TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2082 Chaitra

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

Subject: - Big Data Technologies *(Elective II) (CT 76507)*

- ✓ *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.*

IOE-NCE LIBRARY

1. What is the specific role of a distributed system in big data analysis? Discuss the current trends in big data analytics and explain why traditional data warehousing often fails to meet the demands of modern data stream. [5+5]
2. Explain the metadata stored by the GFS Master and explain why the master stores this information in the memory. Explain the control flow of a write mutation in GFS with a detailed diagram. [4+6]
3. Define functional programming and explain how this programming paradigm is suitable for processing big data. Describe the data flow in MapReduce framework with a suitable example. [4+8]
4. Define the term "NoSQL databases". Explain structured, unstructured and semi-structured data. [2+3]
5. Explain Cassandra database system with examples. [6]

OR

In the MongoDB database called 'syllabus', write MongoDB queries to create a collection called 'course' and insert five different documents for five courses. Your course should have at least 5 attributes like course name, credit hours etc.

6. How does Lucene index resolve read and write congestion? Explain distributed searching with elastic search. [5+5]
7. Explain the role of HDFS and MapReduce in the Hadoop framework and explain about HDFS architecture. [5+6]
8. Write short notes on: [4×4]
   a) Amazon EC2
   b) Hadoop YARN
   c) Scoop and flume
   d) Prescriptive analytics

***

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2082 Shrawan

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

Subject: - Big Data Technologies *(Elective II) (CT 76507)*

- ✓ *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.*

PNCE LIBRARY

1. Differentiate between horizontal scalability and vertical scalability. Explain the 5Vs of Big Data. [5+3]
2. Draw the architecture of Google File System (GFS). Explain the read and write algorithm in GFS. [4+4+4]
3. List down the advantages of functional programming paradigm. What do you understand by higher order functions? Explain. [3+2]
4. Explain the working of MapReduce framework with reference to WordCount problem. [10]
5. What are the implications of CAP theorem? Explain the concept of "Eventual Consistency" in case of NoSQL databases. [5+5]
6. Define Elastic Search. What are the core components of Elastic Search? Explain Distributed Searching with Elastic Search. [2+3+5]
7. Explain in brief about five daemons of Hadoop. Differentiate between Google File System and Hadoop Distributed File System. [8+2]
8. Write short notes on : [3×5]
   a) NoSQL Vs Relational Databases
   b) Role of Data Scientist in Big Data Analytics
   c) Pig, Hive and Spark

***

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2081 Chaitra

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

Subject: - Big Data Technologies *(Elective II) (CT 76507)*

- ✓ *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.*

NCE LIBRARY

1. "The big justice we can do to Big Data is use it." Justify this statement. Explain the role of distributed computing in Big Data analytics. [5+5]
2. Why single master is not a bottleneck in GFS cluster? How does GFS achieve high availability and fault tolerance? What are the advantages and disadvantages of 64 MB chunk size in GFS? [2+4+4]
3. Highlight the significance of lambda function in functional programming paradigm. Briefly explain pure and impure functions. [2+3]
4. How is map reduce library designed to tolerate different machines (map/reduce nodes) failure while executing Map Reduce job? Explain the parallel efficiency of map reduce. [5+5]
5. Define the term "NoSQL". Why NoSQL databases don't follow ACID principle? With a neat diagram, explain the architecture of HBase. [2+3+5]
6. What is Lucene? Describe the typical components involved in search applications. [2+8]
7. Explain any three components of Hadoop Ecosystem. Write down the commands to upload, download, and list files in HDFS. Also, write down the command to display the output of a map reduce job in HDFS. [6+4]
8. Write short notes on following: [3×5]
   a) Optimization and data locality
   b) Cassandra
   c) Amazon cloud

***

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2080 Chaitra

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

Subject: - Big Data Technologies *(Elective II) (CT 76507)*

- ✓ 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.*

NCE LIBRARY

1. What are 5Vs of Big Data? Explain with examples. Justify how Distributed System plays a vital role in Big Data. [4+5]
2. List the various types of NOSQL database with their data structure and usage. [8]
3. Let us assume that YouTube as a client is still using GFS. Few 100 videos in YouTube are really popular and 10,000 of users are accessing it simultaneous. How do you think GFS handles this read scenario? Explain this process along with GFS architecture. [8]
4. What are the main components of Map-Reduce Job? Explain in brief Data Flow technique of Map-Reduce Framework. What happens if Data Node fails during write process? [4+4+3]
5. You are appointed by Ministry of Health as an Engineer to support its plan to perform Health data analytics of all governmental hospitals using a columnar NoSqL database : HBase. Explain the architecture of HBase and discuss how you design and deploy to fully exploit its features in this scenario. [9]
6. Explain the architecture of LUCENE and its role in a typical text search application. Also, Describe different analyzers available in Lucene in brief. [6+5]
7. Compare and Contrast the different mode of Hadoop Enviroment setup. Can we step up Hadoop in the cloud? If so, how and explain the scenarios when these different mode are preferable? [4+4]
8. Write short notes on: [4×4]
   a) Functional Programming
   b) Fault tolerance in GFS
   c) Recent Trends in Big Data
   d) Mongo DB

***

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2079 Chaitra

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

Subject: - Big Data Technologies *(Elective II) (CT 76507)*

- ✓ 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.*

NCE LIBRARY

1. What are the common challenges of Big-Data? Explain the Data Analytics Process. [5+5]
2. Draw the architecture of GFS. Explain the process of data reading and data writing in GFS. [2+8]
3. What is the Map- Reduce Framework and how does it work? Explain with examples. [10]
4. Differentiate SQL Vs No.SQL. Define Data consistency, Availability and partition tolerance interms of No. SQL .Explain which No.SQL supports of the above terms. [3+3+4]
5. Define LUCENE. Describe the typical components involved in the search application. [10]
6. Briefly describe the daemons of Hadoop. Explain the role of HDFS (Hadoop Distributed File System) in Hadoop and write down the syntax for file upload, download, list and view content of file commands for HDFS. [10]
7. Write short notes on: [4×5]
   a) Functional Programming
   b) Cassandara
   c) Elasticsearch
   d) Amazon cloud

***

11:53

2079 Shrawan

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2079 Shrawan

Exam.	[redacted]	Back	
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. Is Big data a problem or an opportunity? Justify. Explain roles of Data Scientist. [4+6]
2. Explain the Read/Write Operation of Google File with a detailed diagram. [10]
3. Why do we require a map-reduce framework? Explain different steps of map-reduce with an appropriate example. [3+9]
4. Define CAP theorem by highlighting it's use case. When do we require columnar databases? Explain with examples. [5+5]
5. What are data indexing steps? Describe the components of search application. [4+6]
6. Explain about the configuration modes of Hadoop. Give an overview of Hadoop ecosystem. [6+6]
7. Write short notes on: [4×4]
   a) Functional programming
   b) Elastic search
   c) Big data in cloud
   d) Characteristics of NoSQL database

***

IOE SYLLABUS

11:53 • 2079 Jestha

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2079 Jestha

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

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

- ✓ 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 Hadoop is used in big Data Analysis? How is Big Data Analytics helpful in increasing business revenue? Explain a suitable example. [10]
2. Explain the five phases of real time Big Data analysis. [10]
3. Explain the respective components of HDFS and YARN. How client writes data into HDFS? Explain with a suitable block diagram. [10]
4. What is the main role of Record Reader and Record Writer in Hadoop Map reduce? How many record Readers are used while 200 MB file size is used as an input file? Explain with the Hadoop Map reduce flow chart. [10]
5. During parallel execution how clock synchronization in Map reduce can be solved? Explain with a suitable database. [10]
6. Explain HBase components in cluster architecture and define the main role of zookeeper in HBase. Write HBase commands to store and select data in HBase database. [10]
7. Why Lucene index is so called TF-IDF? How Lucene index resolve the write and read congestion problem? Explain with a suitable block diagram. [10]
8. Write short notes on: [5×2]
   a) Internet of Things (IoT)
   b) MongoDB
   c) Distributed File System
   d) Hive
   e) Oozie

***

IOE SYLLABUS

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2078 Chaitra

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.*

NCE LIBRARY

1. What is big data? Explain 5 V's of big data by relating with real world use cases. [2+8]
2. Why was Google File System created? What are the assumptions made by GFS? Explain operations performed by master in GFS along with Architecture diagram. [2+2+6]
3. Why do we require Map-reduce framework? Explain in detail how failures are handle in Map-Reduce framework along with an appropriate example. [3+7]
4. List down the categories of NoSQL datases. Give an overview of architecture of it base; a distributed columnar database. [6+6]
5. List the different analyzers used in data indexing along with explanation. Explain the data indexing process. [5+5]
6. List down the different components installed in the hadoop cluster. Explain its workflow. How fault tolerance and scalability is handled by the hadoop cluster? [2+6+4]
7. Write short notes on: [4×4]
   a) Functional Programming
   b) Structured, Semi-structure and Unstructured data
   c) Apache Cassandra
   d) Distributed Searching

***

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2077 Chaitra

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.*

NCE LIBRARY

1. What do you know about the term "Big Data" and what are the five V's of Big Data? How Big Data are helpful in increasing business revenue? [6+6]
2. Explain GFS Architecture. Why single master is not a bottleneck in GFS cluster? [8+5]
3. Explain how MapReduce works with suitable example. Explain distributed execution in MapReduce with example. [6+6]
4. Explain Hbase with architecture. How can you model RDBMS table in Mango dB? Give an example. [4+6]
5. Explain the process of indexing and searching in Lucene with proper diagrams. [8]
6. Explain various components of Hadoop in brief. [10]
7. Write short notes on: [3×5]
   a) Elastic Search
   b) CAP Theorem
   c) JSON and XML

***

NATIONAL COLLEGE OF ENGINEERING

QUESTION BANK

Big Data Technologies

(Elective II)(CT76507)

BCT IV/II

NCE LIBRARY

BD015

BD015

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2076 Bhadra

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

Subject: - Big Data Technologies *(Elective II) (CT 76507)*

- ✓ 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 main characteristics of big data? [5]
2. Define HDFS? How client reads data from HDFS? Explain with the help of suitable block diagram. [10]
3. What are the differences between structured and unstructured data? Explain with suitable examples. [10]
4. How traditional data differ from big data? List out five distinct differences? [5]
5. Explain GFS architecture. How data and control messages flow in GFS architecture. Explain with suitable flow diagram. [10]
6. What is elastic search? What indexes will be used during elastic search. Explain with suitable example. [10]
7. What are the various components of Hbase database. Explain with a suitable block diagram. [10]
8. Why Hbase is called column-oriented NoSQL database built on top of HDFS? What are the commands to STORE, SELECT, MODIFY, and DELETE records from a table of Hbase. [10]
9. Write short notes on: [2×5]
   a) Master-Slave architecture
   b) Zookeeper
   c) Client-Server architerucre
   d) Hadoop Map reduce
   e) Application of Big data analytics

***

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 fiume
   b) Zookeeper
   c) Oozie
   d) Pig and Hive
   e) Client-Server and Master-Slave architecture

***

35F TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2074 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. 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

***

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

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

***

35F TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

Examination Control Division

2074 Magh

Exam.	Back		
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. How big data differ from traditional data? List out five distinct differences? [1+5]
2. What are the sources of structured, semi-structured and un-structured data in real-world? [10]
3. Define DFS. How client writes data in HDFS? Explain with help of suitable block diagram. [10]
4. Clock synchronization in DFS may be the big challenge. How this clock synchronization problem can be solved? [1+5]
5. How data and control messages flow in GFS architecture. Explain with suitable flow diagram. [10]
6. How GFS provides fault tolerance. How it allows tolerating chunk servers failures? [10]
7. How word count job is performed for the following file in HDFS using a Map-reduce flow chart? [10]

   File.txt(file size: 200MB)

   Hi how are you
   How is your job
   How is your family
   How is your brother
   How is your sister
   What is the time now
   What is the strength of Hadoop
8. What are the differences between row and column oriented database? Why Hbase, Cassandra and MongoDB are called column oriented NoSQL database? [10]
9. Write short notes on: [4×2]
   i) Zookeeper and Oozie
   ii) Pig and Hive

***

35E TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2072 Ashwin

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 technical challenges and characteristics of a big data? Who are the data scientists, list out their roles and skills. [6+6]
2. With diagram, explain general architecture of Google File System. [10]

OR

a) Why do we have single master in a GFS and millions of chunk servers? [4]
b) A cluster contains 1500 machines, each having 500GB disc capacity. Calculate approximate the number of the chunck servers, the blocks and the total available size if default chunck replica is 3 and 5 respectively. [6]

3. a) What is a map reduce? Explain the execution overview of the map reduce. [6]
   b) Draw the output of mapreduce of the following lines: [4]
   "big users big volume data cloud contributes bid data"
   "facebook has big users facebook operates big data"
4. a) Explain a CAP theorem. [5]
   b) Differentiate between a RDBMS and a NoSQL Databases. [3]
5. Explain taxonomy of a NoSQL databases. Explain Cassendra database in brief. [10]

OR

Using a MongoDB database,

a) Create a collections named "posts", insert following records: [3]
title: MongoDB, description: MongoDB is a NoSQL database, by: Tom,
Comments: We use MongoDB for unstructured data, likes: 100

i) Now write a query to search title of the post written by Tom. [3]
ii) Write mapReduce function to count number of posts created by various users. [4]

6. What is the Lucene? Describe the typical components involved in the search application. [10]
7. Explain various components of Hadoop in brief. [10]
8. Write short notes on: (any two) [5×2]

i) Combiner Functions
ii) Fault tolerant systems
iii) JSON
iv) Unstructured data

***

25

35F TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2072 Magh

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. What is a "Big Data"? How distributed systems help to solve the Big Data problems? [12]
2. Explain how master implements garbage collection and detects stale replica in a GFS. [10]
3. Why do we have large and fixed sized Chunks in a GFS? What are the demerits of that design? [10]
4. How a MapReduce library designed to tolerate different machines (map/reduce nodes) failure while executing MapReduce job? [8]
5. For following data, list the input to/output from both the map and reduce functions for getting maximum marks of each college. [10]

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

OR

What is the combiner function in mapreduce? Explain its purpose with suitable example.

6. What is the difference between a structured and unstructured data. Explain the eventual consistency and tunable consistency in context of Cassandra. [10]
7. What is an elastic search? Explain various types of analyzers. [2+8]
8. What are the components of the Hadoop? For a hadoop cluster with 128 MB block size, how many mappers will hadoop mapreduce form while performing mapper function on 1 GB of data. Justify with explanation. [10]

***

(25

35B TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
Examination Control Division
2071 Bhadra

Exam.	Regular / Back		
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 BIG-DATA challenges? Explain the data analytics process in terms of BIG-DATA. [3+7]
2. a) Explain the control flow of write mutation with diagram. [7]
   b) Explain the meta data stored by GFS master. [8]
3. a) Explain garbage collection implemented by GFS. Explain its purpose against implementing eager deletion for storage reallocation. [7]
   b) Explain CAP theorem and Eventual consistency. Also, explain the reason why some NoSQL databases like Cassandra sacrifice absolute consistency for absolute availability. [8]
4. How map-reduce works in distributed fashion? Describe the parallel efficiency of map-reduce with suitable block diagram. [3+7]
5. List out the HADOOP daemons. How HADOOP and GFS are similar interms of design architecture. [2+8]
6. Explain the term "NO-SQL". Justify "for distributed scenario normalization contradict the data availability". [3+7]

OR

Write down the map-reduce program to find the word frequency. [10]

7. What are the data indexing steps? Describe the components of search application. [3+7]

***