Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (W), Mumbai : 400058, India

(Autonomous College of Affiliated to University of Mumbai)

Specia Examination
August 2023
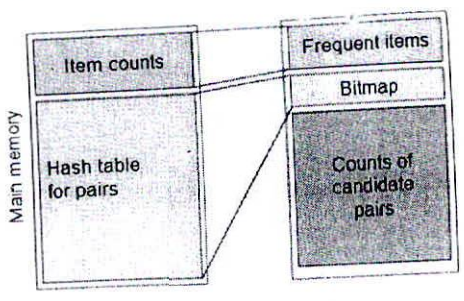
Maxi Marks : 100	Duration : 3 hours
Class : TEIT	Semester : VII
Course code: CS414	Branch : IT
Name of the course : Big Data Analytics	

Q No		Max Marks	CO	BL
1	With the example, explain any 5 Vs of Big data.	10M	CO1	4
2	Draw any 5 different types of data visualization graphs.	10M	CO1	5
3	Consider a movie dataset. Describe the dataset. Write a Hadoop map reduce code in java to print the most popular movies. OR Consider a dataset of online games. Describe the dataset. Write a Hadoop map reduce code in java to print the top 5 long running games.	10M	CO2	3
Q4	(a) You decide to use the page rank algorithm r = Mr, on some web graph whose matrix M is as given below where column represents outlink and row represents inlink.. Numbers in each cell of the matrix represent edge weight. $\begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 20 \end{pmatrix}$ Compute rank vector r, using Eigen vector formulation.	10M	CO5	3
	(b) Draw a web graph with spider trap and a dead node. Using this graph describe the technique that can be used to find page rank of web pages in a web graph that you have drawn.	10M	CO5	3
Q5	Bloom filters can have false positives. Explain with example. OR With example, explain the big data algorithm for counting distinct elements in a data stream.	10M	CO4	3
Q6	Write short note on Hive OR Write short note on HBase	10M	CO3	2
Q7	Write a short note on HDFS	10M	CO3	2

Q8	Following diagram shows technique used to find frequent items in big data. Name this technique and compare it with traditional technique to suit big data requirement.	10M	CO4	2
Q9	Compare traditional database management system vs stream management system.	10M	CO3	2

In the diagram (Main memory): Item counts, Hash table for pairs, Frequent items, Bitmap, Counts of candidate pairs.