Final Term Assignment

Instructions:

- **For this Assignment search all resources accessible to you and utilize accordingly**
- **Use pen and paper to complete this Assignment. Make sure to write your Name, ID and Signature on the paper. If you require multiple papers to complete this work write your Name, ID and Signature on all of them.**
- **Click photos of your completed work from the papers and convert to pdf (you can use any tool/app of your choice).**
- **Submit pdf in link provided in your VUES Student Account.**
- **Name of pdf must be your ID.**
- **Submission Deadline: 27th June 2025, 11:59pm.**

Question:

Explore the following topics and explain in your own words.

- Information Retrieval
- Cloud Database
- Graph Database
- Distributed Database
- Bioinformatics

Name: FARJANA YESMIN OPI ID: 22-47018-1 Sign: Opi

① Information Retrival: Information Retrival is a science and technology that is used for searching relevant information within large collections of data especially unstructured text. The goal of IR is to help users find documents, images, videos or other content that mathes their queries.

How it works: When we type keywords in a search engine like Google or Bing, the system looks through billions of documents and ranks them by relevance using algorithms. It uses techniques like keyword matching, natural language proccessing and machine learning to improve results.

Example: Imagine a library with millions of book but no catalog. IR system acts like a digital catalog and search engine. If we search for "climate change impact on agriculture" the IR systems finds reasearch papers, news articles, and reports related to that topic.

② Cloud Database: A cloud database is a database service provided and managed through cloud computing platforms. Unlike traditional databases run on remote servers maintained by cloud providers. This setup allows easy scalability, remote access and reduces the need of physical hardware.

Benefits: We do not need to buy expensive servers or worry about maintainance. We can scale storage or computing power up and down depending on needs. Accessible from anywhere with internet access.

Example: Amazon's Amazon RDS or Google's Cloud SQL are cloud databases. A company running a online store may use a cloud database to customer data and orders. When sales increase during holidays, the database can automatically expand

to handle more traffic without downtime.

③ Graph Database: A graph database stores data as nodes (entities) and edges (relationships). It is designed to handle highly connected data and complex relationships efficiently. Unlike traditional relational databases that uses table, graph databases exel at exploring connections.

Use Cases: (i) Social Networks (Facebook, LinkedIn) ⇒ Users and their friendships or professional connections.

(ii) Recomandation systems (Netflix, Amazon) ⇒ Item related by user preferences or purchases.

(iii) Fraud Detection ⇒ Finding suspicious pattern by analyzing connections between transactions or accounts.

Example: In a social media app, a graph database can quickly find "friends of friends" or suggest people you might know based on mutual connections which would be more complex and slower ~~than~~ in a traditional database.

④ Distributed Database: A Distributed Database is a single logical database that is physically spread across multiple locations or servers, possibly across different cities or countries. These locations work together to provide data access as if it were a single database.

Advantages:

(i) Increased reliability: If one site fails, others can continue working.

(ii) Faster access: Users get data from the nearest server, reducing delay.

(iii) Scalability: We can add more servers as data grows.

NAME: FARJANA YESMIN OPI ID: 22-47018-1 Sign: Opi

Example: A global e-commerce company like Amazon uses distributed databases to store user data. Customers in US and Europe access their local servers which sync with other ~~services~~ servers worldwide. This avoids delays and prevents data loss if one servers fails.

⑤ Bioinformatics: Bioinformatics is an interdiciplinary field combining biology, computer science and statistic to analyze and interpret biological data. The rapid growth of biological data such as DNA sequences, protein structures and gene expressions requires computational tools to understand them.

Applications:

Genome sequencing: Decoding DNA to study genes and hereditary diseases.

Drug discovery: Identifying molecules that can become medicines.

Evolutionary studies: Comparing genetic information across species.

Example: The Human Genome Project used bioinformatics to sequence the entire human DNA, producing vast amounts of data that needed software tools to analyze genes linked to diseases like cancer or diabetes.