END SEMESTER EXAMINATION - OCTOBER 2025

SEMESTER 7: INTEGRATED M. Sc. PROGRAMME IN COMPUTER SCIENCE – DATA SCIENCE

COURSE : 21UP7CRMCP25: DATA ENGINEERING IN DATA SCIENCE

(For Regular - 2022 Admission and Supplementary 2021 Admission)

Time: Three Hours Max. Weightage : 30

PART A

Answer any 8 questions

1. Given a relation schema R =(x,y,z,w) with functional dependencies F ={x->y, z->w}. All attributes take single and atomic values only. List the normal forms satisfied by this table. What will be the primary key for this table? (A)
2. Differentiate between location transparency and acess transparency. (U)
3. Consider an ordered file with r = 300,000 records stored on a disk with block size B = 4,096 bytes. File records are of fixed size with record length R = 100 bytes. Calculate the following.
   a) Blocking factor
   b) Number of blocks needed for the file. (E)
4. List the properties of big data. (R)
5. Distinguish between a query tree and a query graph. (U)
6. Find the cost of performing sort merge join of relations R and S(files are already sorted on the join attributes) (E)

Relation R: 600 blocks,12000 records
Relation S: 900 blocks,18000 records
js (join selectivity)=0.02
bfr_{RS} =50

7. Give examples for formatting data in python. (U)
8. List the aggregation methods in python. (R)
9. Write python code to read a json file. (Cr)
10. Define a data pipeline. (U)

(1 x 8 = 8 Weight)

PART B

Answer any 6 questions

11. Discuss on various NoSql databases. (U)
12. With the help of a diagram explain the structures of internal node and leaf node in a B+ tree. (An)
13. Explain the various windowing techniques in streaming data. (R)
14. Discuss on the data model of a datawarehouse. (U)

15. Discuss various methods in python to identify missing values and filling missing/na values in columns (U)
16. Discuss the methods for binning of data in python. (Cr)
17. Write a program to web scrape using beautifulsoup. (Cr)
18. Explain screen reading with selenium with a suitable program. (Cr)

(2 x 6 = 12 Weight)

PART C

Answer any 2 questions

19. Explain various indexing methods used in a database. (U)
20. Discuss various join algorithms used in query processing. (U)
21. Explain the various methods for fuzzy matching in python. (Cr)
22. Write a program to illustrate the usage of lxml module for web scraping. (Cr)

(5 x 2 = 10 Weight)