

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, December 2017

Program: BTech CSE + BAO

Semester – VII

Subject (Course): Big Data Analytics

Max. Marks : 100

Course Code: CSIB 385

Duration : 3 Hrs

No. of page/s: 2

*This paper consists of 3 sections as **Section A**, **Section B** and **Section C** having 20, 40 and 40*

Section A (Marks: 20)

Attempt ALL questions. The marks for each question are written alongside with the question.

1. Discuss the differences between API and framework. [4M]
2. What is Job Tracker and Task tracker role in Hadoop Distributed File System? [4M]
3. Define DataNode, NameNode and how does NameNode tackle DataNode failures? [4M]
4. Explain the Fair, FCFS and Capacity scheduling in YARN framework. [4M]
5. Explain the component of Hadoop ecosystem and what are real-time industry applications of Hadoop? [4M]

Section B (Marks: 40)

The marks for each question are written alongside with the question.

6. Draw the spark unified architecture and explain its each component in detail.
[10M]
7. What are the different way to create an RDD ? Explain with its syntax.
[10M]
8. What are the different execution mode available in Spark application?
[10M]

9. Write an short note on:

[10M]

- a. Driver Program
- b. Spark Context
- c. RDD Lineage
- d. Executor

Section C (Marks: 40)

The marks for each question are written alongside with the question.

10. a) What is Spark? State the advantages of using Apache Spark over Hadoop Map Reduce for big data processing with different examples.

b) What is YARN. Draw a neat sketch of architectural diagram of YARN. Explain the function of its various component. [10M + 10M]

11. What is an RDD and its features? Explain the transformations and actions in the context of RDDs with different operations. [6M + 14M]

