



**K. K. Wagh Institute of Engineering Education & Research, Nashik**  
(An Autonomous Institute From A.Y. 2022-23)

InSem Examination-II Summer 2025	
Exam Seat No.:	
Academic Year: 2024-2025	Semester: VI
Class: TY IT	Program: B.Tech
Branch Code: INT	Pattern: 2022
Name of Course: Data Science and Big Data Analytics	Course Code: INT223012
Max. Marks: 30	Duration: 1.15 Hrs.

**Instructions:** Candidates should read carefully the instructions printed on the Question Paper and on the cover page of the Answer Book, which is provided for their use.

1. This question paper contains TWO pages.
2. Answer to each new question is to be started on a new page.
3. Assume suitable data wherever required, but justify it.
4. Draw the neat labelled diagrams, wherever necessary.
5. The last column indicates the Course Outcome and level of Blooms Taxonomy of the Question/sub-question.

**Marks CO**

**Question No. 1**

- 1 a) Explain 3 basic V's of Big Data. (3) CO1  
1 b) Compare shared-everything and shared-nothing architectures. (4) CO1

**Question No. 2**

- 2 a) Explain AI, ML, Data Science domains with a neat diagram. (4) CO1  
2 b) Explain Data warehouse 1.0 infrastructure layers with suitable diagram. (4) CO1

**Group OR**

- 2 c) Explain Big Data and Data Analytics with a neat diagram. (4) CO1  
2 d) What are Data processing infrastructure challenges? (4) CO1

**Question No. 3**

- 3 a) Explain Expectation of a Random Variable. (3) CO2  
3 b) What is hashing? What are desired properties of good hashing scheme. (4) CO2

**Question No. 4**

- 4 a) Explain Bernoulli and Binomial Random Variable with suitable expressions. (4) CO2  
4 b) A petrol station owner is considering the effect on his business (Superpet) of a new petrol station (Global) which has opened just down the road. Currently (of the total market shared between Superpet and Global) Superpet has 80% of the market and Global has 20%. Analysis over the last week has indicated the following probabilities for customers switching the station they stop at each week:

To->	Superpet	Global
Superpet	0.75	0.25

Global	0.55	0.45
--------	------	------

What will be the expected market share for Superpet and Global after another two weeks have past?

**Group OR**

- 4 c) Explain Flajolet Martin Algorithm. (4) CO2
- 4 d) The pattern of sunny and rainy days on the planet Rainbow is a homogeneous Markov chain with two states. Every sunny day is followed by another sunny day with probability 0.8. Every rainy day is followed by another rainy day with probability 0.6. Today is sunny on Rainbow. What is the chance of rain the day after tomorrow? (4) CO2

..... End of question paper.....