# CSCI 4333 Section 2 Design of DB Systems

## 3/25/2024 (self - annotation)

**Introduction to SQL and MySQL**

by K. Yue

**1. Introduction**

* SQL (Structured Query Language): defacto standard for relational databases.
* SQL-like languages are also used in non-relational DBMS.
* Contains core specifications and extensions. Latest SQL standard: 2016.
* Not *pure* relational model: e.g.
  1. Use the terms row, column, and table instead of tuple, attribute and relation.
  2. The results may not be a set.
* Mostly based on Tuple Relational Calculus (TRC) and a little on Relational Algebra (RA).
* SQL is mostly declarative.
* DBMS vendor-specific extensions are common.
* SQL Contains:
  1. Data Definition Language (DDL): define the relation schema (structure)
  2. Data Manipulation Language (DML): manipulate data; CRUD:
     1. Create: Insert
     2. Read
     3. Update
     4. Delete
  3. Data Administration Language: for DB administration such as user and security management.

**2. MySQL**

Toyu: A drastically simplified university.

UML Diagram: [toyu\_uml.jpg](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/model/toyu_uml.jpg)

MS Access: [toyu.accdb](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/access/toyu.accdb)

MySQL creation script: [Createtoyu.sql.txt](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/query/Createtoyu.sql.txt).

**2.1 DDL:**

* Make sure that you are familiar with the core SQL Data Definition Language (DDL) commands. Refer to, for example: <http://www.w3schools.com/sql/default.asp>.
* MariaDB DDL: <https://mariadb.com/kb/en/sql-statements/>
* MySQL DDL manual: <https://dev.mysql.com/doc/refman/8.1/en/sql-statements.html>.
* Basic DDL: some examples
  1. CREATE TABLE
  2. CREATE DATABASE: a database contains a collection of related tables for an application.
  3. CREATE VIEW: a view is a virtual table for users to access a subset of a database.
  4. CREATE INDEX: an index is a data structure to enhance access performance of specific queries.
  5. CREATE PROCEDURE
  6. CREATE FUNCTION
  7. CREATE TRIGGER: a trigger is event-driven procedural code activated by events.
  8. ALTER (No ALTER TRIGGER and ALTER INDEX)
  9. DROP

BNF:

* Note that in MySQL, DATABASE and SCHEMA having the same meaning. Hierarchy:
  1. MySQL: Database = Schema: contains a collection of tables.
  2. Postgres:
     + A database contains a collection of schema
     + A schema contains a collection of tables.
* Constraints: to implement certain constraints in your data model.
  1. NOT NULL: attributes cannot have an null value.
  2. UNIQUE: KEY; the set of attributes must be unique for each row:
  3. PRIMARY KEY: unique, not null, and used for the physical structure of the relation.
  4. FOREIGN KEY
  5. CHECK: for a Boolean condition on the columns.
  6. DEFAULT: define a default value.
* Some other options:
  1. AUTO INCREMENT: automatic increment an integer if a value is not specified. Used for id.

***Example:***

[Createtoyu.sql.txt](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/query/Createtoyu.sql.txt).

***Example:***

Experimenting with MySQL Create Table commands. Execute the following code and ensure that you understand the result. For example,

* A temporary table is not persistent. It is created for a SQL client session. Its scope is the client session.

CREATE TABLE s2  
SELECT \* FROM student;  
  
SELECT \*  
FROM s2;  
  
CREATE TEMPORARY TABLE s3  
SELECT \* FROM student;  
  
SELECT \*  
FROM s3;  
  
CREATE TABLE s4 LIKE student;  
  
SELECT \*  
FROM s4;  
  
INSERT INTO s4  
SELECT \* FROM student;  
  
SELECT \*  
FROM s4;  
  
SHOW TABLES;  
  
-- Note that keys and constraints of student are missing in s2 and S3.  
DESC student;  
DESC s2;  
DESC s3;  
DESC s4;  
  
DROP TABLE s2;  
DROP TABLE s3;  
DROP TABLE s4;  
  
SHOW TABLES;

Column names may include special characters. For example, you cannot use the name 'first name' directly as column name, as spaces are interpreted as separator. You will need special syntax. For example:

1. In MySQL, use back-quote: `first name`
2. In MS SQL Server, use []: [first name]

* For each column, there is a data type and optional specifiers (such as NULL, NOT NULL, default values, etc.)
* Additional constraints and indexes can be defined.
* In general, some important considerations in creating tables:
  1. What are the columns?
  2. What are the data types of the columns?
     + The right domain: be restrictive.
     + Performance consideration.
  3. Nullability of columns
  4. Primary key
  5. Candidate keys
  6. Foreign keys: can they be enforced for the selected storage engine?
  7. Indexes: performance tuning.
  8. Additional constraints: check whether they are enforced by the DBMS.

***Example:***

* MySQL only supports foreign key constraint in the InnoDB database engine.
* Older versions of MySQL ignore the 'check' clause.

**2.2 Data types**

* Data types in MySQL are rich: <https://dev.mysql.com/doc/refman/8.1/en/data-types.html>
* Beside simple data types, two noticeable data types:
  1. JSON
  2. Spatial

**3. DML**

* Basically declarative.

**3.1 Writing to the DB**

* Basic *update* commands (write):
  1. INSERT
  2. UPDATE
  3. DELETE

INSERT INTO <<table>> [<<columns>>]  
VALUES <<expression>>

* If column names are missing, the proper column order during table creation will be used.
* Column names using default values or auto-increment values should not be included in the INSERT statement if they are used.
* NULL and DEFAULT can be used as values in INSERT.
* One may also insert values from a select statement. E.g.

INSERT INTO <<table>> [<<columns>>]  
<<select statement>>

* Delete statement includes a condition for selecting the rows for deletion.

DELETE FROM <<table>>  
WHERE <<condition>>

* The update statement is used to update rows and can have an update condition to identify the rows to be updated.

UPDATE <<table>>  
SET <<update assignments>>  
[WHERE <<update condition>>]

***Example:***

UPDATE Student  
SET major = 'ITEC'  
WHERE StuId = 100000;

* The update and delete statements can be used to affect multiple rows so be very careful.

***Example:***

-- All students will be majoring in CSCI  
UPDATE Student  
SET major = 'CSCI';

* Once changed, the effect is permanent. There is no 'undo' command.

***Example:***

Note the order of the insertions in createtoyu.sql below.

INSERT INTO Grade(grade, gradePoint) VALUES  
    ('A',4),('A-',3.6667),('B+',3.3333),('B',3),('B-',2.6667),  
    ('C+',2.3333),('C',2),('C-',1.6667),  
    ('D+',1.3333),('D',1),('D-',0.6667),('F',0),  
   ('P', NULL), ('IP', NULL), ('WX', NULL);  
     
INSERT INTO School(schoolCode, schoolName) VALUES  
    ('BUS','Business'),  
   ('EDU','Education'),  
    ('HSH','Human Sciences and Humanities'),  
    ('CSE','Science and Engineering');  
  
INSERT INTO Department(deptCode, deptName, schoolCode, numStaff) VALUES  
    ('ACCT','Accounting','BUS',10),  
    ('ARTS','Arts','HSH',5),  
    ('CINF','Computer Information Systems','CSE',5),  
    ('CSCI','Computer Science','CSE',12),  
    ('ENGL','English','HSH',12),  
    ('ITEC','Information Technology','CSE',4),  
    ('MATH','Mathematics','CSE',7);  
  
INSERT INTO Faculty(facId, fname, lname, deptCode, `rank`) VALUES  
    (1011,'Paul','Smith','CSCI','Professor'),  
    (1012,'Mary','Tran','CSCI','Associate Professor'),  
    (1013,'David','Love','CSCI',NULL),  
    (1014,'Sharon','Mannes','CSCI','Assistant Professor'),  
    (1015,'Daniel','Kim','CINF','Professor'),  
    (1016,'Andrew','Byre','CINF','Associate Professor'),  
    (1017,'Deborah','Gump','ITEC','Professor'),  
    (1018,'Art','Allister','ARTS','Assistant Professor'),  
    (1019,'Benjamin','Yu','ITEC','Lecturer'),  
    (1020,'Katrina','Bajaj','ENGL','Lecturer'),  
    (1021,'Jorginlo','Neymar','ACCT','Assistant Professor');  
  
INSERT INTO Course(courseId, rubric, number, title, credits) VALUES  
    (2000,'CSCI',3333,'Data Structures',3),  
    (2001,'CSCI',4333,'Design of Database Systems',3),  
    (2002,'CSCI',5333,'DBMS',3),  
    (2020,'CINF',3321,'Introduction to Information Systems',3),  
    (2021,'CINF',4320,'Web Application Development',3),  
    (2040,'ITEC',3335,'Database Development',3),  
    (2041,'ITEC',3312,'Introduction to Scripting',3),  
    (2060,'ENGL',1410,'English I',4),  
    (2061,'ENGL',1311,'English II',3),  
    (2080,'ARTS',3311,'Hindu Arts',3),  
    (2090,'ACCT',3333,'Managerial Accounting',3);  
  
INSERT INTO Class(classId, courseId, semester, year, facId, room) VALUES  
    (10000,2000,'Fall',2019,1011,'D241'),  
    (10001,2001,'Fall',2019,1011,'D242'),  
    (10002,2002,'Fall',2019,1012,'D136'),  
    (10003,2020,'Fall',2019,1014,'D241'),  
    (10004,2021,'Fall',2019,1014,'D241'),  
    (10005,2040,'Fall',2019,1015,'D237'),  
    (10006,2041,'Fall',2019,1019,'D217'),  
    (10007,2060,'Fall',2019,1020,'B101'),  
    (10008,2080,'Fall',2019,1018,'D241'),  
    (11000,2000,'Spring',2020,1011,'D241'),  
    (11001,2001,'Spring',2020,1012,'D242'),  
    (11002,2002,'Spring',2020,1013,'D136'),  
    (11003,2020,'Spring',2020,1016,'D217'),  
    (11004,2061,'Spring',2020,1018,'B101');  
  
INSERT INTO Student(stuId, fname, lname, major, minor, ach, advisor) VALUES  
    (100000,'Tony','Hawk','CSCI','CINF',40,1011),  
    (100001,'Mary','Hawk','CSCI','CINF',35,1011),  
    (100002,'David','Hawk','CSCI','ITEC',66,1012),  
    (100003,'Catherine','Lim','ITEC','CINF',20,NULL),  
    (100004,'Larry','Johnson','ITEC',NULL,66,1017),  
    (100005,'Linda','Johnson','CINF','ENGL',13,1015),  
    (100006,'Lillian','Johnson','CINF','ITEC',18,1016),  
    (100007,'Ben','Zico',NULL,NULL,16,NULL),  
    (100008,'Bill','Ching','ARTS',NULL,90,NULL),  
    (100009,'Linda','King','ARTS','CSCI',125,1018),  
   (100111,'Cathy','Johanson',NULL,NULL,0,1018);  
     
INSERT INTO Enroll(stuId, classId, grade, n\_alerts) VALUES  
    (100000,10000,'A',0),  
    (100001,10000,NULL,NULL),  
    (100002,10000,'B-',3),  
    (100000,10001,'A',2),  
    (100001,10001,'A-',0),  
    (100000,10002,'B+',1),  
    (100002,10002,'B+',2),  
    (100000,10003,'C',0),  
    (100002,10003,'D',4),  
    (100004,10003,'A',0),  
    (100005,10003,NULL,NULL),  
    (100000,10004,'A-',1),  
    (100004,10004,'B+',NULL),  
    (100005,10004,'A-',0),  
    (100006,10004,'C+',NULL),  
    (100005,10005,'A-',0),  
    (100006,10005,'A',NULL),  
    (100005,10006,'B+',NULL),  
    (100007,10007,'F',4),  
    (100008,10007,'C-',0),  
    (100007,10008,'A-',0),  
    (100000,11001,'D',4);

Note the explicit use of null, which is a keyword in SQL.

***Example:***

Execute the following code and ensure that you understand the result.

INSERT INTO student VALUES  
   (100010,'Bun','Yue',null,null,50,null),  
   (100011,'Paul','Harris','CSCI','ITEC',23,1015);  
    
SELECT \* FROM student;  
    
INSERT INTO student VALUES  
   (100010,'Bun','Yue',null,null,50,null),  
   (100011,'Paul','Harris','CSCI','ITEC',23,1015);  
    
INSERT INTO student VALUES    
   (100020,'Bunno','Yue','GEOG',null,50,null);  
INSERT INTO student VALUES    
   (100021,'Bunna','Yue',null,'GEOG',50,null);  
INSERT INTO student VALUES    
   (100022,'Bunno','Yue',null,null,50,8888);  
  
    
-- Remove the two new rows.  
DELETE FROM Student  
WHERE stuId = 100010 OR stuId = 100011;  
  
SELECT \* FROM student;

**3.2 Querying with the Select Statement**

* Basic data retrieval statement in SQL
* Not to be confused with the select statement in Relational Algebra (RA).
* Basic format, with*conceptual* steps.

SELECT DISTINCT <<result\_columns>> -- [3] construct result columns  
FROM <<source\_tables>> -- [1] conceptually join sources to form a large table  
WHERE <<conditions\_for\_inclusion>> -- [2] Filter rows from [1]

1. <<source\_tables>>: the source tables to gather the result data
2. <<conditions\_for\_inclusion>>: the conditions to be satisfied for results to be included and possibly the conditions how the tables should be joined together.
3. <<result\_columns>>: the result columns or expressions desired to be displayed.

* Built-in functions and operators: <https://dev.mysql.com/doc/refman/8.1/en/built-in-function-reference.html>
* Some common functions:
  1. BETWEEN lower\_range AND upper\_range
  2. IN: membership test for a set/table (binary operation)
  3. EXISTS: not an empty set (unary operation)
  4. IF: ternary operation
  5. LIKE: inexact string matching.
     + wild cards:
       - % match any and all following characters.
       - \_: match any one character.

***Example:***

Execute the following code and ensure that you understand the result.

-- operators:  
-- student with credits in a range.  
SELECT DISTINCT \*  
FROM Student  
WHERE credits BETWEEN 30 AND 70;  
  
-- student in selected majors  
SELECT DISTINCT \*  
FROM Student  
WHERE major IN ('CSCI', 'CINF', 'ITEC');  
  
-- student enrolled in some classes.  
SELECT DISTINCT \*  
FROM Student AS s  
WHERE EXISTS  
(SELECT \*  -- a subquery  
FROM Enroll AS e  
WHERE e.stuId = s.stuId);  
-- or  
SELECT DISTINCT s.\*  
FROM Student AS s INNER JOIN Enroll AS e USING (stuId);  
  
-- students not enrolled in any class.  
SELECT DISTINCT \*  
FROM student AS s  
WHERE s.stuId NOT IN (SELECT DISTINCT e.stuID FROM enroll AS e);  
  
-- students wiht a 'k' in their last name.  
SELECT DISTINCT s.\*  
FROM student AS s  
WHERE s.lname LIKE '%k%';  
  
-- case sensitive version.  
SELECT DISTINCT s.\*  
FROM student AS s  
WHERE s.lname LIKE BINARY '%k%';

-- case sensitive version: a more complicated take.  
-- The mysql client sends the query using cp850.  
-- The default character set of MySQL server is utf8mb4.  
-- It is thus necessary to set the @@character\_set\_connection  
-- in order to use collate if MySQL client is used.  
-- If HeidiSQL is used, it is not necessary.  
SET @@character\_set\_connection=utf8mb4;

SELECT DISTINCT s.\*  
FROM student AS s  
WHERE s.lname LIKE '%k%' COLLATE utf8mb4\_bin;  
  
-- LIKE compares the whole string.  
SELECT DISTINCT s.\*  
FROM student AS s  
WHERE s.lname LIKE 'ng';  
  
-- student with last name of four characters, with ng the last two.  
SELECT DISTINCT s.\*  
FROM student AS s  
WHERE s.lname LIKE '\_\_ng';

**3.3 Join**

* When multiple tables are needed for a query, it is common that foreign keys are used to connect the tables.
* It is thus necessary to ensure that the equality of the foreign key with the referenced key of the parent table.
* A popular style is shown in the example below.

***Example:*** one popular SQL style

SELECT DISTINCT s.fname, s.lname, c.classId, e.grade  
FROM student AS s, enroll AS e, class AS c  
WHERE s.stuId = e.stuId -- Join condition  
AND e.classId = c.classId -- Join condition  
AND c.semester = 'Fall' -- problem condition  
AND c.year = 2019; -- problem condition

Result:

mysql> SELECT DISTINCT s.fname, s.lname, c.classId, e.grade  
    -> FROM student s, enroll e, class c  
    -> WHERE s.stuId = e.stuId   -- Join condition  
    -> AND e.classId = c.classId -- Join condition  
    -> AND c.semester = 'Fall'   -- problem condition  
    -> AND c.year = 2019;        -- problem condition  
+---------+---------+---------+-------+  
| fname   | lname   | classId | grade |  
+---------+---------+---------+-------+  
| Tony    | Hawk    |   10000 | A     |  
| Mary    | Hawk    |   10000 | NULL  |  
| David   | Hawk    |   10000 | B-    |  
| Tony    | Hawk    |   10001 | A     |  
| Mary    | Hawk    |   10001 | A-    |  
| Tony    | Hawk    |   10002 | B+    |  
| David   | Hawk    |   10002 | B+    |  
| Tony    | Hawk    |   10003 | C     |  
| David   | Hawk    |   10003 | D     |  
| Larry   | Johnson |   10003 | A     |  
| Linda   | Johnson |   10003 | NULL  |  
| Tony    | Hawk    |   10004 | A-    |  
| Larry   | Johnson |   10004 | B+    |  
| Linda   | Johnson |   10004 | A-    |  
| Lillian | Johnson |   10004 | C+    |  
| Linda   | Johnson |   10005 | A-    |  
| Lillian | Johnson |   10005 | A     |  
| Linda   | Johnson |   10006 | B+    |  
| Ben     | Zico    |   10007 | F     |  
| Bill    | Ching   |   10007 | C-    |  
| Ben     | Zico    |   10008 | A-    |  
+---------+---------+---------+-------+  
21 rows in set (0.00 sec)

**3.3.1 Inner Join**

* In the Select statement, the FROM clause allows the results of join statement in the table references.
* Using the JOIN operations in the FROM clause is the preferred technique:
  1. Potentially faster performance: better optimization by DB engines, especially when using indexes.
  2. Better style: separation of join conditions and query semantic conditions.
  3. Easier changes between different joins.
* There are many kind of joins, as discussed below.
* You may use the Explain statement in MySQL to find out the execution plan.

***Example:***

Execute the following code and ensure that you understand the result.

SELECT DISTINCT s.fname, s.lname, c.classId, e.grade  
FROM student AS s, enroll AS e, class AS c  
WHERE s.stuId = e.stuId -- Join condition  
AND e.classId = c.classId -- Join condition  
AND c.semester = 'Fall' -- problem condition  
AND c.year = 2019; -- problem condition  
  
SELECT DISTINCT s.fname, s.lname, c.classId, e.grade  
FROM student AS s INNER JOIN enroll e ON (s.stuId = e.stuId) -- Join condition  
    INNER JOIN class AS c ON (e.classId = c.classId) -- Join condition  
WHERE c.semester = 'Fall' -- Problem condition  
AND c.year = 2019; -- Problem condition  
  
-- alternative: using the USING clause.  
SELECT DISTINCT s.fname, s.lname, c.classId, e.grade  
FROM student AS s INNER JOIN enroll e USING (stuId) -- Join condition  
    INNER JOIN class AS c USING (classId) -- Join condition  
WHERE c.semester = 'Fall' -- Problem condition  
AND c.year = 2019; -- Problem condition  
  
-- the ON clause is more general and can be more effective.  
SELECT DISTINCT s.fname, s.lname, c.classId, e.grade  
FROM student AS s INNER JOIN enroll e ON (s.stuId = e.stuId) -- Join condition  
    INNER JOIN class AS c  
   ON (e.classId = c.classId -- Join condition  
      AND c.semester = 'Fall' -- Problem condition  
      AND c.year = 2019); -- Problem condition  
  
**3.3.2 Left and Right Join**

* Left joins are the most popular joins besides (inner) joins.
* R1 Left Join R2: same as inner Join, except that for a tuple t1 in R1 without a matching tuple in R2, t1 will be kept in the result with attributes from R2 being null.
  + All rows in the left table will be in the result at least once.
* A right join is the mirror image of a left join.

***Example***

Execute the following code and ensure that you understand the result.

-- List the names of the students with their minors (in full name).  
-- Student with no department not listed.  
SELECT DISTINCT CONCAT(s.fname, ' ', s.lname) AS student,  
    d.deptName AS `minor department`  
FROM student AS s INNER JOIN department AS d ON (s.minor = d.deptCode);  
  
-- List the names of the students with their minors (in full name).  
SELECT DISTINCT CONCAT(s.fname, ' ', s.lname) AS student,  
    d.deptName AS `minor department`  
FROM student AS s LEFT JOIN department AS d ON (s.minor = d.deptCode);  
  
-- List the names of the students with their minors (in full name).  
-- more readable form.  
SELECT DISTINCT CONCAT(s.fname, ' ', s.lname) AS student,  
    IFNULL (d.deptName, 'N/A') AS `minor department`  
FROM student s LEFT JOIN department d ON (s.minor = d.deptCode);

* Joins are procedural. Join orders can be important. Use parenthesis to enforce the desired order.

***Example:***

(R1 LEFT JOIN R2) RIGHT JOIN R3  
-- may give different result than  
R1 LEFT JOIN (R2 RIGHT JOIN R3)

***Example:***

Problem: List student information and the CSCI class information. Include all students, leaving blanks when appropriate  
(i.e., no CSCI courses enrolled by the student).

+--------+-----------------+---------+-------------+-------+  
| stuId  | student         | classId | CSCI course | grade |  
+--------+-----------------+---------+-------------+-------+  
| 100000 | Tony Hawk       | 10000   | CSCI 3333   | A     |  
| 100000 | Tony Hawk       | 10001   | CSCI 4333   | A     |  
| 100000 | Tony Hawk       | 10002   | CSCI 5333   | B+    |  
| 100000 | Tony Hawk       | 11001   | CSCI 4333   | D     |  
| 100001 | Mary Hawk       | 10000   | CSCI 3333   |       |  
| 100001 | Mary Hawk       | 10001   | CSCI 4333   | A-    |  
| 100002 | David Hawk      | 10000   | CSCI 3333   | B-    |  
| 100002 | David Hawk      | 10002   | CSCI 5333   | B+    |  
| 100003 | Catherine Lim   |         |             |       |  
| 100004 | Larry Johnson   |         |             |       |  
| 100005 | Linda Johnson   |         |             |       |  
| 100006 | Lillian Johnson |         |             |       |  
| 100007 | Ben Zico        |         |             |       |  
| 100008 | Bill Ching      |         |             |       |  
| 100009 | Linda King      |         |             |       |  
| 100111 | Cathy Johanson  |         |             |       |  
+--------+-----------------+---------+-------------+-------+  
16 rows in set (0.001 sec)

Execute the following code and ensure that you understand the result.

-- List student information and the CSCI class information.  
SELECT DISTINCT s.stuId,  
   CONCAT(s.fname, ' ', s.lname) AS student,  
   e.classId,  
   CONCAT(co.rubric, ' ', co.number) AS `CSCI course`,  
   e.grade  
FROM student AS s INNER JOIN enroll AS e USING (stuId)  
   INNER JOIN class AS c USING (classId)  
   INNER JOIN course AS co USING (courseId)  
WHERE co.rubric = 'CSCI';  
  
-- List student information and the CSCI class information.  
-- Include all students, leaving blanks when appropriate  
-- (i.e. no CSCI courses enrolled by the student).  
  
-- These do not do the job. Why?  
SELECT DISTINCT s.stuId,  
   CONCAT(s.fname, ' ', s.lname) AS student,  
   IFNULL(e.classId, '') AS classId,  
   IFNULL(CONCAT(co.rubric, ' ', co.number), '') AS `CSCI course`,  
   IFNULL(e.grade, '') AS grade  
FROM student AS s LEFT JOIN enroll AS e USING (stuId)  
   LEFT JOIN class AS c USING (classId)  
   LEFT JOIN course AS co USING (courseId)  
WHERE co.rubric = 'CSCI';  
  
SELECT DISTINCT s.stuId,  
   CONCAT(s.fname, ' ', s.lname) AS student,  
   IFNULL(e.classId, '') AS classId,  
   IFNULL(CONCAT(co.rubric, ' ', co.number), '') AS `CSCI course`,  
   IFNULL(e.grade, '') AS grade  
FROM student AS s LEFT JOIN enroll AS e USING (stuId)  
   LEFT JOIN class AS c USING (classId)  
   LEFT JOIN course AS co ON (c.courseId = co.courseId AND co.rubric = 'CSCI' );  
  
-- This works. Note the LEFT JOIN and RIGHT JOIN.  
SELECT DISTINCT s.stuId,  
   CONCAT(s.fname, ' ', s.lname) AS student,  
   IFNULL(e.classId, '') AS classId,  
   IFNULL(CONCAT(co.rubric, ' ', co.number), '') AS `CSCI course`,  
   IFNULL(e.grade, '') AS grade  
FROM enroll AS e INNER JOIN class AS c USING (classId)  
   INNER JOIN course AS co ON (c.courseId = co.courseId AND co.rubric = 'CSCI' )  
   RIGHT JOIN student AS s USING (stuId);

**3.4 Subqueries**

* A SQL subquery is a nested/inner subquery within a SQL statement or another query (for SELECT, INSERT, UPDATE or, DELETE).
* Subqueries usually appear in the FROM clause (as derived tables) and the WHERE clause.

***Example***

Execute the following code and ensure that you understand the result.

-- subqueries in the WHERE course  
-- students not enrolled in any class.  
SELECT DISTINCT \*  
FROM student AS s  
WHERE s.stuId NOT IN (SELECT DISTINCT e.stuID FROM enroll AS e);  
  
-- student with the maximum number of ach.  
SELECT DISTINCT MAX(ach)  
FROM student;  
  
-- student within 60 credits of the maximum number of ach any student may have.  
SELECT DISTINCT s.stuId,  
   CONCAT(s.fname, ' ', s.lname) AS student,  
   s.ach AS credits  
FROM student AS s  
WHERE s.ach + 60 >=  
   (SELECT DISTINCT MAX(ach) FROM student);  
  
-- subqueries as derived tables.  
SELECT DISTINCT s.stuId,  
   CONCAT(s.fname, ' ', s.lname) AS student,  
   s.ach AS credits  
FROM student AS s INNER JOIN  
   (SELECT DISTINCT MAX(ach) AS max FROM student) AS m -- an alias is required.  
WHERE s.ach + 60 >= m.max;  
  
**3.5 Common Table Expressions (CTE)**

* Supported by MySQL 8.0.
* Allow the definition of temporary common tables in a sequence before the body of a SELECT statement.
  + WITH t1 AS (definition of t1, a query...), t2 AS (...), ..., tn AS () SELECT ...
* A table defined in CTE can be used immediately until the end of the SELECT statement.
* Support a more natural way to implement *algorithmic solutions*, an (n+1) step solutions.
  + step 1 to n: constructions of t1, t2, ..., tn
  + step (n+1): the body of the SELECT statement.
* Allow recursion.
* May degrade performance.
* It is generally better than subqueries in the FROM clauses.
  + Tables in CTE can be used immediately after their definitions.
  + More natural order.
  + Can use recursion.

***Example:***

-- CTE  
WITH  t1 AS  
   (SELECT MAX(ach) AS max FROM student)  
SELECT s.stuId,  
   s.ach AS `ach credits`,  
   t1.max - s.ach AS `diff from max credits of all`  
FROM student AS s, t1  
ORDER BY `ach credits` DESC;  
  
-- multiple common tables (not efficient; used as demonstration.)  
WITH t1 AS  
   (SELECT MAX(ach) AS max FROM student),  
t2 AS  
   (SELECT s.stuId,  
      s.ach AS `ach credits`,  
      t1.max - s.ach AS diff,  
      s.major  
    FROM student AS s, t1)  
SELECT t2.stuId, t2.`ach credits`,  
   t2.diff AS `diff from max credits of all`,  
   d.deptName AS department  
FROM t2 LEFT JOIN department d ON (t2.major = d.deptCode)  
ORDER BY t2.`ach credits` DESC;

For those interesting in recursive CTE, here is an example. Recursive CTE will not be in the examinations.

Create and populate a simple relation that stores EmpId of an employee and the EmpId of the immediate supervisor.

CREATE SCHEMA CTETinker;  
USE SCHEMA CTEtinker;  
CREATE OR REPLACE TABLE Employee (  
    EmpId CHAR(7) NOT NULL,  
    SupervisorEmpId CHAR(7) NULL,  
    CONSTRAINT Emp\_EmpId\_pk PRIMARY KEY (EmpId),  
    CONSTRAINT Emp\_SupervisorEmpId\_fk FOREIGN KEY (SupervisorEmpId)  
        REFERENCES Employee(EmpId)  
);  
  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E3', null);  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E15', 'E3');  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E50', 'E15');  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E75', 'E50');  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E100', 'E75');  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E102', 'E75');  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E70', 'E50');  
INSERT INTO Employee(EmpId, SupervisorEmpId) VALUES  
   ('E103', 'E70');  
    
SELECT \* FROM Employee;  
  
Result:

MariaDB [temp]> SELECT \* FROM Employee;  
+-------+-----------------+  
| EmpId | SupervisorEmpId |  
+-------+-----------------+  
| E3    | NULL            |  
| E50   | E15             |  
| E15   | E3              |  
| E70   | E50             |  
| E75   | E50             |  
| E103  | E70             |  
| E100  | E75             |  
| E102  | E75             |  
+-------+-----------------+  
8 rows in set (0.002 sec)  
   
A recursive CTE SQL to get all supervisors of employee 'E100':

WITH RECURSIVE Super(SEId) AS  
(  SELECT SupervisorEmpId AS SEId FROM Employee AS e WHERE e.EmpId = 'E100' -- initial condition/action  
   UNION ALL -- union all: add rows created by the recursive action to the result, table Super.  
   SELECT e.SupervisorEmpId AS SEId -- recursive action  
      FROM Employee AS e INNER JOIN Super  
      WHERE e.EmpId = Super.SEId  
      AND e.SupervisorEmpId IS NOT NULL  
      -- exit condition: when the recursive action returns an empty table.  
)  
SELECT \*  
FROM Super;

Result:

+------+  
| SEId |  
+------+  
| E75  |  
| E50  |  
| E15  |  
| E3   |  
+------+

DROP SCHEMA IF EXISTS CTEtinker;

**3.6 Group by and having**

* Useful for group/summary reports: one result row per group, not per row as in the regular select statement.
* Allow aggregate functions (also known as *group functions* and column functions) to be performed by the groups defined.
* Output one row per group.
* A group is defined by an unique value of the columns in the group by clause.
* Example aggregate functions: MAX, MIN, AVG, COUNT, SUM, GROUP\_CONCAT, etc. See: <https://dev.mysql.com/doc/refman/8.0/en/aggregate-functions.html>
* The HAVING clause allows using group functions in the condition. The WHERE clause does not allow using group functions.
* Using GROUP BY, the columns of the SELECT clause can only have:
  1. Columns named in GROUP BY.
  2. Aggregate functions on other columns in the tables.
  3. Constant expressions.
* A number can be used in the GROUP BY and the ORDER BY clauses to refer to the positions of the result columns in the select clauses.

Thus, the conceptual steps and framework for the SELECT statement become

SELECT DISTINCT <<result\_columns>> -- [5] construct result columns  
FROM <<source\_tables>> -- [1] conceptually join tables to form a large table to produce initial rows  
WHERE <<conditions\_for\_inclusion>> -- [2] Filter initial rows  
GROUP BY <<group\_by\_columns>>  
         --[3] group initial rows into groups by values of the group\_by\_column. A group becomes a new row.  
HAVING <<conditions for filtering group>> -- [4] filter groups  
ORDER BY <<columns>>; -- [6] Order the result of [5].

***Example:***

-- Student names and number of classes enrolled.  
-- More than 2 classes to be included in the result.  
SELECT CONCAT(s.fname, ' ', s.lname) AS student,  
   COUNT(e.classId) AS `Enrolled classes`  
FROM student AS s INNER JOIN enroll e ON (s.stuId = e.stuId)  
GROUP BY student  
HAVING `Enrolled classes` > 2  
ORDER BY `Enrolled classes` DESC;

***Exercises***:

[1] Write a query to generate the student names and number of courses enrolled, including those not enrolled?

+-----------------+------------------+  
| name            | Enrolled classes |  
+-----------------+------------------+  
| Tony Hawk       |                6 |  
| Linda Johnson   |                4 |  
| David Hawk      |                3 |  
| Ben Zico        |                2 |  
| Larry Johnson   |                2 |  
| Mary Hawk       |                2 |  
| Lillian Johnson |                2 |  
| Bill Ching      |                1 |  
| Catherine Lim   |                0 |  
| Linda King      |                0 |  
+-----------------+------------------+  
10 rows in set (0.00 sec)  
  
Solution:

SELECT CONCAT(s.fname, ' ', s.lname) AS student,  
    COUNT(e.classId) AS `Enrolled classes`  
FROM student AS s LEFT JOIN enroll e ON (s.stuId = e.stuId)  
GROUP BY student   
ORDER BY `Enrolled classes` DESC; 

[2] Can you write a query to generate the following output?

+----------+------------------------------+------------+----------+----------+  
| deptCode | deptName                     | numFaculty | numMajor | numMinor |  
+----------+------------------------------+------------+----------+----------+  
| ACCT     | Accounting                   |          1 |        0 |        0 |  
| ARTS     | Arts                         |          1 |        2 |        0 |  
| CINF     | Computer Information Systems |          2 |        2 |        3 |  
| CSCI     | Computer Science             |          4 |        3 |        1 |  
| ENGL     | English                      |          1 |        0 |        2 |  
| ITEC     | Information Technology       |          2 |        2 |        2 |  
| MATH     | Mathematics                  |          0 |        0 |        0 |  
+----------+------------------------------+------------+----------+----------+  
7 rows in set (0.00 sec)

Solution:

WITH ma AS  
   (SELECT s.major AS deptCode, COUNT(s.stuId) AS numMajor  
    FROM student AS s  
    GROUP BY s.major),  
mi AS  
   (SELECT s.minor AS deptCode, COUNT(s.stuId) AS numMinor  
    FROM student AS s  
    GROUP BY s.minor),  
f AS  
   (SELECT f.deptCode, COUNT(f.facId) AS numFaculty  
    FROM faculty AS f  
    GROUP BY f.deptCode)  
SELECT d.deptCode,  
   d.deptName,  
   IFNULL(f.numFaculty, 0) AS numFaculty,  
   IFNULL(ma.numMajor, 0) AS numMajor,  
   IFNULL(mi.numMinor, 0) AS numMinor  
FROM department AS d LEFT JOIN ma USING (deptCode)  
   LEFT JOIN mi USING (deptCode)  
   LEFT JOIN f USING (deptCode);

**3.7 Window Functions**

* MySQL 9.x supports Window functions.
* A window function performs a computation on a set of rows (a window frame) in which the current row is in the window frame.
* It is not a clause.
* Unlike the GROUP BY clause, it does not form groups.
* The OVER clause is used to define the window frame.
* OVER(): all rows are in the window frame.
* OVER(PARTITION BY X): each X value defines a window frame.
* Many aggregate functions can be used by Window functions.
* Modern DBMS support a rich set of Window functions.

***Example:***

WITH temp AS  
(SELECT DISTINCT sc.schoolName AS college, d.deptName AS department,  
   COUNT(s.stuId) As deptMajor  
FROM school AS sc INNER JOIN department AS d ON (sc.schoolCode = d.schoolCode)  
   LEFT JOIN student AS s ON (s.major = d.deptCode)  
GROUP BY college, department)  
SELECT temp.college, temp.department,  
   temp.deptMajor AS `major in department`,  
   SUM(deptMajor) OVER(PARTITION BY college) AS `major in college`,  
   SUM(deptMajor) OVER() AS `major in university`  
FROM temp;

Please execute to see the output.

Adding row number and rank:

WITH ma AS  
(SELECT s.major AS deptCode, COUNT(s.stuId) AS numMajor  
FROM student AS s  
GROUP BY s.major),  
mi AS  
(SELECT s.minor AS deptCode, COUNT(s.stuId) AS numMinor  
FROM student AS s  
GROUP BY s.minor),  
f AS  
(SELECT f.deptCode, COUNT(f.facId) AS numFaculty  
FROM faculty AS f  
GROUP BY f.deptCode)  
SELECT ROW\_NUMBER() OVER () AS `#`,  
   RANK() OVER (ORDER BY f.numFaculty DESC) AS `# in descending number of faculty`,  
   d.deptCode,  
   d.deptName,  
   IFNULL(f.numFaculty, 0) AS numFaculty,  
   IFNULL(ma.numMajor, 0) AS numMajor,  
   IFNULL(mi.numMinor, 0) AS numMinor  
FROM department AS d LEFT JOIN ma USING (deptCode)  
   LEFT JOIN mi USING (deptCode)  
   LEFT JOIN f USING (deptCode);

**An Introduction to Python**

by K. Yue

**1. Resources**

* Python: <https://www.python.org/>
* Version 3.11: (navigate to your version).
  + Manual: <http://docs.python.org/3.11/reference/index.html>
  + Tutorial: <https://docs.python.org/3.11/tutorial/interpreter.html>
  + Standard library: <https://docs.python.org/3.11/library/index.html>
* Installation notes:
  + It is recommended that you should install Python*using customization*:
    1. Put Python in the *top level* of your C drive. For example: in c:\Python311\.
    2. Select to let the installer set Python path environment variable.
    3. Select to let the installer disable Windows path length limit (if not, limit is 260 characters). This option is shown after successful installation.
* Python's Beginner Guide for programmers: <https://wiki.python.org/moin/BeginnersGuide/Programmers>
* Style Guide for Python Code:
  + <https://www.python.org/dev/peps/pep-0008/> (dated but still containing useful tips).
  + [Google Python Guideline](https://google.github.io/styleguide/pyguide.html)
* For experienced programmers: <https://learnxinyminutes.com/docs/python3/>

**2. Basics**

* The course assumes that you are reasonable comfortable with Python for program development.
* Python is a high-level, open source, general-purpose, object-oriented, extensible, interpreted scripting programming language.
* Python is a scripting language:
  1. Especially good for scripting and rapid application development
  2. Object-oriented language (more so than many other OO languages)
  3. Dynamic typing: an object bound to a variable can change type dynamically.

Java: static typing: performance, security/reliability

int i; // 4 bytes: stack for storing i.

* 1. Strongly typed: less implicit type conversion.
  2. Native high level data structures: list, tuple, set, dictionary, etc.
  3. Interpreted
  4. Automatic garbage collection
  5. Open source
  6. Use good software engineering principles
* Using the Python's Integrated Development and Learning Environment (IDLE) interpreter is a good way to learn Python.
  1. Use Control-z or quit() to exit Python interpreter.

**2.1 Python programs: some basic concepts**

* A Python program contains *logical lines*.
* A logical line usually contains one *physical line* but can be extended by '\' to contain many physical lines.
* Implicit line joining is performed for (), [] and {} expressions. It is preferred to using '\'.
* Do not use ';' as a statement separator.
* Indentation (white spaces at the beginning of lines) has meaning. Proper indentations are mandatory.
* *Indentation level* is used to group statements.
* Python 2: suggest to not mix tabs and spaces.
* Python 3:
  + Disallow mixing of tabs and spaces.
  + Spaces are preferred.
* Comments start with a '#'. You may also use a multiple line string, such as ''' a comment. '''
* ...

**2.2 Running Python in Windows**

* To start the Python interpreter, open a command line prompt in your working directory and execute, for example:

set path=%path%;C:\Python311  
python

* You may need to replace "C:\Python311" by the location of your python installation.
* You may set the environmental variable PATH in your OS. If so, you will not need to execute the set path command.
* To run a Python program, helloworld.python, use:

python helloworld.py

**2.3 How to learn a new language?**

1. syntax
2. concepts that are familiar to you: note any difference.
3. concepts that are new to you
4. resources and libraries
5. design patterns and best practices

**2.4 Some basics to get you interested (hopefully):**

* No ;
* No i++
* Multiple assignment: a,b = b,a
* strings are automatically concatenated: print("ab" "cd")
* strings are automatically indexed:
  + a='12345'
  + print(a[1:3])
* id() to find out id of an object.
* Python does not support the traditional array data type. Instead, lists can be viewed as a more general array type.

**2.5 Examples of concepts that may be new to you:**

* Immutable or mutable?
  + The *value* of an immutable object cannot be changed.
  + Use the id() function for experimentation.
  + Some data types require immutable objects. E.g., member of a set and key of a dict.
* Automatic garbage collection: "Objects are never explicitly destroyed; however, when they become unreachable they may be garbage-collected."
* Built-in Python data structures: list, set, tuple and dictionary.

***Example:***

Print first 100 Fibonacci numbers: [fib.py.txt](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/emSQL/fib.py.txt) (remove .txt when saving)

#   print first 100 Fibonacci numbers  
a, b, count = 0, 1, 1  
while count <= 100:  
   print(b, ' ', end='')  
   a, b, count = b, a+b, count+1  
print()

Note the use of multiple assignment statements. In other languages, such as Java, you may need to replace:

a, b, count = b, a+b, count+1

by

temp=b;  
b=a+b;  
a=temp;  
count++;

Other examples as will be discussed in the class.

**3. A Non-trivial Example**

***Example:***

Consider the following weather information file: [201401daily\_sample.csv](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/emSQL/201401daily_sample.csv). Write a Python program to read the 'codeSum' column (#23) and output the count of each codeSum. Note that the codeSum column may contain more than one codeSum separated by white spaces. Example:

SN FG+ FZFG BR UP

has five codeSum:

* SN
* FG+
* FZFG
* BR
* UP

[codeSum.py](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/emSQL/codeSum.py.txt):

import sys, getopt  
import re  
from operator import itemgetter  
  
#   getopt: C-style parser for command line options.  
#   sys: System-specific parameters and functions.  
#   re: regular expression operation  
#   operator: methods for built-in operators.  
#       (useful when passing the function as a parameter.)  
  
#   Read and process a weather information file.  
#   It parse the codeSum column (#23) in the CSV file  
#   and show the count of each codeSum.  
  
f = open(sys.argv[1], 'r')  
result = [];  
heading = f.readline().split(',')  
num\_line = 0;  
  
for line in f:  
    line = line.rstrip()    #   strip trailing white spaces  
    #   result is an array of arrays:  
    #   Add the array contains columns of the current line  
    #   to result.  
    result.append(line.split(','))  
    num\_line = num\_line + 1  
f.close()  
  
#   Debug:  
#   for i in range(len(result)):  
#       print (str(i) + ":" + str(result[i][22]))  
  
#   count is a dictionary with the key being the individual CodeSummary  
count = {};  
for i in range(len(result)):  
    #   process one reading.  
    line = result[i][22].strip()  
    if line:  
        #   Get all codeSummary and update their counts.  
        summary = re.split('\s+', line)  
        for j in range(len(summary)):  
            #   Debug:  
            #   print (str(j) + ":" + str(summary[j]) + "---")  
            if summary[j] in count.keys():  
                count[summary[j]] += 1  
            else:  
                count[summary[j]] = 1  
  
#   Print result in the sorted order of codeSum.                 
for key, value in sorted(count.items(), key=itemgetter(0)):  
    print (key + ": " + str(value))  
  
             
Running the program:

>...python codeSum.py 201401daily\_sample.csv  
BR: 18  
FG: 2  
FG+: 5  
FZFG: 8  
HZ: 10  
RA: 2  
SN: 14  
UP: 4  
  
Version 2: more Python-style with some interesting Python features:

[codeSum2.py](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/emSQL/codeSum2.py.txt):

from collections import defaultdict  
import sys, getopt  
import re  
from operator import itemgetter  
  
#   Read and process a weather information file.  
#   It parse the codeSum column (#23) in the CSV file  
#   and show the count of each codeSum.  
  
f = open(sys.argv[1], 'r')  
heading = f.readline().split(',')  
result = [line.strip().split(',') for line in f.readlines()]  
f.close()  
  
#   count is a dictionary with the key being the individual CodeSummary  
count = defaultdict(int)  
for field in filter(lambda a: a, map(lambda w: w[22].strip(), result)):  
   for sym in re.split('\s+', field):  
      count[sym] += 1  
    
#   Print result in the sorted order of codeSum.                 
for key, value in sorted(count.items(), key=itemgetter(0)):  
    print (key + ": " + str(value))

**4. Some useful tools**

**4.1 Logging**

* Python logging: <https://docs.python.org/3/howto/logging.html>
* "Logging is a means of tracking events that happen when some software runs. The software’s developer adds logging calls to their code to indicate that certain events have occurred."
* Logging can be useful in debugging programs.

***Example:***

[logExample.py.txt](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/emSQL/logExample.py.txt):

import logging  
  
# logging  
logLevel = logging.DEBUG  
logPath = "."  
  
logger = logging.getLogger()  
logger.setLevel(logLevel)  
  
logFormatter = logging.Formatter("%(asctime)s::%(threadName)-12.12s::[%(levelname)-5.5s] %(message)s")  
  
consoleHandler = logging.StreamHandler()  
consoleHandler.setFormatter(logFormatter)  
logger.addHandler(consoleHandler)  
  
logger.debug("Program starts.")  
logger.debug("Program continues.")

**4.2 Jupyter Notebook (especially for DASC 5333)**

* <https://jupyter.org/>: "The Jupyter Notebook is the original web application for creating and sharing computational documents."
* Installation and launching:
  + Jupyter Notebook:
    - Installation: using pip: "pip install notebook"
    - Launching server: open a terminal in the work directory and execute the command: "jupyter notebook".
    - Stopping server: [1] crtl-c, or [2] open anther command terminal, and execute: "jupyter notebook stop 8888", where 8888 is the (default) port of the server.
  + JupyterLab Notebook: next generation notebook.
    - Installation: using pip: "pip install jupyterlab"
    - Launching: open a terminal in the work directory and execute the command "jupyter-lab".
    - Stopping server: crtl-c
* If there is time to cover this topic, we will use JupyterLab, which is preferred.

***Example:***

Try out: [CodeSumNB.zip](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2024/notes/emSQL/CodeSumNB.zip). Unzipp to get CodeSumNB.ipynb. Put it in your notebook working directory and open it with juypter-lab.