Research Assignment 1

1. Relational, NO SQL, Cloud, Vector databases

2. Relational databases use tables to store data, with rows representing records and columns representing attributes

3. In a relational database, a primary key is a column or set of columns that uniquely identifies each row in a table
- A foreign key is a column or a set of columns in one table that references the primary key of another table, establishing a relationship between the two tables

4. Database normalization is a process of organizing data within a database to minimize redundancy and improve data consistency and integrity.

5. a database schema is a blue-print that defines the logical structure and organization of a database, outlining how data is arranged within tables, including their fields, data types and relationship.

6. A data mart is a subset of a data warehouse focused on a specific business function or department, while a data warehouse is a centralized repository for the entire organization.

7. A data lake is a storage repository that holds large volumes of raw data in its native format, without a predefined schema. In contrast, a data warehouse stores data that has been cleaned, transformed and structured for analysis

8. A query language is a specialized computer language used to retrieve and manage data from databases
- SQL is the most commonly used query language because it is a standardized language for relational databases and provides a versatile interface for data management

9. The index provides pointers to the actual rows in the database table where data resides

10. In databases, a transaction is a unit of work that modifies the database in a consistent ~~manner~~ manner, ensuring data integrity.
ACID are properties that define how transactions should behave to ensure reliable and predictable data management

11. A database engine is the core software component that handles the storage, retrieval, and manipulation of data within a database
The specific type of database engine can significantly impact database performance, particularly in terms of query speed and efficiency

12. It is the process of designing a database by representing data and its relationships in a structured way.
- It helps ensure data integrity, improves communication between stakeholders and provides a roadmap for implementation and maintenance.

13. A view is a virtual table derives from a query, while a table actually stores data

14. A stored procedures and triggers are database objects that enable the execution of SQL code and automate database tasks

15. Batch processing deals with data in large, discrete chunks at scheduled intervals, while stream processing handles data continuously as it arrives

16. Cloud databases are hosted and managed by a third-party cloud provider, while on-premise databases are managed by the organization itself.

17. Data governance is a crucial aspect of data management that establishes policies, processes, and roles to ensure data quality, security and compliance across the organization.

18. Challenges include managing vast data volumes, ensuring data security and integrity and optimizing performance for efficient data retrieval.

19. - Data integrity guarantees that data is accurate, consistent, reliable throughout lifecycle
- it ensures that data is not corrupted, altered or lost.

20. Popular database platforms include relational database management systems like MySQL, PostgreSQL and microsoftSQL, as well as NOSQL databases such as MongoDB, Redis, Apache Cassandra.
- These platforms are used for a variety of apps, including managing web apps, large data wearehouse and mobile apps.