E/CMPN/Sem-VIII (CBSGS)/Data Warehouse & Mining/Nov-16

Q.P. Code : 733601

(3 Hours) **[Total Marks : 80**

Note: 1. Question No.1 is compulsory
2. Attempt any **Three** questions out of remaining questions
3. Assume suitable data wherever necessary and state them clearly

Q1 a) Consider following dimensions for a Hypermarket chain: Product, Store, Time and Promotion. With respect to this business scenario, answer the following questions. Clearly state any reasonable assumptions you make. Design a star schema. Whether the star schema can be converted to snowflake schema? Justify your answer and draw snowflake schema for the data warehouse (clearly mention the Fact table(s), Dimension table(s), their attributes and measures). [10]

b) Define linear, non-linear and multiple regressions. Plan a regression model for Disease development with respect to change in weather parameters. [10]

Q2 a) What is meant by metadata in the context of a Data warehouse? Explain the different types of meta data stored in a data warehouse. Illustrate with a suitable example. [10]

b) Describe the various functionalities of Data mining as a step in the process of knowledge Discovery. [10]

Q3 a) In what way ETL cycle can be used in typical data ware house, explain with suitable instance. [10]

b) What is Clustering Technique? Discuss the Agglomerative algorithm with the following data and plot a Dendrogram using single link approach. The table below comprises sample data items indicting the distance between the elements. [10]

Item	E	A	C	B	D
E	0	1	2	2	3
A	1	0	2	5	3
C	2	2	0	1	6
B	2	5	1	0	3
D	3	3	6	3	0

TURN OVER

— HUMAN MACHINE INT.
— PARALLEL & DIS. SYS.

Q4 a) Discuss how computations can be performed efficiently on data cubes. [10]

b) A database has five transactions. Let min-support=60% and min-confidence = 80%. Find all Find frequent item sets by using Apriori Algorithm. T_ID is the transaction ID. [10]

T_ID	Items bought
T-1000	M, O, N , K, E, Y
T-1001	D, O, N, K, E, Y
T-1002	M, A, K, E
T-1003	M, U, C, K, Y
T-1004	C, O, O, K, E

Q5 a) Differentiate [10]

i. OLTP Vs. OLAP

ii. Data Warehouse Vs. Data Mart

b) Why naive Bayesian classification is called "naive"? Briefly outline the major ideas of naive Bayesian classification. [10]

Q6 Write short notes on any **four** of the following: [20]

i. Application of Data Mining to Financial Analysis

ii. Fact less Fact Table

iii. Indexing OLAP data

iv. Data Quality

v. Decision Tree based Classification Approach

B.E./ Comp./ Sem VIII (CBSGS) Nov/Dec '16

HUMAN MACHINE INTERACTION

Q.P. Code : 724100

(3 Hours) **[Total Marks : 80**

N.B. : (1) Q **1** is **Compulsory**
(2) Solve any **3** from remaining Questions
(3 Please specify your answers with neat sketches and Example wherever necessary
(4) Assume any data if not specified

1. Solve any **five** of the following. **20**
   (a) What are the three levels of processing? Explain
   (b) Explain Goal Directed design process in brief.
   (c) Mention steps in constructing persona
   (d) What do you mean by direct manipulation and indirect manipulation
   (e) Discuss issues related to Long Term Memory and short Term Memory.
   (f) What do you mean by keyboard accelerators?

2. (A) How reading is important in UI Design? Write your comments related to the quote "Poor Design may affect Reading". **10**
   (B) In the state of Maharashtra, Water Distribution Company want to provide self- help portal for its customers. The portal consists of online meter logging facility, Bill Payments, VDS i.e Voluntary Deposit Scheme for Bill. Complaints and other facilities. Being a Subject Matter Expert (SME) provide the detailed analysis and for the same provide the Interface that will be used by people in all Districts of Maharashtra **10**

3. (A) What are the advantages and Disadvantages of Digital or Graphical systems? Explain in brief. **10**
   (B) It is necessary to provide state of an art digital KIOSK for Rural India where citizens can register for Birth/ Death Certificates, Insurance premium payments, Postal Schemes such as Investments , Money Transfer etc. The application will be easy and multilingual to be configured in Local Language. Provide suitable Analysis and Interface design for the same. **10**

[TURN OVER

4. (A) Provide all factors of Interface design? Provide innovative web application by integrating the technologies that are used in Interface design. **10**

(B) What do you mean by device based and screen based control. Explain? **10**

5. (A) What do you mean by response time? What are various methodologies adopted for Feedback and guidance? **10**

(B) Provide a systematic design analysis for Municipal Corporation's Mobile App; to perform Tax and billing related transactions such as Registration of client, Describe the Property, get Tax and water Charge bills, pay tax, Complaints and many other relevant operations. **10**

6. Write Short Notes on ANY **FOUR** **20**

(a) Qualitative and Quantitative Research
(b) Mental Model
(c) Gestalt's Principles
(d) Graphics Icons and Images
(e) Colors
(f) Web Navigation

B.E./SEM VIII (CBSGS)/COMP. DEC '16

PARALLEL & DIST. SYSTEMS

Q.P. Code : 735403

(3 Hours) **Total Marks : 80**

Note : 1) **Q.No.1** is **compulsory**
2) Attempt **any three** out of remaining **five** questions.
3) Assume suitable data wherever required with justification.

1. (a) Explain with example Amdahl's law for measuring speed up performance of parallel systems. **5**
   (b) Define various pipeline performance measures. **5**
   (c) State the goals of distributed system. **5**
   (d) State the desirable features of global scheduling algorithm **5**

2. (a) Write a note on Pipeline Hazards. **10**
   (b) Explain in brief any three classification of parallel Architecture. **10**

3. (a) Write a note on Election algorithm. **10**
   (b) Explain the concept of Remote Procedure Call **10**

4. (a) Give one example that can be solved effectively with an SIMD architecture. **10**
   (b) Explain in brief the software concept of distributed systems. **10**

5. (a) Explain the need of client centric consistency models as compared to data centric consistency model. Explain any two client centric consistency model. **10**
   (b) Explain Load balancing approach in distributed system. **10**

6. Write short notes on (**Any two**):- **20**
   (i) Andrew file system (AFS).
   (ii) Raymond's Tree based algorithm.
   (iii) Code Migration

B.E. / SEM VIII (CBSGS) COMPUTER NOV '16

~~CBSGS~~

MACHINE LEARNING **Q.P. Code : 724303**

(3 Hours) **Marks: [80]**

N.B : (1) Question No. **1** is **Compulsory**.

(2) Attempt any **Three** questions out of the remaining **Five** questions.

(3) Figures to the right indicate full marks.

(4) Assume any suitable data wherever required, but justify the same.

1) (a) What are the issues in Machine Learning? (05)

(b) Explain Regression line, Scatter plot, Error in prediction and Best fitting line. (05)

(c) Describe the essential steps of K-means algorithm for clustering analysis. (05)

(d) What is SVM? Explain the following terms: hyperplane, separating hyperplane, margin and support vectors with suitable example. (05)

2) (a) Explain in detail Temporal Difference Learning. (08)

(b) Create a decision tree for the attribute "class" using the respective values: (12)

eyecolour	married	sex	hairlength	class
brown	yes	male	long	football
blue	yes	male	short	football
brown	yes	male	long	football
brown	no	female	long	netball
brown	no	female	long	netball
blue	no	male	long	football
brown	no	female	long	netball
brown	no	male	short	football
brown	yes	female	short	netball
brown	no	female	long	netball
blue	no	male	long	football
blue	no	male	short	football

TURN OVER

3) (a) What are the different Hidden Markov Models? (10)

(b) What is Reinforcement Learning? Explain with the help of an example. (10)

4) (a) **Apply K-means algorithm on given data for k=3. Use $C_1(2)$, $C_2(16)$ and $C_3(38)$ as initial cluster centres.**
Data: 2, 4, 6, 3, 31, 12,15,16, 38, 35, 14, 21, 23, 25, 30

(b) Explain with suitable example the advantages of Bayesian approach over classical approaches to probability. (10)

5) (a) Explain in detail Principal Component Analysis for Dimension Reduction. (10)

(b) Find optimal hyperplane for the data points: (10)

{(1,1), (2,1), (1,-1), (2,-1), (4,0), (5,1), (5,-1), (6,0)}

6) Write Short Notes on the following: (**Any two**) (20)

a. Machine Learning applications

b. Classification using Back Propagation Algorithm

c. Issues in Decision Tree

B.E. / SEM VIII COMPUTER NOV '16
CCBSGS)

Q. P. Code : 725403

BIG DATA ANALYTICS.

(3 Hours) **[Total Marks : 80]**

Note : (1) Question No.1 is compulsory.
(2) Attempt any three from the remaining questions.
(3) Assume suitable data.

1. (a) Define the 3 V s of Big Data **05**
   (b) Find Manhattan distance (Ll-norm) and Euclidean distance (L2-norm) for the following points Xl = (1, 2, 2), X2 = {2, 5, 3} **05**
   (c) Explain how are dead ends handled in Page Rank. **05**
   (d) Give problem in Flajolet-Martin (FM) algorithm to count distinct elements in a stream. **05**

2. (a) Explain different NoSQL data architecture patterns. **10**
   (b) Give Hadoop Ecosystem and briefly explain its components. **10**

3. (a) Give 2-step Map Reduce algorithm to multiply two large matrices. **10**
   (b) i. Give Map Reduce algorithm for Natural Join of two relations. **10**
   ii. Give Map Reduce algorithm to perform Intersection of two sets.

4. (a) Explain abstract architecture of Data Stream Management system (DSMS). **10**
   (b) Explain how to compute Page Rank for any web graph. **10**

5. (a) Explain Park Chen Yu algorithm for counting frequent item sets. **10**
   (b) Explain CURE algorithm for large scale clustering. **10**

6. (a) Explain with example collaborative and content based filtering in a recommendation system. **10**
   (b) Write a note on Social Network Graphs. **10**

****