Course Code : CST 406 **CIOU/RS – 16 / 3486**

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 equal marks.
(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.
(5) Mobile phones are prohibited in examination hall.

1. Solve any **Two** :—

(a) Present an example where data mining is crucial to the success of a business. What data mining functionalities does this business need ? Can such patterns be generated alternatively by data query processing or simple statistical analysis ? 5

(b) Write a short note on data warehouse development life cycle. 5

(c) Explain the use of cube and rollup operators with an example. 5

2. Solve any **One** :—

(a) What is a data warehouse ? With the help of a neat sketch explain the various components in a data warehousing system. 10

(b) Explain various types of multi dimensional models. 10

3. (a) Explain query optimization with respect to data ware housing.

OR

Explain hash index and bitmap index with an example. 6

(b) Write a short note on load manager. 4

4. (a) Discuss the following schemes used for integration of a data mining system with a database or a data warehouse system.

(i) No coupling.

(ii) Loose coupling

(iii) Semi tight coupling.

(iv) Tight coupling. 8

(b) Suppose that the data for analysis includes the arrribute *age*. The *age* values for the data tuples are (in increasing order) 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.

(a) What is the mean of the data ? What is the median ?

(b) What is the mode of the data ?

(c) What is the midrange of the data ?

(d) Can you find (roughly) the first quartile (Q_1) and the third quartile (Q_3) of the data ? 2

5. (a) Apply the apriori algorithm to the following data set. State and discuss each step in the Apriori algorithm.

Trans ID	Items Purchased
101	Apple, Orange, Litchi, Grapes
102	Apple, Mango
103	Mango, Grapes, Apple
104	Apple, Orange, Litchi, Grapes
105	Pears, Litchi
106	Pears
107	Pears, Mango

108	Apple, Orange, Strawberry, Litchi, Grapes	
109	Strawberry, Grapes	
110	Apple, Orange, Grapes	

The set of items is {Apple, Orgame, Strawberry, Litchi, Grapes, pears, Mango}. Use 0.3 for the minimum support value.

OR

(b) Explain Bayesian classification with suitable example. 10

6. (a) Describe each of the following clustering algorithms in terms of the following criteria :—

(i) Shapes of clusters that can be determined.

(ii) Input parameters that must be specified and

(iii) Limitations.

(1) K-means

(2) K-medoids

(3) CLARA

(4) BIRCH

(5) DBSCAN 10

OR

(b) With relevant example discuss constraint based cluster analysis. 10

Course Code : CST 407 **CIOU/RS – 16 / 3487**

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

INFORMATION SECURITY

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry equal marks.
(2) Due credit will be given to neatness.
(3) Assume suitable data wherever necessary.
(4) Mobile phones are prohibited in examination hall.
(5) Solve Q. **5** or Q. **6**.

1. (a) Consider a block of plain text arranged in a matrix in row-major form. Can any permutation of element in the block be realized only by some sequence of row or column transposition (possibly interleaved) ? If yes, give an informal but convincing proof. If no, provide a counter example. 5

(b) Differentiate Worms and Virus. State example of Trojan horse problem. 5

OR

(c) What are three goals of security ? Explain various types of attack. 5

2. (a) Explain Fiestel cipher structure with suitable example. 5

(b) (i) The input to S-box 1 is 100011. What is the output ?

(ii) The input to S-box 8 is 000000. What is the output ? 5

3. (a) Demonstrate RSA algorithm. Perform Encryption and decryption using for the following

$P = 7$, $q = 11$, $e = 17$, $M = 8$ implement a code for extended Euclidian to verify the value of private key. 5

(b) Write the time complexities of RSA encryption and decryption (as a function of key size, k). Also justify with valid reason. 5

4. Discuss Man in Middle Attack Problem. Demonstrate a program for calculating primitive roots of any large prime number. Consider a Diffie-Hellman with a common prime q = 13 and a primitive root = 2,

(a) If user A has public key Y_A = 9 what is the A's Private key X_A ?

(b) If user B has public key Y_B = 3 what is the Shared secret key K ? 10

5. (a) Show the implementation of digital signature using any symmetric algorithm. 5

(b) Differentiate between message authentication code and hash function with examples. 5

OR

6. (a) Explain Working of Kerberos in detail. 5

(b) Explain two modes of IPsec Protocol. 5

7. Solve any **one** :—

(a) Show how SSL Handshake and record protocol helps in providing security features to the application. 10

(b) Illustrate an example of payment gateway with neat sketch. 10

(c) (i) State and explain the types of Firewall. 5

(ii) State differences between HTTP and SHTTP protocol functioning and write in brief the significance of trusted systems. 5

Course Code : CST 408-2

CIOU/RS – 16 / 3488

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

Elective – III

DISTRIBUTED AND PARALLEL DATABASES

Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry equal marks.
(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.
(5) Mobile phones are prohibited in examination hall.

1. Solve any **Two** :—

(a) Explain the following with a neat sketch along with the advanatges and disadvantages of each.

(1) Shared nothing Architecture.

(2) Hierarchical Architecture. 5

(b) Explain the following distributed DBMS architecture :—

(1) Client server architecture.

(2) Peer to peer architecture. 5

(c) List out Advantages and Disadvantages of distributed DBMS. 5

2. (a) Consider the following global, fragmentation and allocation schemata :

Global Schema : STUDENT (NUMBER, NAME, DEPT)

Fragmentation Schema : STUDENT 1 = $SL_{DEPT = "EE"}$ STUDENT

STUDENT 2 = $SL_{DEPT = "CS"}$ STUDENT

Allocation Schema : STUDENT 1 at sites 1, 2, 3

STUDENT 2 at sites 1, 2

Assume that "EE" and "CS" are the only possible values for DEPT

(a) Write an application that moves the student (student number is entered from the terminal) from department "CS" to department "EE", at levels, 1, 2, and 3 of transparency.

(b) Consider the case in which application (a) is repeated for many possible values of student number. Write the application accessing the database before collecting inputs from the terminal. 6

(b) What do you mean by transparency in the database ? Explain the following in brief :—

(1) Fragmentation transparency

(2) Location transparency. 4

3. Solve any **Two** :—

(a) Draw and explain the reference model for distributed concurrency control mechanism. 5

(b) What do you mean by distributed transaction ? List and discuss the properties that must be satisfied by a distributed transaction. 5

(c) What is distributed deadlock ? Explain the followings in the context of distributed deadlock prevention.

(1) Non-preemptive method.

(2) Preemptive method. 5

4. Solve any **Two** :—

(a) Define the following terms with proper example :—

(1) Qualified Relation

(2) Algebra of qualified relation. 5

(b) Comment on the following things when operations :—

(a) Selection followed by projection.

(b) Projection followed by selection, are performed on a relation.

(1) Cardinality of the result.

(2) Size of the result

(3) Distinct values for any attribute in result. 5

(c) Explain the different problems in query optimization in distributed query processing. 5

5. Solve any **Two** :—

(a) Write short notes on :—

(1) **Multimaster Replication.**

(2) **Snapshot Replication.** 5

(b) What do you mean by optimistic replication ? Explain various types of replication management policies for distributed database. 5

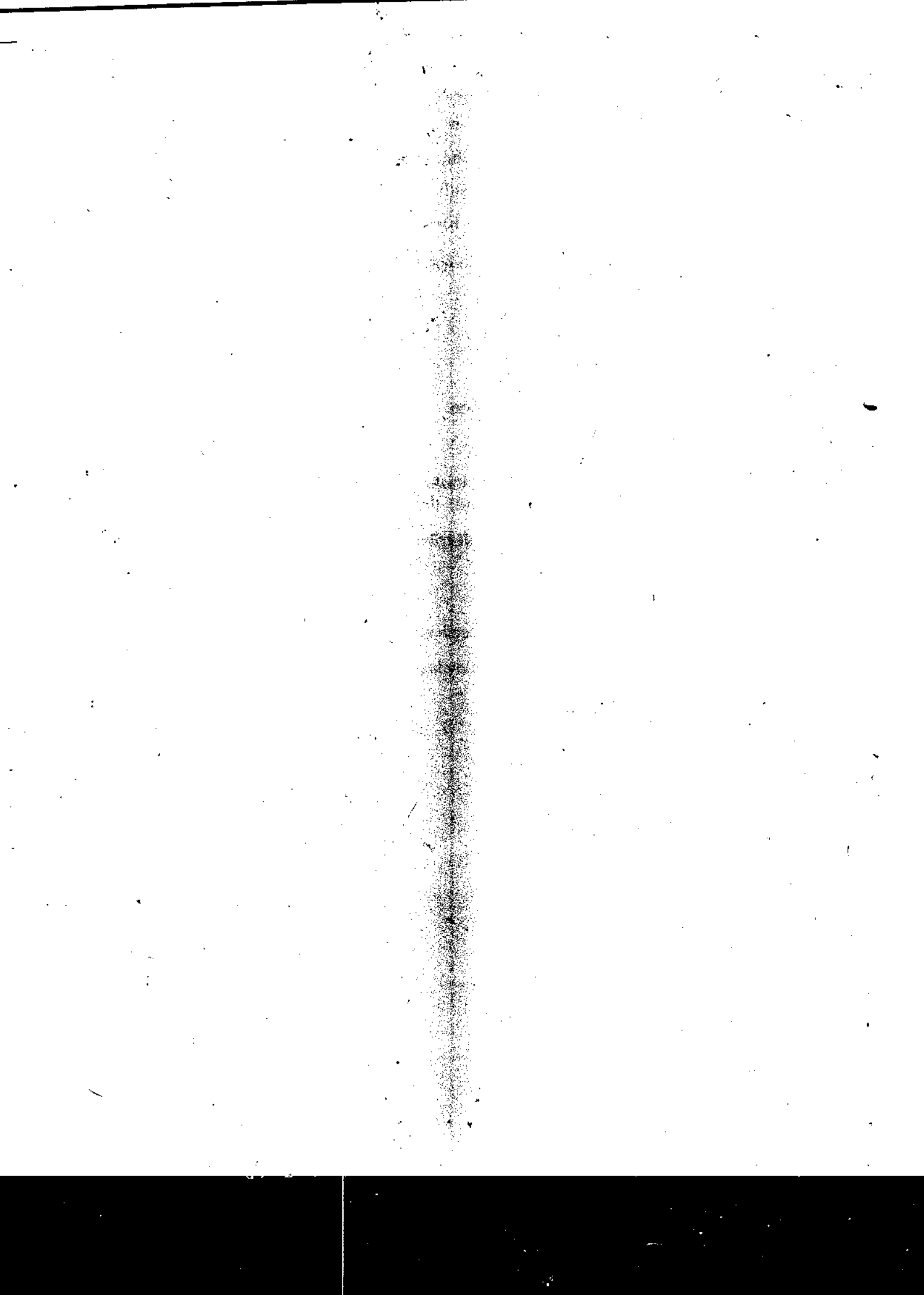
(c) Illutarte reference model for distributed transaction recovery control with the help of neat sketch. 5

6. (a) Discuss use of distributed database in data warehousing. 5

(b) Write short note on the followings :—

(1) Teradata

(2) Gamma. 5



Course Code : CST 409–1

CIOU/RS – 16 / 3489

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 question carry equal marks.
(2) Due credit will be given to neatness and adequate dimensions.
(3) Assume suitable data wherever necessary.
(4) Illustrate your answers wherever necessary with the help of neat sketches.
(5) Retain the construction lines.
(6) Mobile phones are prohibited in examination hall.
(7) Carefully see the internal choices.

1. (a) What are various applications of web intelligence ? How keyword based index works ? Give an example.

OR

(b) Explain how ranking is used to sort the results obtained by an index search in search engines. 4

(c) What is Locality Sensitive Hashing (LSH) ? Evaluate the S–curve $1-(1-s^r)^b$ for $s = 0.1, 0.4$, and 0.6 for the following values of r and b.

(1) $r = 3$, and $b = 10$;

(2) $r = 5$, and $b = 50$;

(3) $r = 4$, and $b = 40$;

(4) $r = 5$, and $b = 100$;

6

2. (a) Suppose there is a repository of one million documents, and word w appears in 1000 of them. In a particular document d, the maximum number of occurrences of a word is 30. Approximately what is the TF.IDF score for w if that word appears (a) 10 times (b) fifty times? 4

(b) Considering following table of statistics of tweets collected over the period; find the sentiment of the tweet : simple courses are kind of boring. Make suitable assumptions.

Count		Sentiment
3000	I really like this course and am learning a lot	positive
2000	I really hate this course and think it is waste of time.	negative
200	The course is really too simple and quite a bore.	negative
4000	The course is simple, fun and very easy to follow.	positive
1000	I'm enjoying this course a lot and learning something too.	positive
1000	I would enjoy myself a lot if I did not have to be in this course	negative
800	I did not enjoy this course enough	negative

4

(c) Define TF, IDF, and TF-IDF. 2

OR

3. (a) Suppose that we use 1024 functions chosen randomly from set of functions F. Next, we shall construct a new family of functions F_1 by performing a 1024-way OR on F. Then.

(1) What is the probability that F_1 will put Fingerprints from the same finger together in at least one bucket ?

(2) What is the probability that two fingerprints from different fingers will be placed in the same bucket ?

(3) Calculate % false negatives and % false positives.

(Use k=3 and p=0.2, q=0.8) 5

(b) What is sparse-distributed memory ? Explain how can it be used for approximate recall ? 5

4. (a) Elaborate on the row oriented and column oriented databases. Describe OLAP and type of database required for Big Data. 5

(b) Explain the process model of map–reduce. Why do you think that this framework will be robust and not fail where sequential systems will fail if used to solve the same problem ? 5

OR

5 (a) For the set of 8 documents show the map–reduce process to find word count of words. Use three mappers and two reducers.

D1 : The game of life is a game of everlasting learning.

D2 : The unexamined life is not worth living.

D3 : Never stop learning.

D4 : Some papers are unexamined.

D5 : Life is a game.

D6 : Life is a full of surprises

D7 : Life is worth living.

D8 : Surprises are part of life. 6

(b) What is Dremel ? Why is it considered important ? 4

6. (a) How to learn facts of the form <subject, verb, object> from text ? Explain with example. 5

(b) Write a note on Mongo DB. Comment why it is popular over other No SQL databases. 5

OR

7. (a) Explain with an example how Bayesian networks are used for drawing inference. 4

(b) What is proposition and predicate logic ? 3

(c) What are fundamental limits of logic ? Also comment on causality. 3

8. (a) What is a long tail phenomenon ? Explain the problems associated with it and how are they handled. 6

(b) What is unsupervised learning ? How classes are emerged ? 4

OR

9. (a) Explain Bayesian network. The prior probability of S being yes is 0.3 and for no 0.7. The prior probability of R being yes is 0.2 and for no is 0.8. Conditional probabilities are given below :

W	S	R	P
Y	Y	Y	0.9
Y	Y	N	0.7
Y	N	Y	0.8
Y	N	N	0.1
N	N	N	0.9
N	N	Y	0.2
N	Y	N	0.3
N	Y	Y	0.1

. Given Evidence 1:Grass is wet, W=Y, find the probability of rain.

. Given evidence W=Y and S=Y, find probability of rain. 8

10. (a) Learning has been an old phenomenon. After adapting it in machine learning what various methods of learning identified and how are they implemented ? 10

OR

11. (a) Elaborating on the significance explain blackboard architecture. State advantaages of this architecture. 5

(b) Illustrate sparse distributed representations and sequence learning. 5