CSI2132-2012

DGD 1:

Introduction

1

Exercise 1: Discuss the capabilities that should be provided by a DBMS

Refer to Chapter 1, slide 16 of the class notes.

- Uniform data administration.
- Reduced application development time.
- Data independence
- Efficient access
- Data integrity and security
- Concurrent access, recovery from crashes.
- Others
  - Web-based access
  - Distributed access

2

A University DB

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

3

Exercise 2: Identify some queries and updates you would run against this DB.

(a) (Query) List the names of all students majoring in Computer Science.

(b) (Query) What are the prerequisites of the Database course?

(c) (Query) Retrieve the transcript of Smith. This is a list of <CourseName, SectionIdentifier, Semester, Year, Grade> for each course section that Smith has completed.

(d) (Update) Insert a new student in the database whose Name=Jackson, StudentNumber=23, Class=1 (First year), and Major=MATH.

(e) (Update) Change the grade that Smith received in Intro to Computer Science section 119 to B.

4

Exercise 3:What is the difference between controlled and uncontrolled redundancy?

- Redundancy is when the same fact is stored multiple times in several places in a database.
  - For example, say the name of the student with StudentNumber=8 is Brown is stored multiple times.
- Redundancy is controlled when the DBMS ensures that multiple copies of the same data are consistent.
  - For example, if a new record with StudentNumber=8 is stored in the database, the DBMS will ensure that this record is for Student Brown.
- If the DBMS has no control over this, we have uncontrolled redundancy.

5

Exercise 4: Specify all the relationships among the records of the database shown in the figure.

(a) Each SECTION record is related to a COURSE record.

(b) Each GRADE_REPORT record is related to one STUDENT record and one SECTION record.

(c) Each PREREQUISITE record relates two COURSE records: one in the role of a course and the other in the role of a prerequisite to that course.

6

Exercise 5: Give some examples of integrity constraints that can apply to the database.

(a) The StudentNumber should be unique for each STUDENT record (key constraint).

(b) The CourseNumber should be unique for each COURSE record (key constraint).

c) A value of CourseNumber in a SECTION record must also exist in some COURSE record (referential integrity constraint).

(d) A value of StudentNumber in a GRADE_REPORT record must also exist in some STUDENT record (referential integrity constraint).

(e) The value of Grade in a GRADE_REPORT record must be one of the values in the set (A, B, C, D, F, I, U, S} (domain constraint).

(f) Every record in COURSE must have a value for CourseNumber (entity integrity constraint).

(g) A STUDENT record cannot have a value of Class=2 (sophomore) unless the student has completed a number of sections whose total course CreditHours is greater that 24 credits (general semantic integrity constraint).

7

Exercise 6a: What is logical data independence and why is it important?

- Users are shielded from changes in the (logical) relation(s) to be stored.
- For example, relation Student(SID, Name, GPA) is replaced by two relations: StudentNames(SID, Name) and StudentGPA(SID, GPA)
- Users are shielded from this change with a view Student(SID, Name, GPA) and application programs are thus unaware of this change

8

Exercise 6b: What is Physical Data Independence?

- Users are shielded from changes in the way that the relation is stored.
- For example, the Student() relation may be stored as
  - a file sorted on SID, or
  - as a heap (random order) file, or
  - as a file which is Indexed on GPA (more about indexes later in the course)

9

Say the name of the 'CS' (Computer Science) Department changes to 'CSSE' (Computer Science and Software Engineering) Department . The corresponding prefix for the course number also changes.
Exercise 7a: Identify the columns in the database that would need to be updated.

The following columns will need to be updated.

Table	Column(s)
STUDENT	Major
COURSE	CourseNumber and Department
SECTION	CourseNumber
PREREQUISITE	CourseNumber and PrerequisiteNumber

10

Exercise 7b: Restructure the columns in COURSE, SECTION, and PREREQUISITE tables so that only one column will need to be updated.

- You should split the following columns into two columns:

Table	Column	Columns
COURSE	CourseNumber	CourseDept and CourseNum
SECTION	CourseNumber	CourseDept and CourseNum
PREREQUISITE	CourseNumber	CourseDept and CourseNum
PREREQUISITE	PrerequisiteNumber	PreReqDept and PreReqNum

- Note that in the COURSE table, the column CourseDept will not be needed after the above change, since it is redundant with the Department column

11

Recent changes in privacy laws have disallowed organizations from using SINs to identify individuals unless certain restrictions are satisfied.
As a result, universities cannot use SINs as primary keys (except for financial data).
In practice, StudentID, a unique ID, a unique identifier, assigned to every student, is likely to be used as the primary key rather than SSN since StudentID is usable across all aspects of the system.

Some database designers are reluctant to use generated keys (also known as *surrogate* keys) for primary keys (such as StudentID) because they are artificial.

12

Exercise 8a: What are the advantages of using a surrogate key (an automatic key with no "meaning")?

- Has no "meaning" and thus no privacy violations
- Can use auto-numbers, which is usually fast
- Avoid that data is ambiguous, due to potential re-use of production keys

13

Exercise 8b: Can you propose any natural choices of keys that can be used to store the student record in a UNIVERSITY database?

- A combination of last name, first name and phone number(s) may be used
- However, what if the name and/or phone number changes?
- Also, if the student has more than one phone number, it would be better to create a separate relation *Phone(SIN, type, number)*

14

Exercise 9: Give examples of systems in which it may make sense to use traditional file processing instead of a database approach.

1. Small internal utility to locate files
2. Small single user application that does not require security (such as a customized calculator or a personal address and phone book)
3. Real-time navigation system (with heavy computation and very little data)

15

When Not to Use a DBMS

- More desirable to use regular files for:
  - Simple, well-defined database applications not expected to change at all
  - Stringent, real-time requirements that may not be met because of DBMS overhead
  - Embedded systems with limited storage capacity
  - No multiple-user access to data

Recall: Advantages of Using the DBMS Approach

- Controlling redundancy
- Restricting unauthorized access
  - Security and authorization subsystem
  - Privileged software
- Providing **persistent** storage for program objects
- Providing storage structures and search techniques for efficient query processing
  - **Indexes**
  - **Buffering** and **caching**
  - **Query processing and optimization**

Recall: Advantages of Using the DBMS Approach (cont'd.)

- Providing backup and recovery
  - **Backup and recovery subsystem** of the DBMS is responsible for recovery
- Providing multiple user interfaces
  - **Graphical user interfaces (GUIs)**
- Representing complex relationships among data
  - May include numerous varieties of data that are interrelated in many ways
- Enforcing **integrity constraints**

Next time

EER diagrams

19