

Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, December 2022

Course: Big Data Ingestion

Program: B.Tech CSE (Big Data)

Course Code: CSBD2006

Instructions:

Semester: III

Time : 03 hrs.

Max. Marks: 100

**SECTION A
(5Qx4M=20Marks)**

S. No.		Marks	CO
Q 1	List different data sources of big data. Explain the structured, semi-structured and unstructured data format with examples.	4	CO1
Q2	In the context of HDFS define following terms: a. heartbeat b. blocks c. racks d. metadata e. Fsimage	4	CO1
Q3	Using Sqoop import command how can we control the number of mappers? Explain its impact on the files generated in HDFS?	4	CO2
Q4	Batch Ingestion vs Stream Ingestion - What's the Difference	4	CO3
Q5	Explain processing solutions developed before big data?	4	CO1

**SECTION B
(4Qx10M= 40 Marks)**

Q6	Explain big data ingestion pipeline and list some factors need to be considered while developing a data ingestion pipeline.	10	CO2
Q7	“A message queue is a form of <u>asynchronous service-to-service communication</u> used in <u>serverless</u> and <u>microservices</u> architectures.” Explain the terms used in the given statement. Or With the help of a use case, explain the concept of event streaming. Also, illustrate the event streaming platform with the help of a diagram.	10	CO3
Q8	What do you understand by the concept of publish/subscribe messaging? Highlight the role of a broker in such a system.	10	CO4
Q9	Justify the growing need of Bigdata in industries, with the help of a use case. Also, list the features of Hadoop that makes it a top solution as a big data platform.	10	CO1

SECTION-C
(2Qx20M=40 Marks)

Q10	<p>In Apache Kafka, explain the concept of topics and partitions using diagrams. Also, explain Kafka Broker and its working mechanism.</p> <p>Or</p> <p>Define the process of data ingestion using Kafka. Also, name various components of Kafka and explain each of them in brief using the Kafka architecture.</p>	20	CO4
Q11	<p>i. Mention the features of Apache Sqoop. Also, explain the process to perform an incremental data load in Sqoop?</p> <p>ii. Create a table in MySQL. Use that table to illustrate:</p> <ul style="list-style-type: none">a. Sqoop import of data from particular rows.b. Sqoop import of data in Avro format.c. Sqoop import using a free form query.	10+10	CO2