CS561: Database Management Systems Notes

Steven DeFalco

Fall 2023

Contents

1 Introduction

Database management systems (DBMS) consist of **data**, **software** (programs such as data interfaces), and **environments** (operating systems). DBMS contains information about a particular enterprise. They are collections of interrelated data, a set of programs to access the data, and an environment that is both *convenient* and *efficient* to use. The *user* should only have to define what it is that they want from the database, whereas the *database* is responsible for defining how this query can be fulfilled; relational databases are good at this.

Drawbacks to using **file systems** to store data include the following:

- data *redundancy* and *inconsistency* (multiple formats, duplication of information, etc.)
- difficulties in *accessing* data
- data isolation (multiple files and formats)
- concurrency issues (among multiple users)
- *integrity* problems
- atomicity of updates
- security problems (hard to provide varied levels of user access)

There are varying ***levels of abstraction*** in a database. The ***physical level*** defines how a record is stored. The ***logical level*** describes data stored in the database and the relationships among data. The ***view level*** is a way to hide details of data types and information for security purposes.

The ***schema*** is the logical structure of the database; this is analagous to type information of a variable in a program. ***Physical schema*** refer to database design at the physical level. ***Logical schema*** refer to database design at the logical level. An ***instance*** is the actual content of the database at a particular time; this is analagous to the value of a variable. ***Physical data independence*** is the ability to modify the physical schema without changing the logical schema.

Data manipulation languages (DML) are languages for accesing and manipulating the data organized in a DBMS. ***Procedural languages*** are ones in which the user specifies what data is required and how to get that data. ***Declarative (nonprocedural) languages*** are ones in which the user specifies what data is required without specifying how to get such data. **SQL** is the most widely used query language.

A ***data definition language (DDL)*** is the specific notation for defining the database schema. The ***DDL compiler*** generates a set of tables stored in a

data dictionary. Data dictionary contains metadata.

A ***relational database*** is based on the relational data model. Data and relationships among the data are represented by a collection of tables. These include both a **DML** and **DDL**. The most common relational database systems employ the **SQL** query language.