

Roll No.

--	--	--	--	--	--	--

H

TCS/TIT-703

**B. Tech. (CS/IT) (Seventh Semester)
End Semester EXAMINATION, 2017**

BIG DATA ANALYSIS

Time : Three Hours] [Maximum Marks : 100

Note : (i) This question paper contains five questions.

(ii) All questions are compulsory.

(iii) Instructions on how to attempt a question are mentioned against it.

(iv) Total marks assigned to each question are **twenty**.

1. Attempt any *two* questions of choice from (a), (b) and (c). (10×2=20 Marks)

(a) What do you mean by big data analysis ?
What are the major factors that motivate companies to go for Big Data Analysis ?
Why is Real-time Analytics important ?

[2]

TCS/TIT-703

- (b) Why is a life cycle important for analysis ?
Explain with suitable diagram big data analytics life cycle.
- (c) What is data preparation ? How are machine learning algorithms helpful in data preparation ?
2. Attempt any *two* questions of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) What do you mean by data discovery in big data analysis ? How is problem domain defined in data discovery ?
- (b) What do you mean by Machine Learning ? What are the statistical features used by machine learning ?
- (c) Why is data visualization important ? List tools for basic visualization techniques.
3. Attempt any *two* questions of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) Explain utility of big data management of business with suitable examples.
- (b) What are the *five* most important methods for statistical data analysis in big data ?
- (c) What are dependent and explanatory variables ? What is statistical modeling ? Explain with suitable example.

C-61

[3]

4. Attempt any *two* questions of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) Why is Clustering important ? Explain K means Clustering with suitable example.
- (b) Define the *three* key design principles of pig latin. Explain the basic building blocks of Hadoop with a neat sketch. Write short note on Hadoop ecosystem for understanding data analysis.
- (c) What is HDFS ? Define the role of combiner and partitioners in a map reduce application. Explain Map Reduce for analysis of unstructured data.
5. Attempt any *two* questions of choice from (a), (b) and (c). (10×2=20 Marks)
- (a) How do we select an appropriate technique of data analysis to achieve business objective ?
- (b) What are the *four* common deliverables of an analytics life cycle project ?
- (c) Write short notes on the following :
- (i) Risk Analysis through big data analysis
- (ii) Big Data Analysis over e-governance

TCS-703/TIT-703

290

C-61