**SIT103-Data and Information Management**

**Survey report**

Prepared by: Iftekhar Qureshi

Student ID: 218676618

Lecturer: Dr. Arash Shaghaghi

*28th August 2019*

# Task 1a

Data is related to information. In fact, we get information from data but raw data is meaningless to us humans. When that data is formatted in a logical way, that allows us humans to utilize it, it becomes an information. Data can be any character, text, word and number. A valid example of the relation would be a computer using formulas, programming scripts, or software applications to turn data into information [1].

We can store, manage and retrieve the manipulated data/information in an organised mechanism, which is using a database. In databases, the information is stored in tables.

The database will hold the data for us but to retrieve, add, delete, and modify the data we need a software. A Database Management System (DBMS) is a software that has the database hosted inside it and all the functionality to manage the data in the database. [2] [3].

Everything that I have mentioned above are placed in an environment called database environment. The environment includes software, hardware, people, procedures and the data itself. Software is everything starting from the operating system to the application programs that includes DBMS. People are the database administrators and procedures are rules, concepts and instructions given to the administrators [4] [5].

# Task 1.b

Advantages of DBMS [6] [7]

1. Better sharing of data:

In DBMS, same data can be shared by many users. Usually, there is a database administrator who is the head of the database management, and this administrator manages and gives rights to other users to access the data. Also, the data inside the same database can be shared between different application programs.

1. Data security:

Since the data is accessible to many people, the possibility of security breaches in DBMS is also higher. Hence, high investments are made by corporations to improve security. Also, DBMS provides a framework for better implementation of security.

1. Data integration:

In a single database, multiple tables can be created to store the data for an organisation. Relationships between these tables can easily be formed which makes retrieving and updating data easier. One section of the company can see how their actions affect other sections giving a clear view of the inside.

1. Minimized data inconsistency:

Data is inconsistent when same data in a database is wrong in different places. For instance, mistake can be made with a person’s name-Iftekhar Qureshi in a school’s register can be recorded as Iftikar Qureshi in the scholarship program database of the same school. DBMS minimises the chances of this discordance by a properly designed database.

Disadvantages of DBMS [6] [7]

1. Increased costs:

We need advanced software and hardware to implement database systems. These are expensive to buy and maintain for a long time. Highly skilled and high paid staff are required for the maintenance.

1. Keeping the system up-to-date:

DBMS vendors frequently upgrade their product to bring new features and improve security. To stay up-to-date with those and maximize efficiency, companies need to update their software and in some cases they even have to change to new hardware.

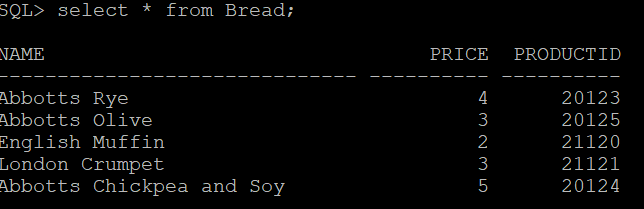
1. Database damage:

Although the security is strong, it is not impossible for hackers to get into the system. Database can get corrupted. Regular backups should be done with any database. Power failure should be taken care of and uninterruptible power supply (UPS) should be implemented.

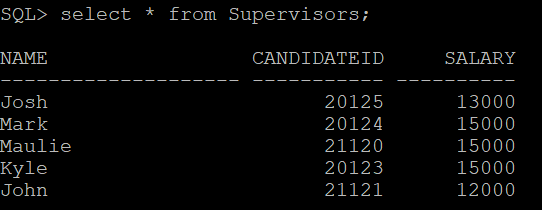
# Task 2.a

Primary key is a unique identifier of a table. We can have only one primary key per table. On the other hand, foreign key is an identifier that is a primary key of another table. We can have more than one foreign key in a table linking to primary keys of other tables.

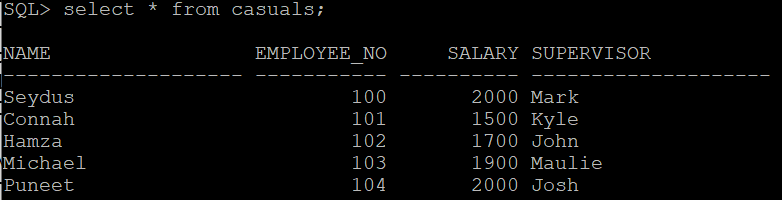
For my database of a bread company, I assigned the **“productid”** column as my primary key for the table named **“bread”** and the primary key is called **“productid\_pk”.**

****

I linked this primary key to **“candidateid**” of the **“Supervisors**” table (name of the foreign key is **“candidateid\_fk”**) so that each supervisor are assigned to work in the section with the right product.



A primary key (which is **“name\_pk”**) is assigned to names of the supervisors so that casuals (which is my third table) can be designated to their correct supervisors. Foreign key in “**Casuals”** table is in the supervisors’ column and named **“supervisor\_fk”.**

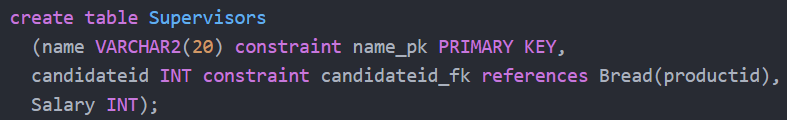


# Task 2.b

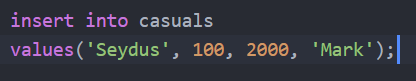
Column constraints are applied to one column only whereas table constraints are applied to a whole table or groups of columns. In my database, all the primary keys are defined at column level. Although there are no table level constraints in my database, I could have, for example, defined two columns-name and productid-in the bread table as the primary key columns. This two keys would make a composite key, which is defined at table level. Not null cannot be defined at table level but at column level we can.

# Task 2.c

DDL is data definition language and DML is data manipulation language. We use DDL when we want to create, delete or modify the structure of database objects in a database. An example of DDL command: CREATE – for creating tables, index, function etc.



DML as its full name suggests, is used to manipulate data which is already present in the database. Example of DML command: INSERT- for inserting data into a table.



TRUNCATE is a DDL command and DELETE is DML command. TRUNCATE deletes entire data from a table. Less information is logged when we truncate a table which means the statement is executed faster. However, each rows are not logged when deleted which means we have a possibility of losing the data forever.

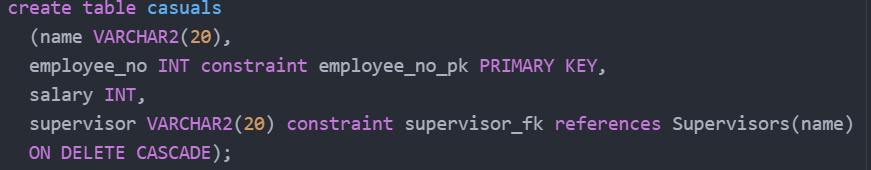
DELETE command is also used to remove records. We can remove one row at a time. We have the luxury of using the WHERE clause to only remove the data that meets a certain criteria. It is easier to recover deleted data as well.

We should only use TRUNCATE when we want to reset an entire table to its default. Apart from that, using DELETE will be a safer approach. [8] [9]

# Task 2.d

ON DELETE CASCADE: if we declare this in the parent table, and then in the future, delete rows from the parent, it will delete all the corresponding rows from the child table. [10]

ON DELETE RESTRICT: if we declare this in the parent table, and then in the future, delete a row from the parent, we will not be able to delete the corresponding row in the child table. [11]



# Task 2.e

DCL commands are used to set up permissions, rights and privileges or anything that requires authorisation to access the database. For example- GRANT can be used in my database to give access to only the privileged people such as the supervisors, not the casuals. REVOKE can be used to take the same privilege off from supervisors provided by GRANT. [12]

TCL commands are used to take care of any transaction within the database. Example- COMMIT which commits the current transaction. I have used commit command to make sure that the data that I have inserted in the tables do not get lost, if my Oracle server disconnects for some reason. ROLLBACK can be used to bring back mistakenly deleted data. [13]

# Task 2.f

MERGE is a combination of three statements-INSERT, DELETE and UPDATE. We can perform all of them at once if we use MERGE. I can use merge in my database to insert new item in the bread table and delete an old item. Also, I can update the price of an existing bread.

# Task 2.g

Recursive expressions are used to join a table with itself [14]. This can be useful in my database. I can use it on my casuals table that has a supervisor column that points to the casual that is the boss of the current casual.

Joining table with aliases is useful when we join two tables but want to shorten a long table name for easier reading. Aliases are not valid outside the query. I can use table aliases to prevent errors when joining bread table and supervisor table because both have the same column called “name” [15].

# Task 3.a

There are six stages of the database development lifecycle typically [16] [17] [18] [19] [20]. These are-

1. Analysis and planning:

In this stage, the group investigates the current system and points out the positive and negative aspects of the system. Then further planning is done on the basis of the study. They see if the new system is what the users want. New objectives are placed and resources, costs, time, benefits etc. are discussed.

1. Database Design:

Design is where the creation of the database starts. The team has to make sure that they design the proper database using the right essential components (hardware, software and networking components) and fulfil the needs of end users.

1. Implementation:

From the previous stage, the team gets an output design which shows the creation of tables, attributes, domains, views, indexes, security constraints, and storage and performance guidelines. On basis of that, DBMS is installed, databases are created and old data is migrated from the prior version of the system.

1. Testing:

In testing, the database administrator tests and fine-tunes the database to ensure that it performs as expected, usually in conjunction with the associated applications. Security is tested and data integrity is enforced through the use of primary and foreign key rules.

1. Operation:

After testing and evaluation, the database is released for operation. The users get access to the system and they start the process of system evolution.

1. Maintenance:

After successful deployment, database admin has to maintain the system by doing regular backups, recovery, adding resources as per user demand and controlling privileges for new and old users.

# Task 3.b

Big data, as its name suggests, means that a lot of data is accumulated at a single place, although it is not necessary that only large amount of data is called big data [21] [22]. Four attributes define most of big data- Volume, variety, velocity (frequency of incoming data), and veracity (trustworthiness) [23]. Having variety of data is an important characteristic of big data. This brings in millions of unstructured data and this data is not possible to store in the traditional relational database systems. Hence, the big data technology is mainly classified into two types: Operational and Analytical [21]. Operational big data is all about the simple transactions that we do every day. For example- online shopping and texting on Facebook. Analytical is more about the complex ones such as weather forecast and space mission.

The technologies are divided into four fields- Data storage, mining, analytics and virtualization. For storage we have popular ones like Hadoop Framework and Mongo DB. For mining we have Presto and Rapid Miner, Kafka and Splunk for analytics part, and Spark and R-language for virtualization. [24]

Some of the difficulties that traditional systems have with big data are [25] [26]-

* Schema-On-Write: Traditional systems are schema-on-write which means that the data has to be validated when it’s being written. A company has to spend a lot of time designing schemas if they want to analyse new data from a source. This process is too slow to deal with unstructured data.
* Cost of storage: Shared storage cost increases as a company starts expanding their storage to host big data.
* Cost of hardware: Large data environments require sophisticated hardware and software which are extremely expensive.
* Complexity: Large teams work behind the database but there is often one DBA for 40 to 50 servers. Handling this is critical as traditional complex systems fail in complex ways.
* Causation: Since data is so expensive to store, a lot of data is thrown out while cleaning the data which makes analysis of the data inaccurate.

# Task 4

Database contain sensitive information. Hence, they are vulnerable to hacking and need good security. Firstly, if a database’s authentication is weak then it will be easy entering the database. There are two authentication mode in sql: Normal authentication and mixed mode [27].

Also, audit can be done to go deeper and make sure any unauthorised person is in control of our database. [28]

Back and recovery: This should be done regularly for some obvious reasons-

* Natural disaster- flood, earthquake, tsunami etc. can cause permanent loss of data.
* Accidental damage- DBA are humans and humans are prone to making mistakes.
* Software or hardware failure- The system can auto turn off or crash causing many important data loss.

Some recovery techniques are[29]:

* Salvation program
* Incremental dumping
* Audit trail
* Differential files
* Backup/current version
* Multiple copies
* Careful replacement

Denial of service (DOS): The hacker takes on a whole database and prevents the actual users from accessing the database usually by flooding attacks or crashing attacks. Early detection of such attacks is the best way of preventing it[30].

# References

|  |  |
| --- | --- |
| [1] | Computer Hope, “Computer Hope,” 2018. [Online]. Available: https://www.computerhope.com/issues/ch001629.htm. [Accessed 23 August 2019]. |
| [2] | M.Chapple, “lifewire,” 10 July 2019. [Online]. Available: https://www.lifewire.com/what-is-a-database-1019737. [Accessed 23 August 2019]. |
| [3] | M.Chapple, “lifewire,” 10 July 2019. [Online]. Available: https://www.lifewire.com/database-management-system-1019609. [Accessed 23 August 2019]. |
| [4] | “W3schools,” 2019. [Online]. Available: https://www.w3schools.in/dbms/database-environment/. [Accessed 23 August 2019]. |
| [5] | “Reference,” 2019. [Online]. Available: https://www.reference.com/technology/database-environment-d298af3eec40be6b. [Accessed 23 August 2019]. |
| [6] | “Tutorialink,” 2019. [Online]. Available: https://tutorialink.com/dbms/advantage-and-disadvantages-of-dbms.dbms. [Accessed August 23 2019]. |
| [7] | Shifali, “COMPUTER NOTES,” 11 September 2012. [Online]. Available: http://datafiletech.blogspot.com/2012/09/advantages-and-disadvantages-of-dbms.html. [Accessed 23 August 2019]. |
| [8] | A.Gaikwad, “MSSQLtips.com,” 20 April 2016. [Online]. Available: https://www.mssqltips.com/sqlservertip/4248/differences-between-delete-and-truncate-in-sql-server/. [Accessed 26 August 2019]. |
| [9] | K.Wenzel, “essentialSQL,” 2019. [Online]. Available: https://www.essentialsql.com/what-is-the-difference-between-truncate-and-delete-in-sql-server/. [Accessed 26 August 2019]. |
| [10] | “Techothenet.com,” 2019. [Online]. Available: https://www.techonthenet.com/sql\_server/foreign\_keys/foreign\_delete.php. [Accessed 26 August 2019]. |
| [11] | “Code Examples,” 2019. [Online]. Available: https://code-examples.net/en/q/13a9fcb. [Accessed 26 August 2019]. |
| [12] | D. Varshni, “GeeksforGeeks,” 2019. [Online]. Available: https://www.geeksforgeeks.org/sql-ddl-dml-dcl-tcl-commands/. [Accessed 26 August 2019]. |
| [13] | “Tutorial Gateway,” 2019. [Online]. Available: https://www.tutorialgateway.org/sql-dml-ddl-dcl-and-tcl-commands/. [Accessed 26 August 2019]. |
| [14] | “SqlSeverTutorial.net,” [Online]. Available: http://www.sqlservertutorial.net/sql-server-basics/sql-server-self-join/. [Accessed 27 August 2019]. |
| [15] | “Quackit,” [Online]. Available: https://www.quackit.com/sql/tutorial/sql\_alias.cfm. [Accessed 27 August 2019]. |
| [16] | “My Reading Room,” 2019. [Online]. Available: http://www.myreadingroom.co.in/notes-and-studymaterial/65-dbms/506-database-development-life-cycle.html. [Accessed 27 August 2019]. |
| [17] | “MariaDB,” 2019. [Online]. Available: https://mariadb.com/kb/en/library/database-lifecycle/. [Accessed 27 August 2019]. |
| [18] | “Innovative Architects,” 2019. [Online]. Available: https://www.innovativearchitects.com/KnowledgeCenter/basic-IT-systems/system-development-life-cycle.aspx. [Accessed 27 August 2019]. |
| [19] | “Fandom,” 2019. [Online]. Available: https://databasemanagement.fandom.com/wiki/SDLC. [Accessed 27 August 2019]. |
| [20] | A.Diwan, “Tutorialspoint,” 2018 July 2018. [Online]. Available: https://www.tutorialspoint.com/Database-Life-Cycle. [Accessed 27 August 2019]. |
| [21] | A.Subramaniam, “edureka!,” 22 May 2019. [Online]. Available: https://www.edureka.co/blog/what-is-big-data/. [Accessed 27 August 2019]. |
| [22] | S.Madden, “IEEE Xplore: Digital Library,” 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6188576. [Accessed 27 August 2019]. |
| [23] | J.Williamson, “dummies: A Willey Brand,” 2019. [Online]. Available: https://www.dummies.com/careers/find-a-job/the-4-vs-of-big-data/. [Accessed 27 August 2019]. |
| [24] | R.Kiran, “edureka!,” 18 June 2019. [Online]. Available: https://www.edureka.co/blog/top-big-data-technologies/. [Accessed 27 August 2019]. |
| [25] | “FlyData,” [Online]. Available: https://www.flydata.com/the-6-challenges-of-big-data-integration/. [Accessed 27 August 2019]. |
| [26] | G.Trujillo, “Pearson IT Certification,” 20 August 2015. [Online]. Available: http://www.pearsonitcertification.com/articles/article.aspx?p=2427073&seqNum=2. [Accessed 27 August 2019]. |
| [27] | “Microsoft,” 2017. [Online]. Available: https://docs.microsoft.com/en-us/sql/relational-databases/security/choose-an-authentication-mode?view=sql-server-2017. [Accessed 27 August 2019]. |
| [28] | D.Asanka, “SQLShack,” 20 November 2017. [Online]. Available: https://www.sqlshack.com/sql-server-auditing-server-database-audit-specifications/. [Accessed 27 August 2019]. |

[29] Infolab.stanford.edu. (2019). *Recovery Techniques for Database Systems*. [online] Available at: http://infolab.stanford.edu/~manku/quals/summaries/spiller-recovery.htm [Accessed 27 Aug. 2019].

[30] Us.norton.com. (2019). *What are Denial of Service (DoS) attacks? DoS attacks explained*. [online] Available at: https://us.norton.com/internetsecurity-emerging-threats-dos-attacks-explained.html [Accessed 27 Aug. 2019].