

VU Learning HUB

BIF601 Quiz File

Chapter-wise Solved MCQs

vulearninghub.xyz

Lesson 1: Introduction to Database Management Systems (DBMS)

Q1. What is the primary role of a Database Management System (DBMS)?

- A) A physical hardware device that stores raw network packets
- B) System software for creating, protecting, reading, updating, and deleting data in a database
- C) An operating system kernel designed specifically for router hardware
- D) A web browser used to display HTML tables

Q2. Which foundational element of a DBMS defines the logical structure of a database?

- A) Storage engine
- B) Database engine
- C) Database schema
- D) Lock manager

Q3. According to the handout, which specific DBMS component is responsible for managing data locking?

- A) Metadata catalog
- B) Optimization engine
- C) Lock manager
- D) Data utilities

Q4. What serves as an interface between databases and end users or application programs?

- A) The physical layer
- B) The DBMS
- C) The router
- D) The command line shell

Answer Key

Q1: B — A DBMS is system software that enables end users to create, protect, read, update, and delete database data

Q2: C — The database schema defines the logical structure of the database

Q3: C — The lock manager is the integrated component responsible for handling database data locks

Q4: B — The DBMS essentially serves as the interface between databases and end users or applications

Lesson 2: Advantages of Database Management Systems

Q1. Data inconsistency exists when:

- A) Multiple users issue ad hoc queries simultaneously
- B) Different versions of the same data appear in different places
- C) Database tables are stored in cloud servers
- D) Data encryption policies are enforced

Q2. An "ad hoc query" in a database context is defined as a:

- A) Scheduled weekly automated backup
- B) Spur-of-the-moment question issued for data manipulation
- C) Permanent structural change to a database schema
- D) Physical connection request between network hubs

Q3. Which benefit of a DBMS provides a framework for better enforcement of data privacy and security policies?

- A) Improved data security
- B) Data sequencing
- C) Packet switching
- D) Hardware abstraction

Q4. Wider access to well-managed corporate data primarily promotes:

- A) Better data integration
- B) Increased hardware overhead
- C) Slower query response times
- D) Data duplication

Answer Key

Q1: B — Data inconsistency occurs when different versions of the same data appear in different locations

Q2: B — An ad hoc query is defined as a spur-of-the-moment request for data manipulation

Q3: A — A DBMS provides a structured framework for enforcing data privacy and security policies

Q4: A — Broad access to well-managed data promotes an integrated view of an organization's operations

Lesson 3: Database Software & User Roles

Q1. Which language commands are typically input by administrators to prompt a database system to perform actions?

- A) HTML
- B) Structured Query Language (SQL)
- C) Assembly Language
- D) Perl

Q2. Which interface component provides database users with a centralized view of data in tabular forms?

- A) Command Line Interface (CLI)
- B) Graphical User Interface (GUI)
- C) UNIX Shell
- D) BIOS Menu

Q3. Which of the following is explicitly listed as one of the primary user roles in a DBMS environment?

- A) Network Repeaters
- B) Administrators
- C) Optical Transceivers
- D) Physical Layer Controllers

Q4. Database software is used to implement cybersecurity measures to protect against:

- A) Hardware overheating
- B) Malware, viruses, and security threats
- C) Data sequencing errors
- D) Circuit switching delays

Answer Key

Q1: B — Administrators use Structured Query Language (SQL) commands to prompt the system to perform database actions

Q2: B — Most database software features a GUI with structured fields and tabular forms for centralized data management

Q3: B — Administrators (along with programmers, end users, and applications) are defined as primary user roles

Q4: B — Database software implements measures to defend databases against malware, viruses, and cybersecurity threats

Lesson 4: Data Management Techniques

Q1. What is the primary purpose of Data Preparation in data management?

- A) To connect local computers over a optical fiber network
- B) To clean and transform raw data into the right shape and format for analysis
- C) To route packets across autonomous networks
- D) To execute low-level machine instructions

Q2. What does the acronym ETL stand for in data pipeline workflows?

- A) Enterprise Transmission Layer
- B) Extract, Transform, Load
- C) Electronic Telecommunication Link
- D) Encapsulated Token Logic

Q3. Which data management technique documents the flow of data through an application or organization?

- A) Data Modeling
- B) Data Pipeline
- C) Data Sequencing
- D) Data Routing

Q4. Automated transfer of data from one system to another is enabled by:

- A) Data Pipelines
- B) Ring Topologies
- C) Manual SQL scripts
- D) Circuit Switching

Answer Key

Q1: B — Data preparation cleanses and transforms raw data into the required format for business analysis

Q2: B — ETL stands for Extract, Transform, Load, which extracts data from a system, transforms it, and loads it into a data warehouse

Q3: A — Data modeling documents how data flows across an application or organizational system

Q4: A — Data pipelines facilitate the automated transfer of data between computing systems

Lesson 5: Computer Networks & Network Categories

Q1. Which type of network is centered around an individual's personal workspace?

- A) Wide Area Network (WAN)
- B) Personal Area Network (PAN)

- C) Metropolitan Area Network (MAN)
- D) Backbone Network (BBN)

Q2. A Wide Area Network (WAN) typically covers a geographical area with a radius greater than:

- A) 1 meter
- B) 10 meters
- C) 1 kilometer
- D) 100 kilometers

Q3. A network connecting devices within a single limited location like an office building or school is a:

- A) Local Area Network (LAN)
- B) Global Area Network (GAN)
- C) Metropolitan Area Network (MAN)
- D) Enterprise Private Network

Q4. Which network category spans globally to connect networks worldwide, with the internet as a prime example?

- A) Home Area Network (HAN)
- B) Global Area Network (GAN)
- C) Campus Network
- D) Local Area Network (LAN)

Answer Key

- Q1: B — A Personal Area Network (PAN) is based on an individual's immediate workspace
- Q2: C — A WAN covers a larger geographical area, usually with a radius exceeding 1 kilometer
- Q3: A — A LAN links computers within a limited restricted geographical area like a school or office
- Q4: B — A Global Area Network (GAN) is a worldwide network connecting systems across the globe

Lesson 6: Data Communication Components & Protocols

Q1. What is defined as a set of rules that governs communication between transmitting devices?

- A) Topology
- B) Protocol
- C) Schema
- D) Medium

Q2. The protocol function of breaking a long message into smaller fixed-size packets and numbering them is called:

- A) Data Routing
- B) Data Sequencing
- C) Flow Control
- D) Data Modeling

Q3. Which of the following is NOT one of the 5 basic components of a data communication system?

- A) Message

- B) Transmission Medium
- C) Compiler
- D) Sender

Q4. In data communication error control, what happens when a received message block contains an error?

- A) The whole network reboots automatically
- B) The erroneous block is discarded by the receiver and retransmitted by the sender
- C) The message is converted to assembly code
- D) Transmission speed is doubled

Answer Key

Q1: B — A protocol is a formal set of rules governing communication between devices

Q2: B — Data sequencing breaks long messages into numbered, fixed-size packets

Q3: C — The 5 core components are Message, Sender, Receiver, Medium, and Protocol (a compiler is not a component)

Q4: B — When an error is detected in error control, the receiver discards the block and the sender retransmits it

Lesson 7: Communication Protocols & Properties

Q1. Which popular protocol is used specifically for file transfers across networks?

- A) HTTP
- B) File Transfer Protocol (FTP)
- C) SMTP
- D) IMAP

Q2. Communication protocols govern which three fundamental aspects of digital messages?

- A) Syntax, Semantics, and Synchronization
- B) Hardware, Software, and Firmware
- C) Extraction, Transformation, and Loading
- D) Rows, Columns, and Tables

Q3. Which protocol is commonly used for sending electronic mail?

- A) Simple Mail Transfer Protocol (SMTP)
- B) UDP
- C) FTP
- D) SQL

Q4. Handshaking, flow control, and address mapping are properties defined by a:

- A) Database Schema
- B) Communication Protocol
- C) Graphical User Interface
- D) Decision Tree

Answer Key

Q1: B — File Transfer Protocol (FTP) is designed specifically for transferring files between hosts

Q2: A — Protocols define the syntax, semantics, and synchronization of communications

Q3: A — SMTP (Simple Mail Transfer Protocol) is used for electronic mail transmission

Q4: B — Properties such as handshaking, flow control, and address mapping are defined by communication protocols

Lesson 8: Network Topologies

Q1. What term refers to the geometric representation of how computers are connected in a network?

- A) Network Architecture
- B) Network Topology
- C) Data Schema
- D) Routing Protocol

Q2. In which network topology are all host devices connected directly to a central Hub?

- A) Ring Topology
- B) Star Topology
- C) Bus Topology
- D) Mesh Topology

Q3. In a Ring Topology, each device has dedicated point-to-point links with exactly how many neighboring devices?

- A) One device
- B) Two devices
- C) All devices in the network
- D) Four devices

Q4. Which network topology utilizes a single central cable with drop lines connecting workstations?

- A) Star Topology
- B) Mesh Topology
- C) Bus Topology
- D) Hybrid Topology

Answer Key

Q1: B — Network topology is the geometric representation of how network devices are linked together

Q2: B — In a Star topology, all devices connect directly to a central hub

Q3: B — Devices in a Ring topology maintain dedicated point-to-point links with only the two adjacent devices on either side

Q4: C — A Bus topology connects host nodes to a single central cable via drop lines

Lesson 9: The OSI & TCP/IP Reference Models

Q1. How many layers are in the OSI (Open Systems Interconnection) reference model?

- A) 4
- B) 5
- C) 7
- D) 9

Q2. Which OSI layer handles data translation, encryption, decryption, and data compression?

- A) Physical Layer
- B) Transport Layer
- C) Presentation Layer
- D) Network Layer

Q3. Logical addressing (IP addresses) and packet routing between networks occur at which OSI layer?

- A) Data Link Layer (Layer 2)
- B) Network Layer (Layer 3)
- C) Session Layer (Layer 5)
- D) Physical Layer (Layer 1)

Q4. How many layers comprise the TCP/IP protocol suite model?

- A) 3
- B) 4
- C) 7
- D) 10

Answer Key

Q1: C — The OSI reference model consists of 7 conceptual layers

Q2: C — The Presentation Layer (Layer 6) handles translation, encryption/decryption, and compression

Q3: B — Logical IP addressing and routing functions take place at the Network Layer (Layer 3)

Q4: B — The TCP/IP model consists of 4 layers: Network Access, Internet, Transport, and Application

Lesson 10: Information Retrieval (IR) & Evaluation

Q1. In Information Retrieval, "Precision" is mathematically calculated as:

- A) Total relevant documents retrieved / Total documents retrieved
- B) Total documents retrieved / Total relevant documents in database
- C) Non-relevant items retrieved / Total items
- D) Total queries / Total execution time

Q2. "Recall" in an Information Retrieval evaluation measures:

- A) The proportion of relevant documents in the database that were successfully retrieved
- B) The speed of query execution
- C) The percentage of non-relevant items in the search results
- D) The size of the database index

Q3. Standard Boolean, Vector-space, and Probabilistic models belong to which category of IR models?

- A) Non-Classical IR Models
- B) Classical IR Models
- C) Alternative IR Models
- D) Neural IR Models

Q4. In Information Retrieval evaluation, "Fallout" is defined as the:

- A) Proportion of non-relevant items retrieved in a given search
- B) Proportion of relevant items retrieved in a search
- C) Server latency time

D) Database error rate

Answer Key

Q1: A — Precision is the ratio of relevant retrieved documents to the total number of documents retrieved

Q2: A — Recall is the proportion of relevant documents in the database that were retrieved

Q3: B — Vector-space, Boolean, and Probabilistic models are classic examples of Classical IR models

Q4: A — Fallout represents the proportion of non-relevant items retrieved during a query search

Lesson 11: Information Retrieval in Biological Databases

Q1. What is the most widely used interface for retrieving information from public biological databases?

- A) NCBI Entrez
- B) Apache Web Server
- C) UNIX Shell
- D) MS Access

Q2. Which bibliographic database provides references and abstracts on biomedical topics accessed via PubMed?

- A) GenBank
- B) MEDLINE
- C) SWISS-PROT
- D) PDB

Q3. SRS (Sequence Retrieval System) was developed to integrate heterogeneous databanks by which entities?

- A) NASA and CERN
- B) EBI and DDBJ
- C) Google and Yahoo
- D) IEEE and ACM

Q4. Which specialized search tool extracts Gene Ontology terms from PubMed abstracts to assist browsing?

- A) GpPubMed
- B) RefMed
- C) CiteSeer
- D) BLAST

Answer Key

Q1: A — NCBI Entrez is the most widely used interface for biological data retrieval

Q2: B — MEDLINE is the bibliographic database of life sciences and biomedical literature

Q3: B — SRS is a data retrieval system that integrates heterogeneous molecular databanks at EBI and DDBJ

Q4: A — GpPubMed extracts Gene Ontology terms from PubMed abstracts for ontological browsing

Lesson 12: Knowledge Discovery & Data Mining

Q1. In a Decision Tree model, what does a "leaf node" represent?

- A) A test on a single attribute

- B) A specific classification or outcome
- C) The root input dataset
- D) An association rule

Q2. A Naive Bayes classifier calculates which probability using prior probabilities and training likelihoods?

- A) Prior Odds Ratio
- B) Posterior Probability
- C) Linear Regression Score
- D) Euclidean Distance

Q3. Decision tree construction algorithms typically build trees in which structural manner?

- A) Top-down recursive divide-and-conquer
- B) Bottom-up random walk
- C) Brute force exhaustive search
- D) Matrix inversion

Q4. Knowledge Discovery in Databases (KDD) developed out of which domain?

- A) Data Mining
- B) Physical Layer Wiring
- C) Optical Switching
- D) Hardware Assembly

Answer Key

Q1: B — A leaf node in a decision tree represents a specific classification outcome

Q2: B — Naive Bayes classifiers assess hypothesis likelihood to yield posterior probability

Q3: A — Basic decision tree algorithms construct trees in a top-down recursive divide-and-conquer manner

Q4: A — Knowledge discovery developed out of the data mining domain

Lesson 13: Machine Learning Approaches & Clustering

Q1. Supervised machine learning algorithms are defined by their use of:

- A) Unlabeled datasets
- B) Labeled datasets
- C) Unconnected physical nodes
- D) Manual SQL queries

Q2. K-means clustering is a prominent example of which category of machine learning?

- A) Supervised Learning
- B) Unsupervised Learning
- C) Semi-Supervised Logic
- D) Assembly Programming

Q3. In unsupervised learning, discovering relationships between variables (e.g., market basket analysis) is known as:

- A) Regression
- B) Association
- C) Classification
- D) Bit Synchronization

Q4. What type of machine learning involves an agent learning through trial and error within an environment?

- A) Reinforcement Learning
- B) Supervised Classification
- C) Decision Tree Induction
- D) Static Parsing

Answer Key

Q1: B — Supervised learning is defined by training algorithms using labeled datasets

Q2: B — K-means clustering is an unsupervised learning technique used for grouping unlabeled data

Q3: B — Association rules are unsupervised learning methods used to find variable relationships in datasets

Q4: A — In reinforcement learning, an agent learns through trial and error by taking actions in an environment

Lesson 14: String Matching & Pattern Finding

Q1. What is a key advantage of the Knuth-Morris-Pratt (KMP) string matching algorithm over Brute Force?

- A) It requires zero memory
- B) It operates in linear time and avoids backtracking on the text string
- C) It processes binary numbers only
- D) It generates phylogenetic trees automatically

Q2. In the KMP algorithm, what does the Prefix Function encapsulate?

- A) Knowledge about how the pattern matches against shifts of itself
- B) The translation of RNA to protein
- C) The physical IP address of nodes
- D) Distance matrices between sequences

Q3. A brute force string searching algorithm finds matches by:

- A) Pre-computing prefix shifts
- B) Trying out all possible shifts and positions
- C) Translating text into bytecode
- D) Clustering similar patterns into dendrograms

Q4. In biological pattern searching, pattern and text symbols are chosen from a finite set called the:

- A) Schema
- B) Alphabet
- C) Topology
- D) Vector Space

Answer Key

Q1: B — KMP is a linear time algorithm that eliminates repetitive comparison/backtracking on the text string

Q2: A — The prefix function encapsulates knowledge of pattern matches against shifts of itself to avoid useless shifts

Q3: B — The brute force algorithm tests all possible shifts to find pattern occurrences

Q4: B — Symbols for genome pattern searching are selected from a predetermined finite set called an alphabet

Lesson 15: Sequence Alignment & Substitution Matrices

Q1. In Dot Plot sequence analysis, what does a jump or shift in the main diagonal indicate?

- A) Perfect sequence match
- B) An Insertion or Deletion (InDel)
- C) High execution speed
- D) Termination of translation

Q2. Which substitution matrix is derived from the BLOCKS database of aligned, ungapped protein regions?

- A) PAM
- B) BLOSUM
- C) Needleman-Wunsch
- D) Smith-Waterman

Q3. BLOSUM62 matrix is particularly recommended for which type of alignments?

- A) Gapped alignments
- B) Ungapped alignments
- C) Global RNA transcription
- D) Phylogenetic tree plotting

Q4. What does the acronym "PAM" stand for in sequence alignment scoring matrices?

- A) Protein Alignment Method
- B) Point Accepted Mutations
- C) Paired Amino Matrix
- D) Pattern Access Model

Answer Key

Q1: B — A jump or shift in a dot plot's main diagonal points out an insertion or deletion (indel)

Q2: B — BLOSUM matrices are derived from aligned ungapped regions in the BLOCKS database

Q3: B — BLOSUM62 is optimal for ungapped alignments, while BLOSUM50 is suited for gapped alignments

Q4: B — PAM stands for Point Accepted Mutations, introduced by Dayhoff et al

Lesson 16: Dynamic Programming Algorithms

Q1. Which algorithm was developed in 1970 to perform global sequence alignment using dynamic programming?

- A) Smith-Waterman Algorithm
- B) Needleman-Wunsch Algorithm
- C) KMP Algorithm
- D) Naive Bayes Classifier

Q2. The Smith-Waterman algorithm is widely used in bioinformatics for:

- A) Global alignment of whole genomes
- B) Local sequence alignment and sequence similarity determination
- C) Compiling C programs into assembly

D) Managing database locks

Q3. What is the distinguishing feature of an "Affine Gap Penalty"?

- A) It penalizes every gap position identically regardless of length
- B) It treats gap opening penalty and gap extension penalty separately
- C) It assigns zero cost to all gap insertions
- D) It operates only on DNA sequences

Q4. What algorithmic technique is used by Needleman-Wunsch to find optimal alignment scores?

- A) Dynamic Programming
- B) Brute Force Enumeration
- C) Heuristic Tree Search
- D) Neural Network Backpropagation

Answer Key

Q1: B — The Needleman-Wunsch algorithm was published in 1970 to perform global sequence alignment

Q2: B — Smith-Waterman is used for local sequence alignment and sequence similarity determination

Q3: B — Affine gap penalties separate the gap opening penalty (d) and gap extension penalty (e)

Q4: A — Dynamic programming is used to determine optimal sequence alignments and scores

Lesson 17: Programming Languages & BioJava

Q1. Which programming language level consists of primitive binary instructions built into computer hardware?

- A) High-Level Language
- B) Machine Language
- C) Assembly Language
- D) Scripting Language

Q2. BioJava was initiated in 1998 at EBI/Sanger by which developers?

- A) Larry Wall and Guido van Rossum
- B) Matthew Pocock and Thomas Down
- C) Saul Needleman and Christian Wunsch
- D) James Gosling and Dennis Ritchie

Q3. What build tool is required for compiling BioJava?

- A) ANT
- B) Maven
- C) CMake
- D) Make

Q4. In Java programming, code enclosed between curly braces { } forms a:

- A) Statement
- B) Block
- C) Comment
- D) Reserved Word

Answer Key

Q1: B — Machine language is the set of primitive binary instructions built directly into computers

Q2: B — BioJava was started at EBI/Sanger in 1998 by Matthew Pocock and Thomas Down

Q3: A — Compiling BioJava requires the ANT build tool

Q4: B — A pair of curly braces in a Java program forms a block grouping code components

Lesson 18: MATLAB & Bioinformatics Toolbox

Q1. What does the acronym "MATLAB" stand for?

- A) Mathematical Analysis Language and Base
- B) Matrix Laboratory
- C) Method for Algorithmic Testing and Logic
- D) Molecular Alignment Toolbox

Q2. In MATLAB, how are basic scalars and vectors represented?

- A) As single-precision strings
- B) As matrices of dimensions 1x1 or 1xn / nx1
- C) As linked lists
- D) As binary hash tables

Q3. Which MATLAB Bioinformatics Toolbox function performs global sequence alignment via Needleman-Wunsch?

- A) nwalgn
- B) swalign
- C) fastaread
- D) seqdotplot

Q4. Which MATLAB function reads sequence data stored in FASTA file format?

- A) genbankread
- B) fastaread
- C) pdbread
- D) multialignread

Answer Key

Q1: B — MATLAB stands for Matrix Laboratory

Q2: B — All variables in MATLAB are treated as matrices, including scalars (1x1) and vectors (1xn)

Q3: A — nwalgn is the MATLAB function for Needleman-Wunsch global alignment

Q4: B — fastaread is used to read FASTA formatted sequence files in MATLAB

Lesson 19: Biopython & PHP Web Scripting

Q1. In Biopython, which object provides parse, read, write, and convert functions for sequence file I/O?

- A) SeqIO
- B) Phylo
- C) Entrez
- D) BioSQL

Q2. Which built-in method of a Biopython Seq object converts a DNA sequence into an RNA sequence?

- A) translate()
- B) transcribe()
- C) reverse_complement()
- D) align()

Q3. PHP is best described as a:

- A) Client-side compiled binary language
- B) Server-side scripting language executed on the server
- C) Machine-level assembly language
- D) Operating system kernel extension

Q4. Biopython was originally developed primarily by:

- A) Chapman and Chang
- B) Needleman and Wunsch
- C) Pocock and Down
- D) Smith and Waterman

Answer Key

Q1: A — SeqIO provides the four core sequence file functions: parse, read, write, and convert

Q2: B — The transcribe() method transcribes a DNA Seq object into an RNA sequence

Q3: B — PHP is a server-side scripting language executed on the web server

Q4: A — Biopython was developed by Chapman and Chang

Lesson 20: Perl, UNIX Commands, & CLUSTALW

Q1. What does the acronym "Perl" stand for?

- A) Practical Extraction and Report Language
- B) Program for Elementary Record Linking
- C) Process Execution and Resource Logic
- D) Protocol for Enterprise Routing Layer

Q2. Which UNIX command is used to list the contents of a directory?

- A) cd
- B) ls
- C) rm
- D) cp

Q3. Multiple Sequence Alignment (MSA) differs from Pairwise Alignment because MSA compares:

- A) Exactly two sequences
- B) Three or more sequences simultaneously
- C) Binary machine code strings only
- D) Only 3D protein coordinate structures

Q4. CLUSTALW is a widely used online tool developed by EMBL/EBI for performing:

- A) Single-user database locking
- B) Progressive Multiple Sequence Alignment

- C) Assembly language compilation
- D) Circuit switching routing

Answer Key

Q1: A — Perl stands for Practical Extraction and Report Language

Q2: B — The ls command lists directory contents in UNIX

Q3: B — MSA involves the alignment and comparison of three or more biological sequences

Q4: B — CLUSTALW is an online tool developed by EMBL/EBI for progressive multiple sequence alignment

End of Quiz

BIF601 — 80 solved MCQs
Prepared by VU Learning HUB
vulearninghub.xyz