

Enrollment No: \_\_\_\_\_ Exam Seat No: \_\_\_\_\_

# C.U.SHAH UNIVERSITY

## Summer Examination-2017

Subject Name: Big Data and Analytics

Subject Code: 5TE02BDA1

Branch: M.Tech (CE)

Semester: 2

Date: 04/05/2017

Time: 02:00 To 05:00

Marks : 70

### Instructions:

- (1) Use of Programmable calculator and any other electronic instrument is prohibited.
  - (2) Instructions written on main answer book are strictly to be obeyed.
  - (3) Draw neat diagrams and figures (if necessary) at right places.
  - (4) Assume suitable data if needed.
- 

### SECTION – I

Q-1 Attempt the Following questions (07)

- a. What is HDFS?
- b. What is Hadoop?
- c. What is map?
- d. What is reduce?
- e. What is the function of sqoop?
- f. What is Pig?
- g. What is Hive?

Q-2 Attempt all questions

- a. Explain the role of data analyst with example. 07
- b. Explain how map reduce model can be used to count words in a file. Given 2 files. First file with content 'hadoop is fast' and second file with content 'hadoop is easy'. 07

OR

Q-2 Attempt all questions

- a. What are 3 V's of big data? Explain with example. 07
- b. List and explain 7 use cases of big data analytics in brief. 07

Q-3 Attempt all questions

- a. What is NoSql? Explain different category of NoSql with example for each. 07
- b. List different map reduce processes. Also explain role of each with necessary diagram. 07

OR

Q-3 a. What is predictive analytics? Explain one use case of predictive analytics. 07  
b. List difference between RDBMS and NoSql. Explain BASE compliance of NoSql 07



in detail.

## SECTION – II

- Q-4 Attempt the Following questions (07)
- What is data node?
  - What is the default block size in HDFS?
  - What is SNN?
  - List 2 Higher Level Abstractions for MR.
  - What is the significance of replication factor in HDFS?
  - How to do load balancing is performed in hadoop cluster?
  - What is the function of name node?
- Q-5 Attempt all questions
- What is the significance of hadoop safe mode? Explain any 4 hadoop safe mode commands. 07
  - With necessary diagram explain architecture of hadoop cluster. 07
- OR
- Q-5 a. List and explain difference between Legacy MR and Next Generation Map Reduce (YARN/MRv2). 07
- b. List and Explain different modes of Developing, debugging and deploying MR programs in brief. 07
- Q-6 Attempt all questions
- Given the sample.txt data below. 07  
1950 0 1  
1950 22 1  
1950 -11 1  
1949 111 1  
1949 78 1  
The 3 column's refers to year, temperature, and quality of data. Assume the sample.txt is stored in local file system.  
Write a Hive QL command to display the content of sample.txt.
  - Explain external and managed Tables in HIVE with example. 07
- OR
- Q-6 Attempt all Questions
- Explain design principles for humanizing Big data. 07
  - What is bag in Pig? Also explain the role of dump command in Pig with example. 07

