Ref. No.: Ex/PG/SWE/T/111A/2024

MASTER OF SOFTWARE ENGINEERING EXAMINASTION, 2024

1st year, 1st semester

INFORMATION SYSTEMS ENGINEERING

Time: 3 Hours Marks: 100

Note: Answer any **10 questions**.

1. a) What is the difference between Data Science and Machine Learning?
   b) Explain the structure of a system?
   c) What are the benefits of model?
   d) How Innovation and Creativity can be implemented in any DSS?

(4x 2.5)

2. a) What is a sequence data type? Give one example.
   b) How can we define and access a class attribute in Python, discuss with example.
   c) Write a Python program to get the largest number from a list.

(3+3+4)

3. a) Describe method overriding in Python with example.
   b) With one example describe the use of python constructor and destructor.
   c) Explain with one example, how we can use static method in Python.

(3+4+3)

4. a) Compare and contrast data-centric versus model-centric DSS.
   b) What are structured, unstructured, and semi structured problems? Give examples.
   c) Why do we need uncertainty? Explain with example.

(3+4+3)

5. a) Discuss about the steps of Process Modeling.
   b) Suppose, there is a lawyer who offers legal advice to his customers. The service works as follows: The customers can ask for legal advice whenever they need it. The lawyer provides the requested advice and puts the billable hours on the customer's time sheet. When the month is over, the lawyer's accountant determines the billable hours based on the time sheet and creates the invoice. Now applying Business Process Model and Notation (BPMN) try to represent the system graphically.

(3+7)

6. a) Write Short Notes on Basic risky decision.
   b) There are some statistics about the risk of kidney problems in the future. A person will have a high risk of kidney problems if the person is obese from childhood. If the person is not obese from childhood but he/she has a family history of hypertension, then the risk of kidney problem will be moderate. If the person does not have any family history of hypertension but does not eat early at night, the risk of kidney problems is low. Otherwise, the chances of kidney problems are nil. With the help of a decision tree show the scenario.

(2+8)

[Turn over

7. a) What are schemaless databases?
   b) Eventual consistency makes no safety guarantees; eventual consistency only guarantees liveness property – Explain.
   c) Explain the concept of sharding with example.
   d) Give one example to show that the number of mappers and reducers have an effect on performance of map reduce solutions.

(4x2.5)

8. Consider the following MongoDB documents in the collection named "Paper":

{"PID":"362", "Author1":" Andrew", "Author2":" James", "Title":" Face classification", "Subject":"Image Processing", "DOI":" 12.32.4", "Cites": "132"}

{"PID":"363", "Author1":" Peter", "Author2":" James", "Title":" Face recognition", "Subject":" Image Processing", "DOI":" 12.32.5", "Cites": "73"}

{"PID":"469", "Author1":" Black", "Author2":" Andrew", "Title": "Object Recognition", "Subject":" ML", "DOI":" 12.32.16", "Cites": "25"}

{"PID":"656", "Author1":" James", "Author2":"George", "Title":" Feature Generation", "Subject":" ML", "DOI":" 12.32.14", "Cites": "58"}

{"PID":"668", "Author1":" Peter", "Author2":"Black", "Title":" Classification of objects", "Subject":" Neural Network", "DOI":" 12.326.6", "Cites": "221"}

Answer the following queries:

   a) Write a query to display the details of the Papers belonging to either ML or Image Processing category.
   b) Write a query to display the details of the Papers where Author name is Andrew and title Object Recognition.
   c) Write a query to add a key value pair of "Like:5" to all the documents where Cite is greater than 50.
   d) Write a command to make PID, the index.
   e) Write the queries that will return the details of the Papers where cites are greater than or equal to 100.

(5x2)

9. a) There is a set of records. Each record has field F and arbitrary number of category labels G = {G1, G2, ...} . Provide a map-reduce solution to count the total number of unique values of field F for each subset of records for each value of any label. Provide pseudo code for the map and reduce functions along with necessary explanations.

Example input:

F=F1, G={a, b}
F=F2, G={a, d, e}
F=F1, G={b}
F=F3, G={a, b}

Expected output: (the commented part is for explanation; it's not part of the output)

a → 3 // F= {F1, F2, F3}
b → 2 // F={F1, F3}
d → 1 // F={F2}
e → 1 // F={F2}

b) Consider the following data:

UserId	TrackId	Shared	listen	Skip
111123	122	0	1	0
111124	126	1	0	0
111127	123	0	1	1
111123	125	1	0	0

Provide a map reduce solution to find the count of listen, share and skip for each tracks. Take advantage of partitioning. Provide pseudo code for the map and reduce functions along with necessary explanations.

(5+5)

10. a) What is HDFS?
    b) Name the services running in master part of hadoop architecture.
    c) What is the output of a mapper task?
    d) What are the key value pairs in case of TextInput Format?
    e) What happens when a Datanode is down?

(5x2)

11. a) Describe the crash recovery mechanism used in Cassandra.
    b) With reference to Cassandra's tuneable consistency feature, describe strict and casual consistencies.

(5+5)