4ᵗʰ SEMESTER Regular EXAMINATION 2016-17
Big Data Analytics (BDA)
Branch:MBA
Max Marks: 100
Time:3 Hours
Q. Code:Z760

Answer Question No.1 and 2 which are compulsory and any four from the rest.
The figures in the right hand margin indicate marks.

Q1   **Answer the following questions:**   **(2 x10)**

a)   What is Big Data? Why do you analyze Big Data ?

b)   Write down any four industry examples of big data.

c)   What does data sand box mean? Write the name of a data sand box platform.

d)   Define predictive analytics and prospective analytics?

e)   Define Web Data.

f)   Differentiate supervised and unsupervised learning.

g)   Define decay window.

h)   Differentiate forward and backward prediction.

i)   What is real time analysis? Give an example of real time data processing?

j)   What is visual data analytics?

Q2   **Fill in the blanks with the appropriate answer from the choices given**   **(2x10)**
         below.

a)   Data in a RDBMS is _____ .
     (i)Structured   (ii)Semi-Structured   (iii)Un-Structured   (iv)Random

b)   e-Mail data is an example of _____ .
     (i)Structured   (ii)Semi-Structured   (iii)Un-Structured   (iv)Random

c)   Real time processing deals with _____characteristics of data
     (i)Variety   (ii)Velocity   (iii)Volatility   (iv)Structured

d)   _____ parameters are mainly used for defining big data.
     (i)Size & Function   (ii)Volume, velocity, variety (iii) Both (i)&(ii)   (iv) None

e)   A collection of independent computers that appears to its users as a
     single coherent system is _____ system
     (i)Client-Server   (ii)Centralised   (iii)Parallel   (iv)Distributed

f) Hadoop support _____type(s) of data formats
(i)Structured   (ii)Semi_Structured   (iii)Un-Structured   (iv)All the above

g) _____ Components of Hadoop is a data warehousing tool.
(i)Pig   (ii)Sqoop   (iii)Hive   (iv)Oozie

h) _____is the characteristics of data dealing with its retention.
(i)Variety   (ii)Velocity   (iii)Volatility   (iv)Variability

i) The example of un-structured data is _____.
(i) XML   (ii) HTML   (iii) Both (i)&(ii)   (iv)None

j) Major big data sources include _____
(i)Conventional Business Activity   (ii)User Generated   (iii)Machine Monitoring   (iv)All the above

Q3 a) Explain three Vs of Big Data ? Give two example of big data case studies. Indicate which Vs are satisfied by these case studies. **(8)**

b) Describe different sources to gather big data. **(7)**

Q4 a) What is big data analytics? Briefly explain different types of analytics. **(8)**

b) Discuss the challenges that are associated with big data. **(7)**

Q5 a) Briefly explain any four methods for statistical data analysis. **(8)**

b) Discuss responsibilities of data scientists and data engineers. **(7)**

Q6 What is Hadoop? What are its advantages ? Briefly explain the components of Hadoop ecosystem. **(15)**

Q7 a) What is MapReduce ? How does it process data ?Discribe MapReduce programming architecture with a neat diagrm. **(8)**

b) Discuss the different ways by which you can improve e-Commerce businesses with big data. **(7)**

Q8 **Write short notes on- (any three)** **(5x3)**

a) Cloud Computing

b) No SQL

c) IoT

d) Impala