

2. (a) What is big data? Discuss it in terms of four dimensions, volume, velocity, variety and veracity.

[7M]

(b) Describe the term big data analytics and what is changing in the realms of big data? [7M]

$\mathbf{MODULE}-\mathbf{II}$

3.	(a)	Describe the core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail. [7M]
	(b)	Discuss in detail about the Hadoop YARN with an example. Explain the four modules that make up the Apache Hadoop framework. [7M]
4.	(a)	Describe interacting process with Hadoop ecosystem in terms of various big data processing technologies. [7M]
	(b)	Recall all the big data storage and processing elements and justify whether Hadoop tackles these challenges. [7M]

$\mathbf{MODULE}-\mathbf{III}$

5.	(a)	What is HDFS? Discuss the I	HDFS architecture and HDFS	commands in brief.	[7M]]
----	-----	-----------------------------	----------------------------	--------------------	------	---

- (b) List the key features of HDFS. Outline the different ways to overwrite the replication factors in HDFS. [7M]
- 6. (a) Create a file in HDFS and explain anatomy of File Read and Write? [7M]
 - (b) Examine the number of spilled records from map tasks far exceeds the number of map output records. The child heap size is 1GB and your io.sort.mb value is set to 1000MB. How would you tune your io.sort? MB value to achieve maximum memory to I/0 ratio. [7M]

- 7. (a) Explain the function of following terms:
 - i) Mapper
 - ii) Reducer
 - iii) combiner.
 - (b) Construct the MapReduce job, under what scenario does a combiner get triggered and write the various options to reduce the shuffling of data in a map reduce job. [7M].
- 8. (a) Describe in detail about the driver class, map and reducer phases with a real time example. [7M]
 - (b) What is role of distributed cache in MapReduce framework? Describe different techniques to optimize MapReduce Job. [7M]

$\mathbf{MODULE}-\mathbf{V}$

- 9. (a) What are joins? How many types of joins are there in pig latin with an examples? [7M]
 - (b) Describe the importance of partitions in Hive with an example. Discuss the Hive commands to create a table with four columns: First name, last name, age, and income? [7M]
- 10. (a) Explain architecture of Apache Hive and various data insertion techniques in Hive with example. [7M]
 - (b) Discuss PIG structure and architecture in brief. Develop a PIG Latin program for an application of word count in a given file. [7M]

 $-\circ\circ\bigcirc\circ\circ-$

[7M]