COMP9120

Week 6: Database Integrity

Semester 1, 2022

Professor Athman Bouguettaya
School of Computer Science

THE UNIVERSITY OF SYDNEY

Contents developed by Dr Lijun Chang

Acknowledgement of Country

I would like to acknowledge the Traditional Owners of Australia and recognise their continuing connection to land, water and culture. I am currently on the land of the Darug people and pay my respects to their Elders, past, present and emerging.

I further acknowledge the Traditional Owners of the country on which you are on and pay respects to their Elders, past, present and future.

Copyright warning

Outline

- **Overview of Integrity Constraints**
- **Static Integrity Constraints**
  - Domain Constraints
  - Key / Referential Constraints
  - Semantic Integrity Constraints
  - Assertions
- **Dynamic Integrity Constraints**
  - Triggers

Integrity Constraints (ICs)

- **Integrity Constraint (IC)**:

a condition that must hold true for every instance of a database

- ICs are an integral part of the initial database schema design (through the **create table** command) to ensure the integrity and consistency of the database.
- ICs can be added/updated later (through **alter table** *table-name* **add** *constraint* command).
  - When such a command is executed, the system first ensures that the relation satisfies the specified constraint. If it does, the constraint is added to the relation; if not, the command is rejected.
- A **legal** instance of a relation is one that (*eventually*) satisfies all specified ICs.
  - Not *necessarily at all times.*

Integrity Constraints (ICs)

- Example integrity constraints
  - Each student ID must be unique.
  - No two lecturers can have the same ID.
  - Every school name in the *unit* relation must have a matching school name in the *school* relation.
  - For every student, a name must be given.
  - The only possible grades are either ‘F’, ‘P’, ‘C’, ‘D’, or ‘H’.
  - Valid lecturer titles are ‘Associate Lecturer’, ‘Lecturer’, ‘Senior Lecturer’, ‘Associate Professor’, or ‘Professor’.
  - Students can only enrol in the units of study that are currently on offer.
  - The sum of all marks in a course cannot be higher than 100.

Automatically Capturing ICs

- Why do we need to capture integrity constraints:
  - Data consistency (e.g., deleting an employee from the **Employee** table should also result in all corresponding tuples from the **Works-on** relation to be deleted).
  - Stored data is more faithful to the real-world meaning (semantics) of the domain application
    - Avoid data entry errors (e.g., inserting a *grade* into the **Student** table which does not exist).
  - Easier application development and better maintainability because ICs are centrally managed by the DBMS.
    - We do not have to worry about "**how**" integrity constraints are enforced/implemented.

Integrity Constraints in a Database

- ICs are *specified* as part of the database *schema design*
  - The database designer is responsible for ensuring that the integrity constraints are not contradicting each other!
    - Could be automated but may introduce unacceptable overhead.
- ICs are *checked* when the related parts of the database are *modified*
  - Can specify when ICs should be checked: after a SQL statement, or at the end of a 'transaction'
    - Transaction: a **group of statements** to be executed atomically (will later look at "ACID" properties of transactions later in the semester)
- Possible *reactions* if an IC is violated:
  - Reject database operation
  - Abort of the 'transaction' – rollback operations part of current 'transaction'
  - Execution of "maintenance" operations to make DB legal again

THE UNIVERSITY OF SYDNEY

An Informal Introduction to Transaction

- A group of statements to be executed **atomically**

  BEGIN;
  A group of SQL statements;
  COMMIT;

- A SQL statement usually starts with the following keywords and ends with a semicolon
  - **SELECT** - extracts data from a database
  - **UPDATE** - updates data in a database
  - **DELETE** - deletes data from a database
  - **INSERT INTO** - inserts new data into a database
  - **CREATE TABLE** - creates a new table
  - **ALTER TABLE** - modifies a table
  - **DROP TABLE** - deletes a table

- Consider an empty table

  R(<u>id</u>: integer, name: varchar(8))

  - What will be the result of the following transaction?

    BEGIN;
    INSERT INTO R **VALUES**(1, ‘Adam’);
    INSERT INTO R **VALUES**(1, ‘Smith’);
    COMMIT;

Types of Integrity Constraints

- **Static Integrity Constraints**
  describe conditions that every *legal instance* of a database must satisfy
  - Inserts / deletes / updates that violate ICs are disallowed
  - Four kinds:
    - *Domain Constraints*
    - *Key Constraints & Referential Integrity*
    - *Semantic Integrity Constraints*
    - *Assertions*

- **Dynamic Integrity Constraints**
  are predicates on database state changes
  - *Triggers*

Outline

- **Overview of Integrity Constraints**
- **Static Integrity Constraints**
  - Domain Constraints
  - Key / Referential Constraints
  - Semantic Integrity Constraints
  - Assertions
- **Dynamic Integrity Constraints**
  - Triggers

Domain Constraints

› Fields must be of the right data domain
  - always enforced for values inserted in the database
  - Also: queries are tested to ensure that the comparisons make sense.

› Most simply, each attribute needs to have a *data type*

› SQL DDL allows domains of attributes to be further restricted in the **CREATE TABLE** statement with the following clauses:
  - **DEFAULT** *default-value*
    default value for an attribute if its value is omitted in an insert statement.
  - **NOT NULL**
    attribute is not allowed to become NULL
  - **NULL**

    the values for an attribute may be NULL (which is the default)

Example of Domain Constraints

```
CREATE TABLE Student
(
    sid       INTEGER      NOT NULL,
    name      VARCHAR(20)  NOT NULL,
    semester  INTEGER      DEFAULT 1,
    birthday  DATE         NULL,
    country   VARCHAR(20)
);
```

Example:

```
INSERT INTO Student(sid,name) VALUES (123,’Peter');
```

Student				
sid	name	semester	birthday	country
123	Peter	1	null	null

Domain Check Constraints

› Limit the allowed values for an attribute by specifying extra conditions with an in-line check constraint

att-name sql-data-type **CHECK(***condition***)**

› Examples:

- Gender can be ‘male’ or ‘female’

  gender **VARCHAR**(6*)* **CHECK(** gender **IN** (‘male’, ‘female’) **)**

- Age must be positive

  age **INTEGER CHECK(** age >= 0 **)**

User-Defined Domains

› New domains can be created from existing data domains, with their own defaults and restrictions

```
CREATE DOMAIN domain-name sql-data-type ...
```

- Example:

```
CREATE DOMAIN Grade CHAR DEFAULT 'P' CHECK(VALUE IN
                                     ('F','P','C','D','H'))
```

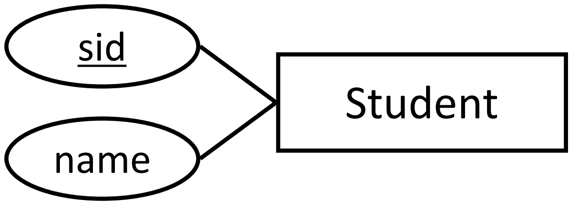
equivalent to:

```
CREATE TABLE Student (
   sid       INTEGER      NOT NULL,
   name      VARCHAR(20)  NOT NULL,
   grade     Grade,
   birthday  DATE
);
```

```
CREATE TABLE Student (
   sid       INTEGER      NOT NULL,
   name      VARCHAR(20)  NOT NULL,
   grade     CHAR DEFAULT 'P' CHECK(grade IN ('F','P','C','D','H')),
   birthday  DATE );
```

Key Constraints

- In SQL, we specify key constraints using the **PRIMARY KEY** and **UNIQUE** clauses:



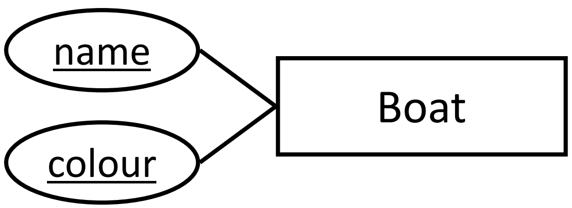

```
CREATE TABLE Student
(
  sid    INTEGER PRIMARY KEY,
  name VARCHAR(20)
);
```

- A primary key is automatically **UNIQUE** and **NOT NULL**
  - A relation can have multiple candidate (unique) keys, but only one primary key

Key Constraints

› Composite keys: a key consisting of multiple attributes

- Must be specified in a separate clause




```
CREATE TABLE Boat
(
  name   VARCHAR(20),
  colour  VARCHAR(20),
  PRIMARY KEY (name, colour)
);
```

```
CREATE TABLE Boat
(
  name   VARCHAR(20) PRIMARY KEY,
  colour VARCHAR(20) PRIMARY KEY
);
```

Foreign Keys and Referential Integrity

› **Foreign key:** set of attributes in a relation that is used to `refer' to a tuple in a parent/referred relation.

- Must refer to a **candidate key** of the parent (i.e., referred) relation

› **Referential Integrity**: for each tuple in the referring relation whose foreign key value is α, there must be a tuple in the referred relation whose value of the referred attribute is also α

- e.g. Enrolled(*sid*: integer, ucode: string, semester: string)
  sid is a foreign key referring to Student:
- If all foreign key constraints are enforced, referential integrity is achieved, i.e., no dangling references

Foreign Keys in SQL

- Only students listed in the Students relation should be allowed to enrol in Units of study.

```
CREATE TABLE Enrolled
(   sid CHAR(5),  uos CHAR(8),  grade VARCHAR(2),
    PRIMARY KEY (sid,uos),
    FOREIGN KEY (sid) REFERENCES Student,
    FOREIGN KEY (uos) REFERENCES Unitofstudy
);
```

Student

sid	name	age	country
53666	Jones	19	AUS
53650	Smith	21	AUS
54541	Ha Tschi	20	CHN
54672	Loman	20	AUS

Enrolled

sid	uos	grade
53666	COMP5138	CR
53666	INFO4990	CR
53650	COMP5138	P
53666	SOFT4200	D
~~54221~~	~~INFO4990~~	~~F~~

??? Dangling reference

Enforcing Referential Integrity in SQL

- SQL-92 and SQL-1999 support all 4 options on deletes and updates.
  - Default is **NO ACTION** (delete/update is rejected)
  - **CASCADE** (also delete/update all tuples that refer to deleted/updated tuple)
  - **SET NULL / SET DEFAULT** (sets foreign key value of referencing tuple)

```
CREATE TABLE Enrolled
(
  -- the sid field default
     value is 12345
  sid CHAR(5) DEFAULT 12345,
  uos CHAR(8),
  grade VARCHAR(2),

  PRIMARY KEY (sid,uos),

  FOREIGN KEY (sid) REFERENCES Student
  -- the on delete cascade conveys
     that an enrolled row should be
     deleted when the student with sid
     that it refers to is deleted
  ON DELETE CASCADE

  -- the on update set default
     will attempt to update the
     value of sid to a default value
     that is specified as the default
     in this Enrolled schema definition
  ON UPDATE SET DEFAULT
);
```

Short 5 mn break:

please stand up, stretch, and move around

Semantic Integrity Constraints (Table Constraints)

- Examples:
  - "Total marks are between 0 and 100"
  - "Only lecturers of a course can give marks for that course."

- Use SQL **CHECK** constraints, in-line like before, or as separate named constraints:

  CHECK (semantic-condition)

Example of Semantic Integrity Constraints

```
CREATE TABLE Assessment
(
    sid   INTEGER     REFERENCES Student,
    uos   VARCHAR(8) REFERENCES UnitOfStudy,
    mark  INTEGER,
    CHECK (mark BETWEEN 0 AND 100)
);
```

SQL: Naming Integrity Constraints

- The **CONSTRAINT** clause can be used to name all kinds of integrity constraints
- Example:

```
CREATE TABLE Enrolled
(
  sid      INTEGER,
  uos      VARCHAR(8),
  grade    VARCHAR(2),
  CONSTRAINT FK_sid_enrolled   FOREIGN KEY (sid)
                      REFERENCES Student
                      ON DELETE CASCADE,
  CONSTRAINT FK_cid_enrolled   FOREIGN KEY (uos)
                      REFERENCES UnitOfStudy
                      ON DELETE CASCADE,
  CONSTRAINT CK_grade_enrolled CHECK(grade IN ('F',…)),
  CONSTRAINT PK_enrolled       PRIMARY KEY (sid,uos)
);
```

Deferring Constraint Checking

- Any constraint - domain, key, foreign-key, semantic - may be declared:
  - **NOT DEFERRABLE**
    The default. It means that every time a database modification occurs to tuples that a DBMS sees as being related, the constraint is checked immediately afterwards.
  - **DEFERRABLE**
    Gives the option to wait until a transaction is complete before checking the constraint.
    - **INITIALLY DEFERRED** wait until transaction end,
      but allow to dynamically change later by

      SET CONSTRAINT name **IMMEDIATE**
    - **INITIALLY IMMEDIATE** check immediate,
      but allow to dynamically change later by

      SET CONSTRAINT name **DEFERRED**

Example of Deferring Constraint Checking

```
CREATE TABLE UnitOfStudy
(
  uos_code      VARCHAR(8),
  title         VARCHAR(20),
  lecturer      INTEGER,
  credit_points INTEGER,
  CONSTRAINT UoS_PK PRIMARY KEY (uos_code),
  CONSTRAINT UoS_FK FOREIGN KEY (lecturer)
    REFERENCES Lecturer DEFERRABLE INITIALLY DEFERRED
);
```

- Allows us to insert a new course referencing a lecturer that is not present at the time, but who will be added later *in the same transaction*.
- Behaviour can be dynamically changed within a transaction with the SQL statement

```
SET CONSTRAINT UoS_FK IMMEDIATE;
```

Add/Modify/Remove Integrity Constraints

› Integrity constraints can be added, modified (only domain constraints), and removed from an existing schema using **ALTER TABLE** statement

ALTER TABLE table-name constraint-modification

› where constraint-modification is one of:

ADD CONSTRAINT constraint-name new-constraint
DROP CONSTRAINT constraint-name
RENAME CONSTRAINT old-name **TO** new-name
ALTER COLUMN attribute-name domain-constraint

› Example (PostgreSQL syntax):

ALTER TABLE Enrolled **ALTER COLUMN** grade **TYPE VARCHAR**(3),
ALTER COLUMN mark **SET NOT NULL**;

› What happens if the existing data in a table does not fulfil a newly added constraint?

Then constraint doesn't get created!

e.g. "SQL Error: ORA-02296: cannot enable (USER.) - null values found"

Assertions

- The integrity constraints seen so far are associated with a single table
- Need for more general integrity constraints
  - E.g. integrity constraints over several tables

- **Assertion**: a predicate expressing a condition that we wish the database to always satisfy.
- SQL-92 syntax:
  CREATE ASSERTION assertion-name **CHECK** (condition)

- When an assertion is made, the system tests it for validity, and tests it again on every update that may violate it
  - This testing may introduce a significant amount of overhead; hence assertions should be used with great care.

Assertion Example

› The number of boats plus the number of sailors should be less than 100.

```
CREATE TABLE Sailors (
   sid    INTEGER,
   sname  CHAR(10),
   rating INTEGER,
   PRIMARY KEY (sid),
   CHECK (rating >=1 AND rating <=10),
   CHECK ((SELECT COUNT(s.sid) FROM Sailors s)
          + (SELECT COUNT(b.bid) FROM Boats b) < 100))
);
```

```
CREATE ASSERTION smallclub CHECK
(      (SELECT COUNT(s.sid) FROM Sailors s)
       + (SELECT COUNT(b.bid) FROM Boats b) < 100) );
```

*PostgreSql does not support ASSERTION
https://www.postgresql.org/docs/9.2/unsupported-features-sql-standard.html

*PostgreSql does not support subquery in CHECK
https://www.postgresql.org/docs/9.1/sql-createtable.html

THE UNIVERSITY OF SYDNEY

Outline

- Overview of Integrity Constraints

- Static Integrity Constraints
  - Domain Constraints
  - Key / Referential Constraints
  - Semantic Integrity Constraints
  - Assertions

- **Dynamic Integrity Constraints**
  - Triggers

Triggers

- A **trigger** is a section of code that is executed automatically if some specified modifications occur to the database AND a certain condition holds true.

- A trigger specification consists of three parts:

  ON *event* **IF** *condition* **THEN** *action*

  - *Event* (what activates the trigger)
  - *Condition* (test the condition's truth to determine whether to execute an action)
  - *Action* (what happens if the condition is true)

Why Triggers?

- Constraint maintenance
  - Triggers can be used to maintain foreign-key and semantic constraints; commonly used with ON DELETE and ON UPDATE
- Business rules
  - Some dynamic business rules can be encoded as triggers
- Monitoring
  - E.g. to react on the insertion of some sensor reading into db

Trigger Example (SQL:1999)

```
CREATE TRIGGER gradeUpgrade
  AFTER INSERT OR UPDATE ON Assessment
  BEGIN
    UPDATE Enrolled E
      SET grade='P'
     WHERE ( SELECT SUM(mark)
               FROM Assessment A
              WHERE A.sid=E.sid AND
                    A.uos=E.uosCode ) >= 50;
 END;
```

Triggering Events and Actions

- Triggering event can be **INSERT**, **DELETE** or **UPDATE**
- Triggers on update can be restricted to specific attributes

CREATE TRIGGER overdraft-trigger **AFTER UPDATE OF** balance
ON Account

- Values of attributes before and after an update can be referenced
  - **REFERENCING OLD ROW AS** name: for deletes and updates
  - **REFERENCING NEW ROW AS** name: for inserts and updates
  - In PostgreSQL: separate **OLD** and **NEW** variable automatically generated with a trigger function (PL/pgsql).

Trigger Granularity

- **Granularity**
  - *Row-level trigger*: A row-level trigger is fired for each row that needs to be updated.
  - *Statement-level trigger*: A statement-level trigger is fired for each SQL statement which may involve a set of rows that need to be updated.

- Statement-level trigger can be more efficient when dealing with SQL statements that update many rows…

Statement vs Row Level Trigger

- Example: Assume the following schema

  Employee(name, salary)

with *1000 tuples* and a **AFTER INSERT** *trigger* on salary…

- Now let us give employees a pay raise of 2.5%:

```
Create Trigger Pay_raise
  AFTER INSERT  OF salary ON Employee
  For Each Row
      UPDATE Employee SET salary = salary*1.025
```

Statement vs Row Level Trigger

- Update Costs:
  - How many rows are updated?
  - How often is a ***row-level*** trigger executed?
  - How often is a ***statement-level*** trigger executed?

Trigger levels - Syntax

- Instead of executing a separate action for each affected row, a single action can be executed for all rows affected by a statement

  - Use **FOR EACH STATEMENT** instead of **FOR EACH ROW** (actually, the default)

- Statement-level triggers can be more efficient when dealing with SQL statements that update many rows…

After Trigger Example (statement-level triggers)

Keep track of salary averages in the log

```
CREATE FUNCTION Salary_Average() RETURNS trigger AS $emp_stamp$
BEGIN
………….
END;

CREATE TRIGGER RecordNewAverage
  AFTER UPDATE OF Salary ON Employee
  FOR EACH STATEMENT
  BEGIN
    EXECUTE PROCEDURE Salary_Average()
  END;
```

Triggers

```
CREATE [OR REPLACE] TRIGGER trigger-name

  ⎧ BEFORE     ⎫ ⎧ INSERT           ⎫
  ⎨ AFTER      ⎬ ⎨ DELETE           ⎬  ON table-name
  ⎩ INSTEAD OF ⎭ ⎩ UPDATE OF attr   ⎭

  REFERENCING   ⎧ OLD ⎫  TABLE AS variable-name -- optional
                ⎩ NEW ⎭

  FOR EACH ROW              -- optional; otherwise, a statement trigger
  WHEN (condition)          -- optional
DECLARE
    <local variable declarations>
BEGIN
   <PL/SQL block>
END;
```

In PostgreSQL, this is replaced by a trigger procedure

PostgreSQL trigger procedure: https://www.postgresql.org/docs/9.5/static/plpgsql-trigger.html

Some Tips on Triggers

- Use BEFORE triggers
  - Usually for checking integrity constraints
- Use AFTER triggers
  - Usually for integrity maintenance and update propagation

- Good overviews:
  - Ramakrishnan – Brief overview Section 5.8, 5.9
  - Kifer/Bernstein/Lewis: “Database Systems - An Application-oriented Approach”, 2nd edition, Chapter 7.
  - Michael v. Mannino: “Database - Design, Application Development and Administration”

You Should Now be Able To

- Capture Integrity Constraints in an SQL Schema
  - Including key constraints, referential integrity, domain constraints and semantic constraints
- Formulate complex semantic constraints using Assertions
- Know when to use Assertions, and CHECK constraints
- Know the semantic of deferring integrity constraints
- Be able to formulate simple triggers
  - Know the difference between row-level & statement-level triggers

References

- Ramakrishnan/Gehrke (3rd edition - the 'Cow' book)
  - **Sections 3.2-3.3 and Sections 5.7-5.9**
  - *Integrity constraints are covered in different parts of the SQL discussion; only brief on triggers*
- Kifer/Bernstein/Lewis (2nd edition)
  - Sections 3.2.2-3.3 and Chapter 7
  - *Integrity constraints are covered as part of the relational model, but a good dedicated chapter (Chap 7) on triggers*
- Ullman/Widom (3rd edition)
  - Chapter 7
  - *Has a complete chapter dedicated to both integrity constraints&triggers. Good.*
- Michael v.Mannino: "Database - Design, Application Development and Administration"
  - *Include a good introduction to triggers.*

Next Week

- **Transaction Management**
  - Transaction Concept
  - Serializability

- Readings:
  - **Ramakrishnan/Gehrke (Cow book), Chapter 16**
  - Kifer/Bernstein/Lewis book, Chapter 18
  - Ullman/Widom, Chapter 6.6 onwards

See you next week!