

Seat No.: \_\_\_\_\_

Enrolment No. \_\_\_\_\_

**GUJARAT TECHNOLOGICAL UNIVERSITY**  
**MCA Integrated - SEMESTER- VIII • EXAMINATION – SUMMER 2017**

**Subject Code: 4480601**

**Date: 27/04/2017**

**Subject Name: BIG DATA**

**Time: 10:30 am to 1:00 pm**

**Total Marks: 70**

**Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1** (a) i) Define the terms: “BI”, “NoSQL”, and “cluster”. **03**  
ii) Compare the terms “data analysis” and “data analytics” in brief. **02**  
iii) Compare the terms “file system” and distributed file system”. **02**  
(b) Describe CAP theorem along with appropriate example. **07**
- Q.2** (a) Define Big Data and state its characteristics. Why do we need Big Data analysis? Explain. **07**  
(b) State the developments in ICT that have acted as catalysts towards adoption of Big Data in businesses and explain each of them in brief. **07**
- OR**
- (b) i) State the stages of Big Data analytics lifecycle. **04**  
ii) State the difference between traditional data visualization and data visualization for Big Data. **03**
- Q.3** (a) State various categories of NoSQL databases and explain each of them in brief. **07**  
(b) Describe eventually consistent non-relational databases in detail. **07**
- OR**
- Q.3** (a) What is HBase? Explain how data is stored in HBase using appropriate example. **07**  
(b) What do you mean by the term “horizontal scaling”? Describe how horizontal scaling is achieved in MongoDB. **07**
- Q.4** (a) What is MapReduce? Explain various map tasks of MapReduce with appropriate example. **07**  
(b) i) Explain the terms parallel data processing and distributed data processing. **03**  
ii) State and describe the types of workloads that are processed in Big Data. **04**
- OR**
- Q.4** (a) Describe the concepts “task parallelism” and “data parallelism” along with appropriate example. Which of these concepts is adopted by MapReduce? **07**  
(b) Explain the reason why MapReduce is not considered suitable for real-time data processing and how it can be enabled to work in a near-real-time scenario. **07**
- Q.5** (a) Mention the types of data analysis and explain each of them in brief. **07**  
(b) State the techniques of machine learning and explain each of them in brief. **07**
- OR**
- Q.5** (a) i) Explain the terms “quantitative analysis” and “qualitative analysis”. **03**  
ii) Compare the terms correlation and regression. How are they used in Big Data analytics? Explain. **04**  
(b) i) Explain how group operation is performed while querying data in MongoDB. **04**  
ii) Compare mongodump and mongoexport utilities in brief. **03**

\*\*\*\*\*