1. **Pune Institute of Computer Technology, Pune-43**

DEPARTMENT OF INFORMATION TECHNOLOGY

(Academic Year – 2022-23)

Question Bank (Unit I, II, III)

Ques. No.	Question	Max Marks	CO Mapped	Bloom's Learning Level
	Unit -I Questions			
1	Explain Big Data Processing flow and Big data processing cycle with a neat diagram.	05	CO-1	L1
2	Shared everything architecture suffers from limited scalability. Explain this by stating the difference in Shared everything and shared nothing architectures.	05	CO-1	L3
3	Explain shared-everything and shared-nothing architectures in detail with respect to Big Data.			
4	Explain 5V's for defining Big Data along with the factors responsible for data explosion?			
5	List and explain choices for reengineering the data warehouse.			
6	Discuss the processing complexities associated with the big data.			
7	Justify your answer with an example, "Data Science and Big Data" is the same or different.			
8	Define with Example Big data with 5 V's			
9	Enlist the impact of learning approaches in Big Data/ Explain different kinds of learning approaches.			
10	What is Data explosion in Big Data? Explain with examples.			
11	Explain the big data infrastructure requirement for mobile computing.			
12	Define Big Data. Enlist the differences and similarities in Big Data and Data Science with examples.			
13	What are the different learning approaches in Big Data? Explain with examples.			
	Unit-II Questions			
1	1% of the people have certain genetic defect. 90% of tests for gene detect the defect. 9.6% of the tests are false positive. If the person gets positive test result, what are the odds that they actually have the genetic defect?	05	CO-2	L5
2	In my town, it's rainy one third of the days. Given that it is rainy, there will be heavy traffic with probability 1/2, and given that it is not rainy, there will be heavy traffic with probability 1/4. If it's rainy and there is heavy traffic, I arrive	05	CO-2	L5

Question Bank (Unit I, II, III)

	late for work with probability 1/2. On the other hand, the probability of being late is reduced to 1/8 if it is not rainy and there is no heavy traffic. In other situations (rainy and no traffic, not rainy and traffic) the probability of being late is 0.25. You pick a random day. a)What is the probability that it's not raining and there is heavy traffic and I am not late? b)What is the probability that I am late?			
3	Assume that a man's profession can be classified as professional, skilled labourer or unskilled labourer. Assume that of the sons of professional men, 80 percent are professional, 10 percent are skilled labourers, and 10 percent are unskilled labourers. In the case of sons of skilled labourers, 60 percent are skilled labourers, 20 percent are professional and 20 percent are unskilled. Finally, in the case of unskilled labourers, 50 percent of the sons are unskilled labourers, and 25 percent each are in the other two categories. Assume that every man has at least one son, and form a Markov chain by following the rofession of a randomly chosen son of a given family through several generations. Set up the matrix of transition probabilities. Find the probability that a randomly chosen grandson of an unskilled labourer is a professional man.			
4	Explain Flajolet Martin Algorithm. List the limitations of algorithm and how will you overcome these limitations?			
5	A computer system can operate in two different modes. Every hour, it remains in the same mode or switches to a different mode according to the transition probability matrix P = [0.4, 0.6 0.6 0.4] i) Compute the 2-step transition probability matrix. ii) If the system is in mode I at 5:30pm, what is the probability that it will be in mode I at 8 : 30 pm on the same day?			
6	Explain following terms. i) Expectation ii) Pair wise independence			
7	Find the first 3 powers of following transition matrix with Marcov chain.			

Pune Institute of Computer Technology, Pune-43

DEPARTMENT OF INFORMATION TECHNOLOGY

(Academic Year – 2022-23)

Question Bank (Unit I, II, III)

	D= [0.9 0.1 0.2 0.8]			
8	Determine the distinct elements in below input stream of integers using Flajolet Martin algorithm. Consider hash function h(X)=6X+1 X=1,3,2,1,2,3,4,3,1 ,2,3,1			
9	Explain Bloom filter with Proper example.			
10	Explain Pairwise independent Hashing.			
11	Write a note on i) Random Variables and Joint Probability ii) Marcov Chains and Random Walks iii) Pair-wise independence and universal hashing iv) Data Streaming Models			
12	Explain Mean, Mode, Median, Variance, standard deviation.			
13	Explain the need of Correlation analysis and Analysis of Variance.			
	Unit-III Questions			
1	Explain how a single point of failure of a name node is managed in Hadoop?	05	CO-3	L4
2	What is NoSQL? What is its need? List advantages and disadvantages.	05	CO-3	L2
3	Explain the following terms i) Google File System ii) Heartbeat mechanism in HDFS			
4	Differentiate between SQL and NoSQL databases with example. What is the need to develop Big Data applications using NoSQL databases?			
5	Explain HDFS Read & Write operations in detail.			
6	What is the role of Sorter, Shuffler and Combiner in Map reduce Paradigm?			
7	Explain Hadoop ecosystem in detail.			
8	Explain the need of MapReduce in of Big Data. Define the architecture of MapReduce in Hadoop.			
9	Write a note on Textual ETL Processing.			
10	Neatly draw Hadoop Architecture and explain.			
11	List any 5 Hadoop shell commands with examples.			

Question Bank (Unit I, II, III)

12	Explain the role of Job Tracker and Task Tracker with neat diagram.			

----------******----------