# SIT103 – Data and Information Management

# Assessment Task 1 - Survey Report

**Student Name: Satvik Sharma**

**Student ID: 218595095**

*Task 1A:*

Data can be defined as the raw facts that needs processing and it can be sometimes seen random and completely useless until it is arranged in a proper way. When the data is processed and organized to make it useful, it is then known as information [1]. Database is the organized collection of data which is stored and electronically from a computer system and can quickly store, manage and receive data [2]. When an application defines, retrieves, and manages data in database, it is known as Data Base Management System or DBMS[3]. The database environment means the system of component that consists and regulates data and its uses with the help of software and hardware. Thus, all these are interconnected with the roots starting from the raw data, and further if defined properly gives the information and the organization of the data leads to the creation of database and the software used to create a database is DBMS and all this is done in a data base environment.

*Task 1B:*

Advantages of DBMS:

1. Data redundancy: Sometimes multiple copies of the same file are created in the computer. This can be prevented with the help of DBMS as only one file is created and if there is any change, it is reflected immediately.[4]
2. Data sharing: With the creation of database, it creates an ability to share the data to the database users. The data can be shared based on the correct protocols being followed.[4]
3. Backup and recovery: The data is backed up periodically by the DBMS, which makes it easier for the user for future. Moreover, DBMS also restores database after a crash or system failure to its previous condition.[4]
4. Data integrity: Since there are several databases in a DBMS, all the data that is stored should be consistent and accurate as it is visible to multiple users, thus making the data integral.[4]

Disadvantages of DBMS:

1. Cost: firstly, the DBMS requires sophisticated applications and costly hardware. Moreover, the cost of maintaining the hardware is very high. Furthermore, sometimes the licenses of the DBMS are very high.[4]
2. Frequent updates: DBMS software require updates frequently and sometimes require hardware updates or buying new software, which costs money and requires its user to understand the application again.[4]
3. Management: different DBMS requires different hardware requirements and if not properly managed and updated timely, users can face issues with the application or security issues.[4]

*Task 2:*

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

*Task 2A:*

Primary key is used to identify the not null records in the table which by default can only be used once, whereas the foreign key is the type of key which is a primary key in another table and can be entered more than once in a table. [5]in table 1, the primary key is the student ID. In table 2, the unit code is primary key and student id is the foreign key. In table 3, unit code is the foreign key and no. of students failed is the primary key.

*Task 2B:*

The main difference between the column constraint is that the column constraint has an extent only over to the specified column for example in table no. 3, the column constraint can be expressed for the number of students failed, whereas in a table level constraint the scope is over the whole table.[6]

*Task 2C:*

DDL or data definition language, deals with database schemas and descriptions, of how the data should reside in the database[7] for example, TRUNCATE, which is used to remove data from the tables.[8]DML or data manipulation language deals with the manipulation of data and consists of functions[7] such as DELETE which, as its name suggests, deletes all records from data base.[9]

The DML command DELETE \* FROM TABLE*X* [without WHERE] and the DDL command TRUNCATE TABLE*X* differ as written above with the type of command, i.e. one is DDL and other is DML.[7]

TRUNCATE is used when the user wants to delete entire table and DELETE is used when single row is needed to get delete.

*Task 2D:*

ON DELETE CASCADE is used to delete the column automatically in another table when the first table is deleted [10]. For example, if I delete the first table that is, student details, the student ID (which is the foreign key in second table) will automatically get deleted. ON DELETE RESTRICT is used to protect the parent table from deletion, not the child table [11].

*Task 2E:*

DCL or data control language is responsible for access restrictions inside of the database. Main commands of DCL are GRANT and REVOKE [7]. For example, if I create user xyz@localhost and GRANT all permissions on all tables, the user will be able to see and edit the tables. Similarly, if I use REVOKE, it will cancel all permissions. The TCL or transaction control language is used to manage transactions in databases. For example, like in table2, the transaction will start with START TRANSACTION and further finishes command with COMMIT. ROLLBACK will be used to get back all the changes made [12].

*Task 2F:*

The MERGE statement is used to make changes in one table based on the values matched from another table [13] for example if I merge student names from table 1 into table 2, the table 2 will show student names as well.

*Task 2G:*

Recursive join is used to obtain the hierarchical data in the database[14], for example, if someone uses the third table and wants to know which student failed, this option will link it to the second table and will allow the user to know about the student. Joining table using alias is used to assign another name to the table and the joining it with another one, for example if I name my table 1 as stu\_dtls, and then join with table 2, It will join both the tables [15].

*Task 3A:*

The steps that are required to create a new database via planning, requirement gathering, conceptual design, local design et cetera is known as Database Development Life Cycle.[16]

1. Initial Study: the first step in a database development lifecycle is the initial study, which means to study about the present situation, defining problems and constraints, defining objectives and setting up scopes and boundaries.[17]also it requires consultation with and agreement among the stake holders about what the want from the system.[18]
2. Analysis: It begins with the consideration of what system should do and terminates with the productions of a specified system.[18]
3. Detailed system design: all the design of the system’s processes, including all the technical specifications for screen, menus, reports and other devices.[19]
4. Implementation: the physical realization of the database and application design are to be done during this phase, that is why it is called the programming phase of the systems design.[20]
5. Testing: During this process, the database is running to find the errors and show that the database and application programs are working according to the requirement.[21]
6. Maintenance: it involves dealing with the changes required with the time such as the operating system or the hardware requirements. The life cycle is repeated in this step to bring more features to the already developed software in the market.[22]

*Task 3B:*

Big data can be defined as extremely large datasets that may be analysed computationally to reveal patterns, trends, and associations.[23] Big data can be distinguished by:

1. Volume: the quantity of the stored data.[24]
2. Variety: the type and nature of data.[24]
3. Velocity: the speed at which the data is generated in order to make the demands.[24]
4. Veracity: it refers to the quality and value of the data.[24]

Some of the technologies that work with big data are the Hadoop ecosystem[25], which uses a network of many computers to solve big data problems[26], the programming language R, which is another open source project designed for working with statistics[25], predictive analytics, which predicts the data for the future based on the historical data[27], data lakes, which makes it easier to access vast stores of data.

problems faced by traditional databases are listed below:

1. Cost of storage: more storage means more money.[27]
2. Complexity: traditional database is full of complex silos of system administration, DBAs et cetera. Often there is one DBA for every 40-50 database servers.[27]
3. Causation: since the cost of storage is too much, data is filtered out and the remaining is thrown out, which can lead to inaccuracy of results.[27]
4. Cost of proprietary hardware: Large proprietary hardware solutions can be cost prohibitive when deployed to process extremely large volumes of data.[21]

*Task 4:*

Present day technology has made it easier to receive and store large amounts of data. Database systems are the most common means of maintaining and utilizing the data.[29] Since the database consists of data, it can sometimes lead to its breaching and can be further used by hackers. The data leaking of Facebook is the best example of hacking data because of a poor data base. More than 540 million records about Facebook users were publicly exposed on amazon cloud’s computing services [30]. If certain measures are taken, this kind of data breach can be prevented. First step in securing a database is to make it authentic and only authorized can enter the systems. Authentication is an act of proving an assertion, such as identity of a computer system user [31], for example, mobile phone asks for PIN and performs authentication [32]. There are several kinds of authentications. Most basic form of authentication authentication can be passwords, which can be used to prevent unauthorized use of database [33]. All types of authentications are based on three factors. First one is something the user knows, like passwords [34]. Second type is something the user is having, like smart cards. Last type is something the user is, which includes part of the human body such as fingerprint [35]. Second step to make a database more secure is to get its backup which means copy of data from the database is saved that can be used to reconstruct that data [36]. There are mainly two types of backups, physical backup and logical backup and the main difference between these two is that physical backup is to copy for backing up all the physical files, such as data files, log files et cetera, either when the database is up or when the database is shut down whereas logical backup extracts just the data from data files into dump files [37]. Next step is to protect the data from the DOS attacks or denial of service attacks which is basically a cyber-attack [38] in which attackers attempt to prevent legitimate users from accessing the service [39]. Different types of attacks are (S)SYN flood[40], in which an attacker sends a series of SYN requests, teardrop attacks[41], in which the hacker sends broken and disorganized IP fragments, Low-rate DOS attacks[42], which slows down the machine, ICMP flood[43], in which the hacker sends in a large number of abnormalities, and peer to peer attacks[44], in which network act as both supplier and consumer[45]. Lastly, the insider protection is necessary which, means the breaching of data by the people working inside the company [46].

*Bibliography:*

1. Diffen.com. (2019). *Data vs Information - Difference and Comparison | Diffen*. [online] Available at: [https://www.diffen.com/difference/Data\_vs\_Information](https://www.diffen.com/difference/Data_vs_Information%20) [Accessed 23 Aug. 2019].
2. En.wikipedia.org. (2019). *Database*. [online]Available at: <https://en.wikipedia.org/wiki/Database> [Accessed 23 Aug. 2019].
3. Techopedia.com. (2019). *What is a Database Management System (DBMS)? - Definition from Techopedia*. [online] Available at: <https://www.techopedia.com/definition/24361/database-management-systems-dbms> [Accessed 24 Aug. 2019].
4. Tutorialspoint.com. (2019). *Advantages of Database Management System*. [online] Available at: <https://www.tutorialspoint.com/Advantages-of-Database-Management-System> [Accessed 24 Aug. 2019].
5. "Difference between Primary Key and Foreign Key", *Dotnettricks.com*, 2019. [Online]. Available: <https://www.dotnettricks.com/learn/sqlserver/difference-between-primary-key-and-foreign-key>. [Accessed: 26- Aug- 2019].
6. Ask TOM Difference between Table level and Column level constraints?", *Asktom.oracle.com*, 2019. [Online]. Available: <https://asktom.oracle.com/pls/apex/f?p=100:11:0::::P11_QUESTION_ID:2665518700346300293>. [Accessed: 26- Aug- 2019].
7. MySQL What is DDL, DML and DCL?", *W3schools*, 2019. [Online]. Available: <https://www.w3schools.in/mysql/ddl-dml-dcl/>. [Accessed: 26- Aug- 2019].
8. TRUNCATE TABLE (Transact-SQL) - SQL Server", *Docs.microsoft.com*, 2019. [Online]. Available: <https://docs.microsoft.com/en-us/sql/t-sql/statements/truncate-table-transact-sql?view=sql-server-2017>. [Accessed: 26- Aug- 2019].
9. S. Tutorial, "SQL DELETE - Deleting Data in a Table or Multiple Tables", *ZenTut*, 2019. [Online]. Available: <https://www.zentut.com/sql-tutorial/sql-delete/>. [Accessed: 26- Aug- 2019].
10. Techonthenet.com. (2019). *SQL Server: Foreign Keys with cascade delete*. [online] Available at: <https://www.techonthenet.com/sql_server/foreign_keys/foreign_delete.php> [Accessed 27 Aug. 2019].
11. D. Downey, "How to use RESTRICT for Foreign Key in mysql?", *Database Administrators Stack Exchange*, 2019. [Online]. Available: <https://dba.stackexchange.com/questions/14155/how-to-use-restrict-for-foreign-key-in-mysql>. [Accessed: 27- Aug- 2019].
12. Shvechikov, D. (2019). *DDL, DML, DCL & TCL Commands in SQL ᐈ Examples*. [online] ExplainJava.com. Available at: <https://explainjava.com/dml-ddl-dcl-tcl-sql-commands/> [Accessed 27 Aug. 2019].
13. "Merge (SQL)", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Merge_(SQL)>. [Accessed: 27- Aug- 2019].
14. "SQL - Recursive JOINs | sql Tutorial", *Riptutorial.com*, 2019. [Online]. Available: <https://riptutorial.com/sql/example/11142/recursive-joins>. [Accessed: 27- Aug- 2019].
15. "SQL Server Self Join by Practical Examples", *SQL Server Tutorial*, 2019. [Online]. Available: <http://www.sqlservertutorial.net/sql-server-basics/sql-server-self-join/>. [Accessed: 27- Aug- 2019].
16. “The database development lifecycle", *Linkedin.com*, 2019. [Online]. Available: <https://www.linkedin.com/learning/learning-relational-databases-2/the-database-development-lifecycle>. [Accessed: 27- Aug- 2019].
17. "Database Development Life Cycle", *Myreadingroom.co.in*, 2019. [Online]. Available: <http://www.myreadingroom.co.in/notes-and-studymaterial/65-dbms/506-database-development-life-cycle.html>. [Accessed: 27- Aug- 2019].
18. A. Watt, "Chapter 13 Database Development Process – Database Design – 2nd Edition", *Opentextbc.ca*, 2019. [Online]. Available: <https://opentextbc.ca/dbdesign01/chapter/chapter-13-database-development-process/>. [Accessed: 27- Aug- 2019].
19. "Database Design", *Shsu.edu*, 2019. [Online]. Available: <https://www.shsu.edu/~csc_tjm/summer2000/cs334/chapter06/chapter6.html>. [Accessed: 27- Aug- 2019].
20. Haider, "Database development life cycle", *Slideshare.net*, 2019. [Online]. Available: <https://www.slideshare.net/AfrasiyabHaider/database-development-life-cycle>. [Accessed: 27- Aug- 2019].
21. *Fid.cl*, 2019. [Online]. Available: <http://fid.cl/courses/db/dbms_lifecycle/dbmsLifecycle.pdf>. [Accessed: 27- Aug- 2019].
22. "The database development life cycle: View as single page", *Open.edu*, 2019. [Online]. Available: <https://www.open.edu/openlearn/ocw/mod/oucontent/view.php?printable=1&id=2786>. [Accessed: 27- Aug- 2019].
23. "big data | Definition of big data by Lexico", *Lexico Dictionaries | English*, 2019. [Online]. Available: <https://www.lexico.com/en/definition/big_data>. [Accessed: 27- Aug- 2019].
24. "Big data", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Big_data>. [Accessed: 27- Aug- 2019].
25. "Top 15 Big Data Technologies", *Datamation.com*, 2019. [Online]. Available: <https://www.datamation.com/big-data/big-data-technologies.html>. [Accessed: 27- Aug- 2019].
26. [5]"Apache Hadoop", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Apache_Hadoop>. [Accessed: 27- Aug- 2019].
27. "What are Predictive Analytics?", *PAT RESEARCH: B2B Reviews, Buying Guides & Best Practices*, 2019. [Online]. Available: <https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/>. [Accessed: 27- Aug- 2019].
28. G. Trujillo, R. Garcia, S. Jones, C. Kim and J. Murray, "Traditional Data Systems | Understanding the Big Data World | Pearson IT Certification", *Pearsonitcertification.com*, 2019. [Online]. Available: <http://www.pearsonitcertification.com/articles/article.aspx?p=2427073&seqNum=2>. [Accessed: 27- Aug- 2019].
29. A. Saeed and J. Hall *Accelerated pavement testing*. Washington, D.C.: Transportation Research Board, 2003.
30. "Hundreds of millions of Facebook user records were exposed on Amazon cloud server", *Cbsnews.com*, 2019. [Online]. Available: <https://www.cbsnews.com/news/millions-facebook-user-records-exposed-amazon-cloud-server/>. [Accessed: 27- Aug- 2019].
31. "Authentication", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Authentication>. [Accessed: 27- Aug- 2019].
32. "What is Database Authentication? - Definition from Techopedia", *Techopedia.com*, 2019. [Online]. Available: <https://www.techopedia.com/definition/27388/database-authentication>. [Accessed: 27- Aug- 2019].
33. "4 Authenticating Users to the Database", *Docs.oracle.com*, 2019. [Online]. Available: <https://docs.oracle.com/cd/B12037_01/network.101/b10777/authuser.htm>. [Accessed: 27- Aug- 2019].
34. En.wikipedia.org. (2019). *Multi-factor authentication*. [online] Available at: [https://en.wikipedia.org/wiki/Multi-factor\_authentication](https://en.wikipedia.org/wiki/Multi-factor_authentication%20) [Accessed 27 Aug. 2019].
35. "The Three Types of Multi-Factor Authentication", *Globalknowledge.com*, 2019. [Online]. Available: <https://www.globalknowledge.com/ca-en/resources/resource-library/articles/the-three-types-of-multi-factor-authentication/>. [Accessed: 27- Aug- 2019].
36. Docs.oracle.com. (2019). *Backup and Recovery Overview*. [online] Available at: <https://docs.oracle.com/cd/B19306_01/backup.102/b14192/intro001.htm> [Accessed 27 Aug. 2019].
37. "What is the difference between physical and logical backup?", *Stack Overflow*, 2019. [Online]. Available: <https://stackoverflow.com/questions/39123563/what-is-the-difference-between-physical-and-logical-backup/39124218>. [Accessed: 27- Aug- 2019].
38. "Denial-of-service attack", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Denial-of-service_attack>. [Accessed: 27- Aug- 2019].
39. "What is a Denial-of-Service Attack (DoS)? - Definition from Techopedia", *Techopedia.com*, 2019. [Online]. Available: <https://www.techopedia.com/definition/24841/denial-of-service-attack-dos>. [Accessed: 27- Aug- 2019].
40. "SYN flood", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/SYN_flood>. [Accessed: 27- Aug- 2019].
41. *Security.radware.com*, 2019. [Online]. Available: <https://security.radware.com/ddos-knowledge-center/ddospedia/teardrop-attack/>. [Accessed: 27- Aug- 2019].
42. L. Zhou, M. Liao, C. Yuan and H. Zhang, "Low-Rate DDoS Attack Detection Using Expectation of Packet Size", *Security and Communication Networks*, vol. 2017, pp. 1-14, 2017. Available: 10.1155/2017/3691629.
43. 2019. [Online]. Available: <https://www.cloudflare.com/learning/ddos/ping-icmp-flood-ddos-attack/>. [Accessed: 27- Aug- 2019].
44. *Security.radware.com*, 2019. [Online]. Available: <https://security.radware.com/ddos-knowledge-center/ddospedia/peer-to-peer-attack/>. [Accessed: 27- Aug- 2019].
45. H. Dalziel, "5 Major Types of DOS Attack | Learn & Know Them", *Concise Courses*, 2019. [Online]. Available: <https://www.concise-courses.com/5-major-types-of-dos-attack/>. [Accessed: 27- Aug- 2019].
46. "Insider threat", *En.wikipedia.org*, 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Insider_threat>. [Accessed: 27- Aug- 2019].