

SVKM's
Dwarkadas J. Sanghvi College of Engineering
Acad .Year 2022-2023
YEAR III / Semester VI

Program: B.Tech in Computer Engineering
Subject/Course: Big Data Infrastructure
Date: 31.05.2023

Max. Marks: 75
Time: 09:00-12:00
Duration: 03:00 Hrs

FINAL EXAMINATION

Instructions: Candidates should read carefully the instructions printed on the question paper and on the cover page of the Answer Book, which is provided for their use.

- (1) This question paper contains 2 pages.
- (2) All Questions are Compulsory.
- (3) All questions carry equal marks.
- (4) Answer to each new question is to be started on a fresh page.
- (5) Figures in the brackets on the right indicate full marks.
- (6) Assume suitable data wherever required, but justify it.
- (7) Draw the neat labelled diagrams, wherever necessary.

Question No.		Max. Marks
Q1 (a)	What are the challenges with big data?	[05]
Q1 (b)	Explain the responsibilities of Name node, Data node and Secondary node in HDFS. OR How Mapper, combiner, partitioner and reducer work in map reduce workflow.	[10] [10]
Q2 (a)	Write pseudo code for word count by map reduce. Illustrate with an example showing all the steps?	[10]
Q2 (b)	Write Pig commands for following: i. Join two tables ii. Filter records on the basis of a condition OR HDFS Commands for following: i. Copy a file from local file system to HDFS ii. Set replication factor of a file/directory in HDFS to 5.	[05] [05]

Q3 (a)	What are the Different types of znodes in zookeeper, mention the operations performed by zookeeper.	[05]
Q3 (b)	Compare HBase vs RDBMS. How “write” mechanism works in HBase. OR Explain the concept of bucketing and partitioning with the help of query.	[10] [10]
Q4 (a)	What are the various interaction patterns in streaming data? OR What are the various deployment modes of spark? Draw any explain spark ecosystem.	[10] [10]
Q4 (b)	Explain the concept of commit and offset in Kafka stream processing?	[05]
Q5 (a)	Assume a database named “Company”. Write Mongodb queries for following: i. Create a COLLECTION “EMPLOYEE” ii. Insert data for 2 employees: name, address, salary, designation, contact_no iii. Modify salary of all developers to 70000 iv. Display all employee details whose designation is “developer” in descending order of name v. Display average salary	[10]
Q5 (b)	What are the advantages and disadvantages of NOSQL database? OR Explain “Lazy Evaluation” in spark framework with its advantages.	[05] [05]

All the Best!