uploaded.

Course Code : CST 406 EIQU/RW – 16 / 1679

Eighth Semester B. E. (Computer Science and Engineering) Examination

DATA WAREHOUSING AND MINING

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry marks as indicated against them.
(2) Due credit will be given to neatness and adequate dimensions.
(3) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) Describe data warehouse development life cycle with neat sketch.

OR

(b) What is CUBE ? If we create CUBE for retail application with three dimensions for time, product and store, illustrate with an example how the subcubes in the lattice can be created. 4

(c) Explain the following data warehouse model :—

(a) Enterprise warehouse.

(b) Data Mart.

(c) Virtual warehouse. 6

2. (a) In a STAR schema to track the shipment for a distribution company, the following dimension tables are found :

(i) Time,

(ii) Customer ship–to,

(iii) Ship–from,

(iv) Product,

(v) Type of deal, and,

(vi) Mode of shi[illegible] Review these [illegible]sions and list the possible attribute for each dimensio[illegible]les. Also, designate a primary key for each table.

OR

(b) Analyze that a da[illegible]use consists of the three dimensions time, doctor, and patient, [illegible]two measures count and charge, where charge is the fee that a [illegible] charges a patient for a visit.

(i) Draw a S[illegible] schema diagram for the above data warehouse.

(ii) Starting wi[illegible]base cuboid [day, doctor, patient], what specific OL[illegible]ations should be performed in order to list the to[illegible] collected by each doctor in 2004 ?

(iii) To obtain [illegible] list, write an SQL query assuming the data is sto[illegible] relational database with the scheme fee (day, mont[illegible] doctor, hospital, patient, count, charge.

Suppose each dim[illegible]s four level associate with it. How many cuboids will this [illegible]ain (including the base and apex cuboids)? 5

(c) What are the dif[illegible]AP operations that can be performed on a cube ? As a se[illegible] on the project team of publishing company exploring the opti[illegible] data warehouse describe the merits of OLAP and how it will [illegible]ntial in this environment. 5

3. (a) Bring out the di[illegible]between :—

(i) Traditional [illegible]and Index Organized Tables (IOT).

(ii) Bitmap in[illegible] B–tree index. 5

(b) Define Partitioning[illegible] partitioning is essential in data warehouses? Narrate each parti[illegible]technique with examples.

OR

Bring out query o[illegible]on in the context of data warehouse. Explain how query optimi[illegible]an be performed in data warehouse system. 5

4. (a) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are :

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

(i) What is mean and median of data ?

(ii) What is the mode of data ? Comment on the data modality (i.e. bimodal, trimodal, etc). What is themidrange of the data?

(iii) Find first quartile (Q1), third quartile (Q3), IQR of the data ?

(iv) Give the five-number summary of the data.

(v) Show a boxplot of the data. 5

(b) Describe data mining ? Answer the following :—

(i) Is it a simple transformation of technology developed from databases, statistics, and machine learning ?

(ii) Describe the steps involved in data mining when viewed as a process of knowledge discovery. 5

5. (a) A database has five transactions. Let min_sup=60% and min_conf=80%.

TID	Items_bought
T100	{M,O,N,K,E,Y}
T200	{D,O,N,K,E,Y}
T300	{M,A,K,E}
T400	{M,U,C,K,Y}
T500	{C,O,O,K,I,E}

(i) Find all the frequent item sets using Apriori algorithm.

(ii) List all the storng association rules (with support s and confidence c). 7

(b) Compute accuracy, er [illegible] sensitivity, specificity, precision and recall for the following co [illegible] matrix.

Classes	[illegible] computer=yes	Buys_computer=no
Buys_computer=yes	7954	146
Buys_computer=no	512	3588

3

[illegible]R

(c) Clustering is recogn [illegible] an important data mining task with broad applications. Give [illegible] ation example for each of the following cases :

(i) An applicati [illegible] uses clustering as a major data mining function.

(ii) An applicati [illegible] uses clustering as a prepocessing tool for data pr [illegible] for other data mining tasks. 3

6. (a) Given the followin [illegible] ments for the variable age : 18, 22, 25, 42, 28, 43, 33, 35 [illegible] standardize the variable by the following:

Compute :

(i) The mean [illegible] deviation of age.

(ii) The z-score [illegible] first four measurements.

[illegible]R

(b) Both K-means and [illegible] s algorithm can perform effective clustering. Illustrate the stren [illegible] eakness of k-means in comparison with k-medoids. 4

(c) Describe each of the [illegible] algorithm in terms of the following criterion:

(1) Shape of th [illegible] r that can be determine.

(2) Input param [illegible] must be specified.

(3) Limitations.

(4) Time complexity of the algorithm.

(i) K–means.

(ii) K–medoid.

(iii)CLARA. 6

Course Code : CST 407 EIQU/RW – 16 / 1680

Eighth Semester B. E. (Computer Science and Engineering) Examination

INFORMATION SECURITY

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) Question **1** is compulsory.

(2) Solve Q. **4** or Q. **5**.

(3) All questions carry marks as indicated against them.

(4) Due credit will be given to neatness and adequate dimensions.

(5) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) List and briefly define types of cryptanalytic attacks based on what is known to the attacker. Also define avalanche effect. 7

(b) Decrypt the following using Single columnar transposition if

Keyword = APPLE

Ciphertext : TSUTPIILRSTSOANIHAMROOICNASN. 3

2. (a) Relate any real life applications to each block cipher Modes of operations, and discuss the concept in brief.

OR

(b) Write down the process for function key generation in DES Encryption. Write the difference between Conventional and asymmetric cryptography. 10

3. Solve any **One** :—

(a) Demonstrate Man in-middle attack in Diffie-Hellman key exchange algorithm. Derive the proof of equations for showing two keys calculation used at sender and receiver that produces identical result. And solve the following :—

user A and user B wants to establish secret key using the Diffie-Hellman Key exchange Protocol. assuming the values as $n = 11$ $g = 5$ $x = 2$ and $y = 3$ Find out the values of A, B and secret key (k1, k2). 10

(b) Can you use RSA [illegible] it Public and private key ? If so describe how. Consider RSA [illegible] n with public key 55 and public exponent e = 3.

(i) How many [illegible] are in Z*55 ?

(ii) Compute the [illegible] exponent d.

(iii) Compute the [illegible] on of the message m = 6.

(iv) Compute the [illegible] on of the cipher text c = 2 10

(c) For a user workstati[illegible] ical business environment, Discuss the potential locations of confide[illegible] ck ? What is FEPs Function ? Give its sketch. Design the relatio[illegible] een Encryption and protocol levels. 10

4. (a) Define the properti[illegible] h function. Comment on the security of Hash function with a su[illegible] mple code i. e. attack complexity of Weak collision and strong [illegible] . Also draw the application use of MAC that has an implementa[illegible] ssage Authentication and confidentiality where authentication is ti[illegible] text. 6

(b) Define trap door [illegible] Differentiate MD5 and SHA-I. 4

OR

5. (a) List the disadvant[illegible] MAC as a authenticator function. Write a valid reason to justify [illegible] HMAC cannot be trusted to be used in digital signature. 3

(b) Write the algorith[illegible] to perform a digital signature for any electronic document similarl[illegible] he mathematical formulation used to verify the signature by pre[illegible] at sketch with set of equation to analyse the difficulty level of [illegible] be tried for modifying the signature. 7

6. Solve any **Two** :—

(a) How to achieve ce[illegible] authentication, State the application of Kerberos? Answer can we h[illegible] ltiple Kerberos system installed in a distributed environment. 5

(b) Explain AH and [illegible] ocol in brief. 5

(c) Discuss Electronic payment process. State how tightly protocol security is built on such E-commerce transactions. 5

7. Solve any **Two** :—

(a) Can you produce a sample of Virus Structure. Show compression logic for virus programs. 5

(b) State two common techniques used to protect a password file and Explain distributed intrusion detection with the help of agent architecture. 5

(c) Define three classes of intruders. What is audit record analysis ? 5

Course Code : CST 408-2 EIQU/RW - 16 / 1681

Eighth Semester B. E. (Computer Science Engineering) Examination

Elective – III

DISTRIBUTED AND PARALLEL DATABASES

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) Each Question carry marks as indicated.
(2) Due credit will be given to neatness.
(3) Assume suitable data wherever necessary.
(4) Illustrate your answers wherever necessary with the help of neat sketches.

1. (a) Explain the shared nothing and Hierarchical architectures of parallel Database with neat sketch along with advantages and disadvantages. 6

Attempt any **One** question :—

(b) What do you mean client server architecture ? How this methodology is adopted in distributed DBMS processing ? 4

(c) How Inter-Operator parallelism is different from Intra-Operator parallelism ? with example. 4

2. (a) Explain the need of transparency in distributed database along with all three types of transparency levels. 4

Attempt any **One** question :—

(b) Consider Global Schema -> PLAYER (NUMBER, NAME, GAME)

Fragmentation Schema ->

PLAYER 1 = $SL_{\text{GAME = "CRICKET"}}$ PLAYER

PLAYER 2 = $SL_{\text{GAME = "VOLLEYBALL"}}$ PLAYER

Allocation Schema – >

PLAYER 1 : Site 1, 2

PLAYER 2 : Site 3, 4

(Assume that CRICKET and VOLLEYBALL are the only games) 4

(i) Write an application that moves a player having number 10 and Game "Cricket' to game "Volleyball" at level 2 and 3 of transparency. 4

(ii) Consider th[illegible] which selects the information of player for many possi[illegible]es of player number. Write this application accessing th[illegible]se after having collected several inputs from the terminal[illegible] 2

(c) Differentiate betw[illegible]l fragmentation and vertical clustering. Describe benefit equations fo[illegible] fragmentation. How can we reconstruct the global relation in vertical[illegible]ation ? 6

3. (a) What is distributed [illegible] Explain the following in the context of distributed deadlock preventio[illegible]

(i) Non-Preem[illegible]thod.

(ii) Preemptive[illegible] 5

(b) Explain the rules o[illegible]stamp mechanism of concurrency control Consider a data item x. [illegible](x) = 25 and WTM (x) = 20. Let the pair (Ri (x), TS) (Wi(x)[illegible]ote the read and write request of transaction Ti on the item x [illegible]estamp TS. Indicate the behavior of the basic timestamp method [illegible]llowing sequence of requests : (R1 (x), 19), (R2(x), 22), (W3 (x), 21), ([illegible] (R5 (x), 28), (W6 (x), 27)

OR

Discuss the two ph[illegible]nent protocol for supporting atomicity of distributed transaction. Also [illegible] it behaves in case of site failures. 5

4. (a) Consider the jo[illegible]$_{B}$ S; assume that R and S are at different sites and disreg[illegible]t of collecting the result of the join. Let C0 = 0 and C1 = [illegible]llowing profiles are given :
Size (R) = 50; car[illegible] val (A[R]) = 50; size (A) = 3
Size (S) = 5; car[illegible] val (B[S]) = 50; size (B) = 3
R $SJ_{A=B}$ S has [illegible] p = 0.2
S $SJ_{B=A}$ R has [illegible] p = 0.8
Give the transmi[illegible] of performing the join at the site R using Semi Join reducti[illegible] 5

OR

Consider Global S[illegible]
DOCTOR(DNUM, [illegible] DEPT)
PATIENT (PNUM, [illegible] DEPT, TREAT, DNUM)

CARE (PNUM, DRUG, QUAN)
Fragmentation Schema :
DOCTOR 1 = SL $_{\text{DEPT="SURGERY"}}$ DOCTOR
DOCTOR 2 = SL $_{\text{DEPT="PEDIATRICS"}}$ DOCTOR
DOCTOR 3 = SL $_{\text{DEPT\#"SURGERY AND DEPT\#"PEDIATRICS'}}$DOCTOR
PATIENT 1= SL $_{\text{DEPT = "SURGERY" AND TREAT=INTENSIVE}}$ PATIENT
PATIENT 2 = SL $_{\text{DEPT="SURGERY" AND TREAT\#INTENSIVE}}$ PATIENT
PATIENT 3 = SL $_{\text{DEPT\#'SURGERY"}}$ PATIENT
CARE1 = CARE SJ $_{\text{NUM=PNUM}}$ PATIENT 1
CARE2 = CARE SJ $_{\text{NUM=PNUM}}$ PATIENT 2
CARE3 = CARE SJ $_{\text{NUM=PNUM}}$ PATIENT 3

Translate the following global queries into fragment queries and use criteria 1 to 6 to simplify them.

(i) PJ $_{\text{NAME}}$ SL $_{\text{DRUG="ASPIRIN AND TREAT = "INCENTIVE"}}$ (DOCTOR JN $_{\text{DNUM = DNUM}}$ PATENT NJN CARE)

(ii) GB $_{\text{AVG(QUAN)}}$ SL$_{\text{DRUG="ASPIRIN"}}$ (CARE NJN SL $_{\text{TREAT="INTENSIVE"}}$ PATIENT)

(b) Explain the need of parametric queries. Also explain the use of cut operator in a parametric query with suitable example. 5

5. Attempt any **Two** questions :—

(a) Describe different types of failures in a distributed system. What actions should be taken to overcome them ? 5

(b) In case of a fully redundant database, how does a strict replica control protocol works ? 5

(c) How does voting based protocol behaves in network portioning ? 5

6. Attempt any **Two** questions :—

(a) Explain the features of TERADATA and GAMMA relational database. 5

(b) Explain the mapping and implementation of data warehousing on parallel system. Also explain the advantages and disadvantages of it. 5

(c) Explain the distributed and parallel approach for data mining. Also explain importance of parallel and distributed processing in the context of data mining techniques. 5

Course Code : CST 409 - 1

EIQU/RW - 16 /1682

Eighth Semester B. E. (Computer Science and Engineering) Examination

Elective - IV

WEB INTELLIGENCE AND BIG DATA

Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry equal marks. Figures to the right indicate marks.

(2) Carefully see the internal choices.

(3) Which Course Objectives (COs) are satisfied by the question is mentioned against each question.

(4) Assume suitable data and illustrate your answer with the help of neat sketches wherever necessary. Due credit will be given to neatness.

1. (a) Enlist various applications of Big Data. Give an example of any one application which utilizes intelligence of the system. 5

(b) What are different steps in building intelligent systems using Big data ? Elaborate on prediction. 5

OR

2. (a) What is a page rank ? How Google has been able to explore page rank in its search engine ? 5

(b) Elaborate on implementation of LSH technique for search space minimization. 5

3. (a) What is mutual information ? When is it combined with TF - IDF ? Does it give good measure of relevance ? 5

(b) There was a murder and an investingating team found a finger prints from a crime spot. There are 1000000 FPs available in the database of the angency against which they have to match the FP. Suppose the probability of finding minutia in random grid square of a finger print (FP) is 15%.

Contd.

If a grid is having [illegible] in all squares of a grid, then the corresponding grid of other **FP** [illegible] have the minuta with a probability of 85% if the FP is taken [illegible]e same finger. Consider each function f in a family of **F** is de[illegible] a 3 grid squares. f says 'yes' if both FPs have minutia in al[illegible] squares otherwise it says 'no'. If we choose 1000 such functio[illegible]ly chosen from F, find :—

(1) What is th[illegible]ility that F_1 will put finger-prints from the same [illegible]gether in at least one bucket ?

(2) What is th[illegible]ility that two finger-prints from different fingers wil[illegible]aced in the same bucket ?

(3) Calculate [illegible] negatives and % false positives. 5

OR

4. (a) Explain Sparse d[illegible] Memory. Can it be used for search space optimization ? Ho[illegible] 5

(b) Define mutual info[illegible]alculate in the given table the mutual information for the features [illegible] *movie* towards behaviour *sentiment*.

Count		Sentiment
2000	I liked the [illegible] am loving it.	positive
800	I hate thi[illegible] think it is waste of money.	negative
200	The movi[illegible]ne old story and quite a bore.	negative
3000	The movi[illegible]rated, fun and very interesting.	positive
1000	I'm enjoy[illegible]vie a lot and learning something too.	positive
400	I would e[illegible] a lot if I did not have to go for this movie.	negative
600	I did not [illegible]movie enough.	negative

Also justify that [illegible]formation can be a measure for selection of a feature. 5

5. (a) In the library of ten [illegible] documents a word w appears in 500 of them. In a particular d[illegible]d, the maximum number of occurences of a word is 20. Approxi[illegible]at is the TF.IDF score for w if it appears (a) once (b) ten tim[illegible] 4

(b) What is the significance o[illegible]nal probability for classification ? Considering the table given in Que[illegible]b, find the sentiment of the statement "I have hated this movie [illegible] boring and waste of money". using Naïve Bayes classification. Co[illegible]itable features to be included. 6

6. (a) Explain the approach for [illegible]reduce for preprocessing in many applciations. 5

(b) What is big table ? [illegible]es it store the data ? Explain. 5

7. (a) Explain the process m[illegible] map – reduce. 5

(b) What are sharded indice[illegible] Mongo DB gets an advantages of Sharded indexing ? 5

(c) With the help of suit[illegible]ple explain eventual consistency. 5

8. (a) Why Dremel is succe[illegible] proved to be the technology of future ? 5

(b) EXPLAIN Naïve Baye[illegible] and multiple Naïve classificers. Show the evolution of Bayesian [illegible]from the multiple Bayes classifer. 5

(c) What is a long tail [illegible]n ? Explain the problems associated with it. How are they h[illegible] 5