

DOF  
17112116

**KIIT UNIVERSITY, BHUBANESWAR**  
**MID-SEMESTER EXAMINATION- 2016**

Semester: 7th

Subject: DATA ANALYTICS (DA)

Paper Code: IT-3002

*Answer any FIVE questions including Question No.1 which is compulsory.*

*The figures in the margin indicate full marks.*

Time: 2Hrs

Full Marks: 25

- 
- Q1. Answer the following questions briefly** [1X5=5]
- a. Name the five key stakeholders of an analytic project and define their role?
  - b. Define Virtualization? How it handles the big data?
  - c. Explain the block concept in HADOOP with suitable example?
  - d. Explain the challenges in distributed computing and how it is handled by HADOOP?
  - e. Distinguish between Job tracker and Task tracker?
- Q2.a.** Explain the platform management layer of big data architecture? Discuss the technologies available in this layer? [2.5]
- b.** Explain the role of analytic engine in big data architecture? Which types of engines are used for analyzing data? [2.5]
- Q3. a.** Explain the architecture of HADOOP with diagram ? [2.5]
- b.** Write the steps of MAP REDUCE used to handle large data problem [2.5]
- Q4. a.** Explain the data model of HBase? Give suitable example of it [2.5]
- b.** Create a two dimensional array (4x4) and perform following operations [2.5]
- using R function
- i. display the matrix
  - ii. display only 2nd column
  - iii. change all elements of 2nd column
  - iv. display only 3rd row
  - v. display the new matrix
- Q5. a.** Suppose you have a user data in one file, website data in another file and you need to find top 5 most visited site by users aged 18-25. Draw the data flow diagram and the Pig query for the same? [2.5]
- b.** Explain the grouping and joining operators in Pig? [2.5]
- Q6.** Write short notes on any two of the following: [2.5 x 2]
- a. Zookeeper
  - c. NOSQL
  - b. HDFS (HA)