

25 F TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2075 Bhadra

Exam.	Regular		
Level	BE	Full Marks	80
Programme	BCT	Pass Marks	32
Year / Part	IV / II	Time	3 hrs.

Subject: - Big Data Technologies (*Elective II*) (CT76507)

- ✓ Candidates are required to give their answers in their own words as far as practicable.
- ✓ Attempt **All** questions.
- ✓ The figures in the margin indicate **Full Marks**.
- ✓ Assume suitable data if necessary.

1. Why distributed computing is necessary for big data? [5]
2. Define DFS. How client writes data in HDFS? Explain with the help of suitable block diagram. [10]
3. The data in big data warehouse is called hybrid data. Explain with suitable examples. [10]
4. How GFS differ from other File Systems? List out five distinct differences. [5]
5. What is the main role of GFS Master during read and write processes? How data and control messages flow in GFS architecture. Explain with suitable flow diagram. [10]
6. Map Reduce is the heart of Hadoop eco-system? Define work flow of Map reduce with suitable examples. [10]
7. Clock synchronization in DFS may be the big challenge. How this clock synchronization problem can be solved? [10]
8. Hbase, Cassandra and MongoDB are called column-oriented NoSQL database? How row-oriented database differ from column-oriented database? Explain with suitable examples. [10]
9. Write short notes on: [5×2]
 - a) Scoop and fume
 - b) Zookeeper
 - c) Oozie
 - d) Pig and Hive
 - e) Client-Server and Master-Slave architecture

Exam.	Regular		
Level	BE	Full Marks	80
Programme	BCT	Pass Marks	32
Year / Part	IV / II	Time	3 hrs.

Subject: - Big Data Technologies (Elective II) (CT76507)

- ✓ Candidates are required to give their answers in their own words as far as practicable.
- ✓ Attempt All questions.
- ✓ The figures in the margin indicate Full Marks.
- ✓ Assume suitable data if necessary.

1. a) Explain with example about the distributed system in Big Data. [8]
b) What is the role of Data Scientist? [4]
2. a) Explain the architecture of Google File System (GFS). [8]
b) What is availability and fault tolerance in Google File System? [5]
3. a) Explain in brief Data Flow technique of Map-Reduce Framework. [8]
b) What is Optimization and Data Locality in Map Reduce? [4]
4. Differentiate between structured and unstructured data and discuss the Taxonomy of NoSQL. [8]
5. Explain the components of Indexing and searching. [8]
6. a) Explain in brief five daemons of Hadoop. [8]
b) What is the role of Hadoop Distributed File System in Hadoop? [4]
7. Write short notes on: [5×3]
 - i) Elastic Search
 - ii) Hbase Architecture
 - iii) Functional Programming

Exam.	New Back (2066 & Later Batch)		
Level	BE	Full Marks	80
Programme	BEX, BCT	Pass Marks	32
Year / Part	IV / II	Time	3 hrs.

Subject: - Big Data Technologies (Elective II) (CT76507)

- ✓ Candidates are required to give their answers in their own words as far as practicable.
- ✓ Attempt All questions.
- ✓ The figures in the margin indicate Full Marks.
- ✓ Assume suitable data if necessary.

1. Why do we need data analytics process? Explain the role of Distributed computing in Big data. [5+5]
2. Why do we have large and fixed sized Chunks in GFS? What can be the demerits of that design? [10]
3. How is MapReduce library designed to tolerate different machines (map/reduce nodes) failure while executing MapReduce job? [10]
4. For following data, list the input to/output from both the map and reduce functions for getting maximum marks of each college. [10]

Student Name	College Name	Final Marks in %
Ram	ABC	70
Sita	ABC	80
Hari	ABC	60
Gita	XYZ	90
Rita	XYZ	80
Shyam	PQR	90
Laxmi	PQR	70
Gopal	PQR	60

OR

- What is the combiner function in mapreduce? Explain its purpose with suitable example. [10]
5. Explain the term NO-SQL. Explain CAP theorem with suitable block diagram. [3+7]
 6. Describe the typical components involved in search application. [10]
 7. What are different daemons in HADOOP cluster? Explain each in details. [3+7]
 8. Write short notes on any two of following. [2×5]
 - a) Shadow Master and Cluak services
 - b) Analyzers available in Lucene
 - c) Vertical and Horizontal Scalability

35 F TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2073 Bhadra

Exam.	Regular		
Level	BE	Full Marks	80
Programme	BEX, BCT	Pass Marks	32
Year / Part	IV / II	Time	3 hrs.

Subject: - Big Data Technologies (*Elective II*) (CT76507)

- ✓ Candidates are required to give their answers in their own words as far as practicable.
- ✓ Attempt **All** questions.
- ✓ The figures in the margin indicate **Full Marks**.
- ✓ Assume suitable data if necessary.

1. What are the current trends in big data analytics? What are the technical challenges and characteristics of big data? [10]
2. Explain the GFS Architecture. Why single master is not a bottleneck in GFS cluster. [5+5]
3. How does MAP-REDUCE work? Explain each step with suitable example. [5+5]
4. Discuss the architecture of Hbase in short. Explain eventual consistency and tunable consistency in context of Cassandra. [10]
5. Explain LUCENE architecture and its data indexing approach. [10]
6. What are the components of Hadoop? Explain each in brief. [10]
7. How do you find max and min occurrence of the words in a given text document. Explain. [10]
8. Write short notes on: (any two) [2×5]
 - a) CAP theorem
 - b) Role of Data Scientist in Big data
 - c) Amazon cloud
