Part A: Assignment 03:

Question 1

1.1   Database Lifecycle main phases are:
- Database initial study and it <u>actions</u> are analysing the company situation, define problems and constraints, define objectives, define scope and boundaries.
- Database Design and it <u>actions</u> are create the conceptual design DBMS software selection, create the logical design, create the physical design.
- Implementation and loading and it actions are Installing the DBMS, creating the databases and loading or converting the data.
- Testing and Evaluation and it <u>actions</u> are testing the database, fine-tune the database, evaluating the database and its application programs**.**
- Operation and Maintenance and it <u>actions</u> are producing the required information flow.
- Evolution and it <u>actions</u> will be to introduce changes and making enhancements.

1.2 ACID Test

ACID means Atomicity, consistency, isolation, and durability. Database transaction management uses these transactional properties or set requirements that all single transactions that are processed in a DBMS must comply to when completing a transaction. ACID also applies serializability technique that all scheduled transactions executed must yield consistent completed result. The DBMS manages all the transactions using a concurrency control technique to achieve a desirable result, otherwise, the incomplete transaction in the database must be rolled back to its previous or initial state before execution.

1.3 Relationship between DBMS Performance Tunning and Query Optimisation.

DBMS resides on the server-side which applies a technique of performance tuning that ensures that end-user queries are processed and returned with the correct results within a minimum amount of time. However, the response time is dependent on several factors or computer resources like the CPU, Memory,

internet bandwidth, network, hard disk space, server-side environment, client applications and so on.

On the other hand, Query Optimisation is an SQL query request technique on the client-side that enables programmers and end-users to optimize queries in a way that provides the DBMS with efficient sequential way of executing set of queries to allow the DBMS to find the best possible efficient means of processing client requests.

They are both database concepts that is designed based on a client-server architecture, and both serve similar objective by finding the shortest best lowest cost of optimizing queries.

Question 2

2.1 The main problems are:

- *Performance degradation* because of a growing number of remote locations over greater distances.
- *High costs* associated with maintaining and operating large central (mainframe) database systems and physical infrastructure.
- *Reliability problems* created by dependence on a central site (single point of failure syndrome) and the need for data replication.
- *Scalability problems* associated with the physical limits imposed by a single location, such as physical space, temperature conditioning, and power consumption.
- *Organizational rigidity* imposed by the database, which means it might not support the flexibility and agility required by modern global organizations.

2.2 The main levels are:

- Single-site Processing and Site-site Data: The DBMS on a single site processes data on   single host machine where all files with data are stored on the local disk and data processing are handled locally by the host machines.

- Multiple-site Processing and Single-site Data: Data processing workload is distributed amongst different computers connected by a network file server application in various geographical locations while sharing a single data repository. Data processor and transaction processors at multiples sites helps to reduce the workload that one central computer for processing.

2.3 BI - Business Intelligence components are:

- **Extraction Transformation and Loading Tools:** This tool allows BI to collect, filter, integrate data for data storage and used for decision making in an organisation.
- **Data Store:** Data warehouses and data mart store data that be used to make decision to support data analysis and businesses for decision making.
- **QUERY and Reporting:** This Data component allows analysts to create queries that access databases and generate reports for stakeholders of a company and in turn used for decision making.
- **Data Visualization:** This data component tool presents data to users in a format that is understandable and presentable such as pie chart, graph, dashboard and so on.
- **Data Monitoring and Alerting:** This data component allows businesses to monitor their business activities in real-time and performing actions such as emailing shop manager, starting an application and other actions.
- **Data Analytics:** This component performs data analysis and data-mining tasks using the data in the data store. It assists advises user to build a business data model.

Question 3

3.1 Big Data

Big data simply refers to a set of data that is characterized by 3 V's (Volume, Velocity, Variety) with a combination of various types of data (text, graphics, video, audio) that makes it complex for a traditional SQL relational database management system to handle however No-SQL databases such as mongo dB and other types of No-SQL have been designed to handle such data types.

3.2 The 3'Vs are Volume, Velocity, Variety.

The internet and social media space has experienced a boom in data collected in the last years, organisation store different type of data in huge number of sizes collected from their customers ad this data types contains characteristic of the 3'Vs.

<u>Volume</u> as a characteristic refers to the amount of data stored and this data are stored in different capacities such a s Bit, Byte, Kilobyte, Megabyte and to mention a few.

<u>Velocity</u> refers to the speed at which data collected and processed into information used to make business decisions.

<u>Variety</u> refers to how different unstructured data types are collected, formatted, structured, and stored to be processed into information.

3.3 Key-Value Databases

The key-value databases are databases that adopted the NoSQL concept model for storing data in form of a key-value pair. The key is associated with the value which can be any data type of value, such as text, xml, images, video, audio, wave files etc. it is a database that does not follow the traditional SQL relational database model. The keys must be unique, and all queries are processed by accessing the key using three basic operations which are Get, Store and Delete. Some key-value databases are Dynamo, Riak, Redis, Voldermot.

3.4 The innovative Services are as follows:

- Permit rapid responses to competitive pressures by bringing new services and products to market quickly.
- Increase customer satisfaction through the creation of innovative data services such as mapping data combined with GPS (Global Positioning System) information to provide location-aware services. These applications present end users with information or services located near the users' current location.
- Allow anywhere, anytime data access using mobile smart devices via the Internet.
- Yield fast and effective information dissemination through universal access from across the street or across the globe.

3.5 Data Administered and Database Administrator activities and characteristic are:

- The Database Administrator (DBA) is responsible for managing the corporate database and managing computerized data and other data and DBA activities include supporting end users, defining and enforcing policies, procedures, and standards for the database, data security, privacy, and integrity and providing data backup and recovery services with monitoring distribution and use of the data in the database.
- Data Administered has more managerial tasks than the more technically oriented DBA and the DA must have a mix of technical and managerial skills.

Question 4

4.1

Advantages

- *Low initial cost of entry*. Cloud computing has lower costs of entry when compared with the alternative of building in house.
- *Scalability/elasticity*. It is easy to add and remove resources on demand.
- *Support for mobile computing*. Cloud computing providers support multiple types of mobile computing devices.

- *Ubiquitous access*. Consumers can access the cloud resources from anywhere at any time, if they have Internet access.
- *High reliability and performance*. Cloud providers build solid infrastructures that otherwise are difficult for the average organization to leverage.
- *Fast provisioning*. Resources can be provisioned on demand in a matter of minutes with minimal effort.
- *Managed infrastructure*. Most cloud implementations are managed by dedicated internal or external staff. This allows the organization's IT staff to focus on other areas.

Disadvantages

- *Issues of security, privacy, and compliance*. Trusting sensitive company data to external entities is difficult for most data cautious organizations.
- *Hidden costs of implementation and operation*. It is hard to estimate bandwidth and data migration costs.
- *Data migration is a difficult and lengthy process*. Migrating large amounts of data to and from the cloud infrastructure can be difficult and time-consuming.

- *Complex licensing schemes*. Organizations that implement cloud services are faced with complex licensing schemes and complicated service-level agreements.
- *Loss of ownership and control*. Companies that use cloud services are no longer in complete control of their data. What is the responsibility of the cloud provider if data are breached? Can the vendor use your data without your consent?
- *Organization culture*. End users tend to be resistant to change. Do the savings justify being dependent on a single provider? Will the cloud provider be around in 10 years?
- *Difficult integration with internal IT system*. Configuring the cloud services to integrate transparently with internal authentication and other internal services could be a daunting task.

Question 5

5.1

The fourth industrial revolution (4IR) and its main components is a combination of advanced technologies in artificial intelligence, robotics, internet of things, block chain quantum computing and many other kinds of technology. Modern day businesses are fast becoming indispensable as they sale product and services to their customers. It brings significant boost in manufacturing efficiency and effectiveness by applying innovative advanced techniques using automated applications which provides competitive advantage to companies who embrace it. Fourth Industrial Revolution (4IR) is different from the Third Industrial Revolution (3IR)? Well, everything. It is marked by exponential thinking where linear solutions no longer apply. The digital version replaces the analogue version. Knowledge and invention are cumulative. Evolution is just the re-encoding of information, after all. Every person, culture, industry, and country are affected. All forms of production, management, systems, and governments will be transformed.

The opportunities are unlimited: faster prototyping and time-to-market with 3-D printing and production, conquering disease and illness with nanotechnology, micro-financing using robot-advisors and advanced algorithms, more efficient and affordable connected homes, safer and more convenient travel with autonomous vehicles. Not to mention other improvements made in human longevity, energy, material sciences, entertainment, consumerism—the list goes on and on. All these advances will be predicated by developments made in Artificial Intelligence (AI), machine learning, algorithms, massively large data sets, and robotics.

5.2

Data science is a study of data that uses the combination different techniques such as mathematics, statistics, and programming languages such as python to collect and analyse various data types like, digital sound waves, converted analogue to digital audios, images, videos, text, and many other kinds of data. Understanding how all this structured and unstructured data can be stored in various kinds of relational and non-relational DBMS technologies is vital for a database designer.

Data scientists could be considered as the science of the data world with an array of math and data skills, data scientists translate raw data into valuable insights *and* make predictions for the future.

Since data science continues to become a more complex field as technology advances, data scientists are highly valued professionals in most companies. If a company wants to pull

advanced analysis and make sound predictions based on their sets, then the organization will pour resources into their data science strategy. If you're a data science professional, you're guaranteed challenging, but rewarding work.

5.3

Open education resources as a medium of education provides free learning materials that are accessed by the public for research, teaching, and learning purposes. These educational contents are published under an open licence public domain for teachers and learners who want to consume free learning materials. This free learning materials are stored in a public database which can be access by all kinds of learners.