



GL BAJAJ

Institute of Management & Research
Approved by A.I.C.T.E., Ministry of HRD, Govt. of India

Roll No.....

Plot No. 2, Knowledge Park-III, Greater Noida (U.P.) –201306

POST GRADUATE DIPLOMA IN MANAGEMENT (2019-21) END TERM EXAMINATION (TERM -V)

Subject Name	Big Data Analytics	Time: 02.30 hrs
Sub. Code	PGIT 03	Max Marks: 60

Note:

All questions are compulsory. Section A carries 10 marks: 5 questions of 2 marks each, Section B carries 30 marks having 3 questions (with internal choice question in each) of 10 marks each and Section C carries 20 marks one Case Study having 2 questions of 10 marks each.

SECTION - A

Attempt all questions. All questions are compulsory.

2×5 = 10 Marks

- Q. 1 (A): Discuss the role of Semi Structured data in decision making with Big Data Platforms?
Q. 1 (B): Discuss the role of Data Analyst in the Big Data Environment.
Q. 1 (C): Discuss the Veracity Aspect of data in Big Data Analytics.
Q. 1 (D): What are the major applications of Big Data for Production planning in any organization?
Q. 1 (E): Discuss the role of Task Shuffling in MapReduce.

SECTION - B

10 x 3 = 30 Marks

All questions are compulsory (Each question has an internal choice. Attempt any one (either A or B) from the internal choice)

Q. 2A: Compare the Traditional Data Analysis with Big Data Analytics. How CRISP-DM and SEMMA Methodology are not sufficient for the Big Data Analytics environment?

Q. 2B: Discuss the MapReduce Architecture for data manipulation and transactions in Big Data Environment. Also discuss the major section of MapReduce Master and Slave nodes for data processing.

Q. 3A: Write the R Programme for Reading the data from CSV file (Extracted from some database) with the following attributes:

(Product No., Product Name, Date of Manufacturing, Price, Batch No., Date of Expiry)

Once the data has been imported in R, then, Identify which products are on the verge of expiry as per the current date. **Export the data back to MS Excel file.**

Q. 3B: Write the R Programme for creating a Histogram with the following data about the sales by twenty salespersons in the retail store in a week:

36	25	38	46	55	68	72	55	36	38
67	45	22	48	91	46	52	61	58	55

Q. 4A: Write a R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from two given two vectors. Print the second row of the second matrix of the array and the element in the 3rd row and 3rd column of the 1st matrix.

Q. 4B: You have been provided two dataset with the values: A= (9.0,9.5,9.6,10.2,11.6) and B= (9.9,8.7,9.8,10.5,8.9,8.3,9.8,9.0). You need to conduct the Statistical analysis to identify the various aspects of the two datasets, including the T test. Also discuss the outcomes.

SECTION - C

Read the case and answer the questions

10×02 = 20 Marks

Q. 5: Case Study:

Shell: Big Data and Big Oil: *How Shell uses Big Data in practice*

With rising costs and dwindling natural resources, it's crucial that drilling takes place in the locations that will provide the biggest rewards. With this in mind, Shell collects data that allows them to predict the likely size of oil and gas resources by monitoring low-frequency seismic waves below the earth's surface. These waves register differently on sensors as they travel through the earth's crust, depending on whether they are passing through solid rock, liquids or gaseous material –indicating the probable location of hydrocarbon deposits.

Until recently, this method often proved unreliable; time-consuming, costly exploratory drills would usually be needed to confirm findings. Now, with an increased ability to capture, monitor, store and interpret large volumes of data, Shell's capability to identify reserves has improved. While a previous survey might have involved taking a few thousand readings, today it will typically involve more than a million readings. This data is then uploaded to analytics systems and compared with data from other sites around the world. The more closely the data matches the profiles of other drilling sites with abundant reserves, the higher the probability that a full-scale drilling operation will pay off.

When surveying potential new sites, Shell uses fibre-optic cables and sensor technology from Hewlett-Packard. All this data is stored and analysed using a Hadoop infrastructure that runs on Amazon Web Service servers. While the company doesn't publicise data volumes, it is known that the first test of the system collected around one petabyte of information, and it is estimated that Shell's data-driven oilfield approach has so far generated around 46 petabytes. Their analytics team is thought to consist of around 70 people.

As one of the largest companies in the world, Shell is undoubtedly able to invest heavily in new technology. But their reason for moving to a data-driven approach will ring true among companies of any size: a need to achieve better, more reliable results while reducing costs and risk. With the vast volumes of data available these days, this is a goal that almost any company can aim for and achieve. And this doesn't have to mean investing in expensive proprietary technology – third party data, whether it's freely available public data or part of a low-cost paid-for data solution, can provide a wealth of business-critical insights.

Question

Q5(A): Discuss the importance of Big Data Analysis for enhancing the performance of Oil exploration and productivity enhancement as mentioned in the case study. What are the other application areas for the company?

Q5(B): Identify the another application areas for the Shell company for application of Big Data Analytics and which kind of decisions may be supported by analytics?

Question Number	COs	Marks Allocated
Q. 1:	CO3	10 marks
Q. 2:	CO2	10 marks
Q. 3:	CO5	10 marks
Q. 4:	CO4	10 marks
Q. 5:	CO1	20 marks

Note: Font: Times New Roman, Font size: 12.