)**

2. (a) Explain Hadoop Ecosystem with core components. Explain its architecture ?   **(10)**

(b) What are the different frameworks that run under YARN? Discuss the various YARN Daemons.   **(10)**

3. (a) What is NoSQL? Explain various NoSQL Data architecture patterns?   **(10)**

(b) What is RDD? How is data partitioned in RDD ?   **(10)**

4. (a) (i) What are the advantages of Apache Spark over Map Reduce ?   **(10)**
(ii) D3 and big data

(b) Explain Bulk Synchronous Processing (BSP) and graph processing wrt to Apache spark ?   **(10)**

5 (a) What is significance of Apache Pig in Hadoop Context? Explain the main components and the working of Apache Pig with the help of diagram ?   **(10)**

(b) What is Big Data and its types? List out the difference between Traditional Data Vs. Big Data with the help of example ?   **(10)**

6 (a) Discuss the Apache Kafka fundamentals. Explain the Kafka Cluster Architecture with suitable diagram   **(10)**

(b) Write short note on any one: **(10)**

    a. Master slave vs peer to peer Architecture in No SQL
    b. Partitioner and Combiner wrt to Map Reduce

_____

16823   **Page 1 of 1**