

Name: Enrolment No:		
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES Online End Semester Examination, May 2021		
Course: Big Data Search Program: B. Tech. (CSE) SPZ Big Data Course Code: CSBD 3008	Semester: VI Time: 03 hours Max. Marks: 100	
SECTION A		
1. Each Question will carry 5 Marks 2. Instruction: Complete the statement / Select the correct answer(s)		
Q1	Which is not in category of unstructured data? (a) Health records (b) Audio files (c) Images (d) Excel files	CO1
Q2	Which of the following provides Java based indexing and search technology? (a) Hadoop (b) Lucene (c) Oracle (d) Kafka	CO2
Q3	Which component has the simple task of removing common words from the stream of tokens that become index terms? (a) Stemming (b) Stopping (c) Lemmatizing (d) Crawling	CO2
Q4	----- is the unit at which Elasticsearch distributes data around the cluster. (a) Segment (b) Shard (c) file (d) Replica	CO3
Q5	The----- is a basic unit of information that can be indexed in Elasticsearch. (e) Node (f) Document (g) Tokenizer (h) Replica	CO3
Q6	What is a broker in Kafka? (a) Zookeeper (b) Producer (c) Consumer (d) Server	CO4

SECTION B		
<p>1. Each question will carry 10 marks 2. Instruction: Write short / brief notes</p>		
Q7	How would you define Lucene? How does Lucene build an index? Give an example. [2+8]	CO2
Q8	Define Faceting in Solr. Explain Apache Solr architectural components with diagram. [3+7]	CO2
Q9	How would you define ELK stack? Explain roles of all technologies in the Elasticsearch ELK Software Stack with help of an architectural diagram. [2+8]	CO3
Q10	Explain the main architectural components of Kafka with a diagram. How does records flow in Apache Kafka? Explain. [6+4]	CO4
Q11	<p>Define Spark streaming. What is a DStream in Apache Spark? What are the common transformations on DStreams supported by Spark Streaming? [2+2+6]</p> <p style="text-align: center;">OR</p> <p>How would you define RDD in Spark? What are the main operations of Spark RDD? Explain. [2+8]</p>	CO4
Section C		
<p>1. Each Question carries 20 Marks. 2. Instruction: Write long answer.</p>		
Q12	<p>What are the major issues involved in search engine design? Write building blocks of search engine architecture. Explain major components of building blocks with architectural diagram. [5+5+10]</p> <p style="text-align: center;">OR</p> <p>Document filtering is an application that stores a large number of queries or user profiles and compares these profiles to every incoming document on a feed. Documents that are sufficiently similar to the profile are forwarded to that person via email or some other mechanism. Describe the architecture of a filtering engine and how it may differ from a search engine. [10+10]</p>	CO1