**Raymond You**

**CS 3200**

**Homework 1 (Chapter 1 -3 Connolly & Begg)**

**Each question is worth 10 points each**

**Discuss the differences between DDL and DML. What operations would you typically expect to be available in each language?**

-A data sublanguage consists of two parts: A Data Definition Language (DDL) and a Data Manipulation Language (DML).

-The Data Definition Language (DDL) is used to define a schema or to modify an existing one. It cannot be used to manipulate data. The database schema is specified by a set of languages expressed through DDL. Because of compiling the DDL statements, a set of tables stored in special files collectively called the system catalog is created. The system catalog integrates the metadata, which is data that describes the objects in the database and makes it easier for those objects to be accessed or manipulated. The DBMS normally consults the system catalog before the actual data is accessed in the database. The terms data definition and data direction also describe the system catalog.

-Data Manipulation Language (DML) is a language that provides a set of operations to support the basic data manipulation operations on the data held in the database. DML operations include insertion of new data into the database, modification of data stored in the database, retrieval of data contained in the database, and deletion of data from the database. The DML that involves data retrieval also has a high-level special-purpose language that satisfies diverse requests for the retrieval of data held in the database. This is called a query language. The DML is also distinguished by their underlying retrieval constructs: procedural DMLs and nonprocedural DMLs. Procedural languages specify how the output of a DML statement is to be obtained, while nonprocedural DMLs describe only what output is to be obtained.

**Define data security and data integrity. Describe the differences between data security and data integrity.**

-Data integrity is the validity and consistency of stored data. Integrity is expressed in terms of constraints, which are consistency rules that the database is not permitted to violate. These constraints may apply to data items within a single record r to relationships between records. For example, an integrity constraint could state that a member of staff’s salary cannot be greater than $40,000. Integration allows the DBA to define integrity constraints, and the DBMS to enforce them.

-Data security is the protection of the data from authorized users. Without suitable security measures, integration makes the data more vulnerable than file-based systems. Integration allows the DBA to define database security, and the DBMS to enforce it. This security may be user names and passwords to identify people authorized to use the database. Each user can also have different amounts of restricted access such as by the operation type (retrieval, insert, update, and delete)

**Describe the main characteristics of the database approach and contrast it with the file-based approach.**

-The file-based approach is where there is a collection of application programs that perform services for the end-users, such as the production of reports and each program defines and manages its own data. File-based systems were an early attempt to computerize the manual filing system that we were all familiar with. Each file is simply a collection of records, which contains logically related data and each record contains a logically connected set of one or more fields, where each field represents some characteristic of the real-world object that is being modeled. Limitations of the File-based approach are separation and isolation of data, duplication of data, data dependence, incompatible file formats, and fixed queries/proliferation of application programs.

-The database approach is where there is a shared collection of logically related data and its description, designed to meet the information needs of an organization. Data abstraction allows one to change the internal definition of an object without affecting the users of the object, provided that the external definition remains the same. The database has a Database Management System (DBMS) which is a software system that enables users to define, create, maintain, and control access to the database. The DBMS is the software that interacts with the users’ application programs and the database. A typical DBMS contains DDL and DML including query languages. A Database Application Program is a computer program that interacts with the database by issuing an appropriate request (typically an SQL statement) to the DBMS. The five major components in the DBMS environment are hardware, software, data, procedures, and people.

-Advantages of DBMS are control of data redundancy, data consistency, more information from the same amount of data, sharing of data, improved data integrity, improved security, enforcement of standards, economy of scale, balance of conflicting requirements, increased productivity, and increased concurrency. The disadvantages of DBMSs are complexity, size, cost of DBMSs, additional hardware costs, cost of conversion, performance, and greater impact of a failure.

**Provide a definition for a data administrator and a database administrator. What types of interactions would these two users of the database have?**

­-The Data Administrator (DA) is responsible for the management of the data resource, including database planning; development and maintenance of standards, policies and procedures; and conceptual/logical database design. The DA consults with and advises senior managers, ensuring that the direction of database development will ultimately support corporate objectives.

-The Database Administrator (DBA) is responsible for the physical realization of the database, including physical database design and implementation, security and integrity control, maintenance of the operational system, and ensuring satisfactory performance of the application for users. The role of the DBA is more technically oriented than the role of the DA, requiring detailed knowledge of the target DBMS and the system environment. In some organizations, there is no distinction between these two roles.

**Name three record-based data models. Discuss the main differences between these data models.**

1. Relational data model – In the relational mode, data and relationships are represented as tables, each of which has a number of columns with a unique name. the relational data model requires only that the database be perceived by the user as tables.
2. Network data model – In the network data model, data is represented as collections of records, and relationships are represented by sets. The records are organized as generalized graph structures with records appearing as nodes (also called segments) and sets as edges in the graph.
3. Hierarchical data model – The hierarchical model is a restricted type of network model where data is represented as collections of records and relationships are represented by sets. However, the hierarchical model allows a node to have only one parent. A hierarchical model can be represented as a tree graph, with records appearing as nodes (also called segments) and sets as edges.

**What are the advantages of a relational database when compared to the file-based approach to storing data?**

-Advantages of a relational database when compared to the file-based approach to storing data are:

1. Control of data redundancy – doesn’t completely eliminate redundancy but controls the amount of redundancy inherent in the database
2. Data consistency – by storing a data item only once in the database, only need to preform updates to the value once
3. More information from the same amount of data
4. Sharing of data – database belongs to the entire organization and can be shared by all authorized users
5. Improved data integrity – usually expressed in terms of constraints, which are consistency rules that the database is not permitted to violate
6. Improved security – protection of the database from unauthorized users. Without suitable security measures, integration makes the data more vulnerable than file-based systems
7. Enforcement of standards – integration allows the DBA to define and the DBMS to enforce the necessary standards
8. Economy of scale
9. Balance of conflicting requirements - Because the database is under the control of the DBA, the DBA can make decisions about the design and operational use of the database that provide the best use of resources for the organization as a whole.
10. Improved Data accessibility and responsiveness
11. Increased productivity
12. Improved maintenance through data independence
13. Increased concurrency
14. Improved backup and recovery services

**What is concurrency control and why does a DBMS need a concurrency control facility?**

-Concurrency control allows the DBMS to be updated correctly when multiple users are updating the database concurrently. Concurrent access is relatively easy if all users are only reading data, as there is no way that they can interfere with one another. However, when two or more users are accessing the database simultaneously and at least one of them is updating data, there may be interference that can result in inconsistencies. The DBMS must ensure that when multiple users are accessing the database, interference cannot occur.

**What is a transaction?  Give an example of a transaction.**

-A transaction is a series of actions, carried out by a single user or application program, which accesses or changes the contents of the database. For example, some simple transactions for the DreamHome case study might be to add a new member of staff to the database, to update the salary of a member of staff, or to delete a property from the register.

-A more complicated example might be to delete a member of staff from the database and to reassign the properties that he or she managed to another member of staff.

**What is meant by the term ‘client-server architecture’ and what are the advantages of this approach? Compare the client-server architecture with two other architectures.**

-Client-server architecture is where there are software components which interact to fore a system. There is a client process, which requires some resource, and a server, which provides the resource. The advantages of this type of architecture are that it enables wider access to existing databases, increased performance, hardware cost reduced maybe, communication costs reduced, increased consistency, and that it maps on to open systems architecture quite naturally.

-The Three-Tier Client-Server Architecture appeared to solve the problem of enterprise scalability for the traditional two-tier client-server model. There are three layers: the user interface layer, which runs on the end-user’s computer (the client), the business logic and data processing layer that runs on a server and is often called the application server, and a DBMS which stores the data and may run on a database server.

-N-Tier Architectures - the three-tier architecture can be expanded to n tiers, with additional tiers providing more flexibility and scalability.

**What is a Transaction Processing Monitor? What advantages does a TP Monitor bring to an OLTP environment?**

-A Transaction Processing Monitor is a program that controls data transfer between clients and servers in order to provide a consistent environment, particularly for online transaction processing.

-TP Monitors provide significant advantages including:

1. Transaction routing - increase scalability by directing transactions to specific DBMSs
2. Managing distributed transactions – manage transactions that require access to data held in multiple, possibly heterogenous DBMSs
3. Load balancing – balance client requests across multiple DBMSs on one or more computers by directing client service calls to the least loaded server
4. Funneling – establish connections with the DBMSs as and when required, and can funnel user requests through these connections
5. Increased reliability – act as a transaction manager, performing the necessary actions to maintain the consistency of the database, with the DBMS acting as a resource manager.