Name: **Enrolment No:**	UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2018

Course: Bigdata Storage **Semester: III**
Programme: BTech- CSE Bigdata **Course Code: CSBD 2001**
Time: 03 hrs. **Max. Marks: 100**
Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Define the term Storage Media. Explain the three most prevalent forms of data storage.	4	CO1
Q 2	Explain the following terms with respect to File System: a) Superblock b) I-node c) Volume d) Partition	4	CO2
Q 3	Explain at least four limitations of traditional file system.	4	CO2
Q 4	Explain the following terms with respect to Filesystem operations: a) Mounting b) Unmounting c) Symbolic links and Hard links d) Indexing and Logging	4	CO2
Q 5	Give two examples each of the common file systems used in Windows and Linux based systems. Also differentiate between FAT32 and NTFS.	4	CO2
	SECTION B		
Q 6	Explain the four major categories of NoSQL databases. Explain in detail the two NoSQL database types that are used for Content Management Systems.	5+5	CO1
Q 7	What are the five major advantages of NoSQL approach over SQL approach? Explain how NoSQL approach helps in improving Throughput and Scalability?	5+5	CO1
Q 8	Explain in detail the difference between Master-Slave and Peer to Peer Data Replication techniques.	10	CO3
Q 9	Differentiate between Algorithmic Sharding and Dynamic Sharding. Explain at least five challenges faced during sharding. **OR** Differentiate between Lookup strategy and Range strategy used in Sharding. What is the need of sharding a database and how does it differ from replication?	5+5	CO3

SECTION-C			
Q 10	How are NFS and GFS related to DFS? Explain NFS architecture in detail.	**10+10**	**CO4**
Q 11	Explain the concept of file snapshot in HDFS. Explain various goals and assumptions behind the design of HDFS. Explain in detail the concept of data replication in HDFS. [5+5+10] OR Write a note on HDFS explaining its evolution and architecture. Explain the process of journaling and client interaction in HDFS. [10+10]	**20**	**CO5**

Name: Enrolment No:	UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2018

Course: Bigdata Storage **Semester: III**
Programme: BTech- CSE Bigdata
Time: 03 hrs. **Max. Marks: 100**
Instructions:

SECTION A

S. No.		Marks	CO
Q 1	Draw the block diagram of a Disc and explain the terms track, sector and cluster.	4	CO1
Q 2	Explain the terms Metadata and I-node. Also explain the relationship between Direct, Indirect and double-indirect blocks with the help of diagram.	4	CO2
Q 3	Explain the following terms with respect to Filesystem operations: e) Mounting f) Unmounting g) Symbolic links and Hard links h) Indexing and Logging	4	CO2
Q 4	Give two examples each of the common file systems used in Windows and Linux based systems. Also differentiate between FAT32 and NTFS.	4	CO2
Q 5	What are the two major classification of storage media? Explain the terms Tape Drive, HDD and SDD.	4	CO2
	SECTION B		
Q 6	What are five major reasons favoring NoSQL approach? Explain how NoSQL approach helps in improving Throughput and Availability.	5+5	CO1
Q 7	Explain the four major categories of NoSQL databases. Explain in detail the two NoSQL database types that are used by Blogging Platforms.	5+5	CO1
Q 8	Explain in detail the two most commonly used Data Distribution Models. Also differentiate between the two models used for data replication.	5+5	CO3
Q 9	Differentiate between Algorithmic Sharding and Dynamic Sharding. Explain at least five challenges faced during sharding. **OR** How does Range sharding strategy differ from Lookup sharding strategy? Explain the following terms : a) Shard Key b) Logical Shard	5+5	CO3

	c) Physical Shard d) Heirarchical Key e) Entity Groups		
	SECTION-C		
Q 10	What are the two major goals behind the design of DFS architecture? Explain the concept of replication in DFS. Explain GFS architecture in detail.	**5+5+10**	**CO4**
Q 11	How does the client-server DFS architecture differ from symmetric DFS architecture. Explain the HDFS architecture along with detailed explanation of NameNode, Secondary NameNode and DataNode. [5+15] OR What are the various factors responsible for maintaining Robustness and Fault Tolerance in HDFS? What is Checkpointing and how is it done in HDFS? Explain the terms Replica Placement, Replica Selection and Safemode. [10+5+5]	**20**	**CO5**