



# MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL  
(A constituent unit of MAHE, Manipal)

II SEMESTER MTECH END SEMESTER EXAMINATIONS APRIL-MAY 2019

SUBJECT: BIG DATA ANALYTICS (CSE 5233)

Date of Exam: 29-04-2019

09:00 AM – 12:00 PM

Max. Marks: 50

### Instructions to Candidates:

- Answer **ALL** questions & missing data may be suitably assumed.

- 1A** Explain 5V's of BigData **3M**  
**1B** With example list different types of NoSQL databases. **2M**  
**1C** Write a Map-Reduce program to search a word in files. **5M**
- 2A** Briefly explain the components of YARN **2M**  
**2B** What is a mapper? Explain different phases of it. **3M**  
**2C** What are the features of data serialization formats? Compare different file formats with respect to these features. **5M**
- 3A** With the help of code snippet explain types of partition in Hive. **4M**  
**3B** What is serder? **2M**  
**3C** How skew is handled in Hive? Draw the diagram and write the code snippet. **4M**
- 4A** What are the features of PIG? **2M**  
**4B** How MapReduce chain is formed in RHadoop? Write the wordcount program using RHIPE. **5M**  
**4C** What are the components of web search? Explain any one of the components. **3M**
- 5A** For the graph shown in the Fig 5A. Apply Dijkstra's algorithm and find the shortest path using 0 as source node. Trace the parallel breadth first search algorithm. **5M**

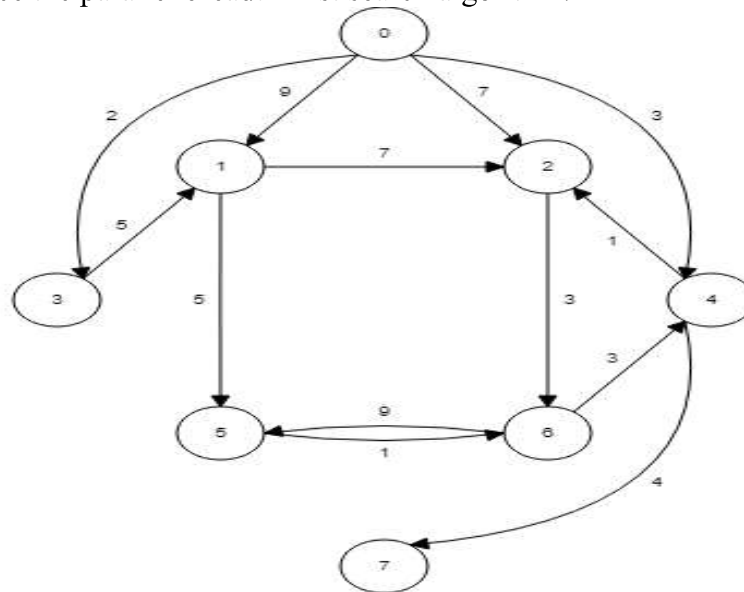


Figure 5A.

- 5B** Compress the digits from 1 to 10 using unary, Gamma, and Golomb code with  $b=5$  and  $b=10$ . **5M**