**DASC 5333**

**3/25/2025**

**More SQL**

by K. Yue

**1. More SQL Features**

**1.1 Prepared Statements**

* MySQL supports server-side prepared statements.
* A prepared statement is used in the following sequence.
	1. The statement is prepared.
	2. The prepared statements are executed one to many times, including using different values on placeholders.
	3. The statement is deallocated.
* Prepared Statements have the following benefits:
	1. They are compiled once and can be executed many times.
	2. With the use of placeholders, they protect against SQL injections.
	3. They can be used to execute dynamic SQL statements.
		1. However, dynamic SQL statements can be serious security holes and should be used carefully.
* In MySQL, prepared statements can be used in
	1. Applications: such as SQL clients (e.g. MySQL prompt) or Python programs.
	2. SQL scripts: such as invoked by the source command.
* There are vendor-dependent restrictions in prepared statements.
* For examples, in MySQL,
	1. Constants or user variables should be used as the parameter values to the placeholders of the prepared statements.
	2. user variables should be used in the INTO clause of the SQL query in the prepared SELECT statement.

***Example:***

Try the following code in MySQL prompt.

-- Prepared statements.
SET @sql = "SELECT \* FROM toyu.student";
PREPARE stmt FROM @sql;
EXECUTE stmt;
DEALLOCATE PREPARE stmt;

-- with placeholders.
SET @sql = "SELECT \* FROM toyu.student WHERE major = ? AND ach >= ?";
PREPARE stmt FROM @sql;
SET @major = 'CSCI';
SET @ach = 38;
EXECUTE stmt USING @major, @ach;

EXECUTE stmt USING 'CSCI', 38;

SET @major = 'CINF';
SET @ach = 15;
EXECUTE stmt USING @major, @ach;

SET @major = 'ITEC';
SET @ach = 25;
EXECUTE stmt USING @major, @ach;

DEALLOCATE PREPARE stmt;

**2. Views**

* Views are *virtual* tables *derived* from other tables.
* In MySQL, "views are stored queries that when invoked produce a *result set*. A view acts as a virtual table." See: <https://dev.mysql.com/doc/refman/8.2/en/views.html>
* Some advantages of using views:
	1. Better data abstraction: hiding unnecessary information.
	2. Logical data independence
	3. Better consistency
	4. More security control
	5. Possibly more efficient



* Some disadvantages:
	1. More work
	2. Complicated especially when views are updated.
	3. Performance: view processing may not be optimized, e.g.,
	4. A view can be treated as a subquery by the db engine.
	5. Cannot create index for view in MySQL.
* See the [three layered DB architecture](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2025/notes/intro/DBBasics.html).
* MySQL Create View Manual: search for "mysql view manual"

CREATE
    [OR REPLACE]
    [ALGORITHM = {UNDEFINED | MERGE | TEMPTABLE}]
    [DEFINER = { user | CURRENT\_USER }]
    [SQL SECURITY { DEFINER | INVOKER }]
    VIEW view\_name [(column\_list)]
    AS select\_statement
    [WITH [CASCADED | LOCAL] CHECK OPTION]

* There are some limitations in views. For example:
	+ In earlier versions of MySQL, subqueries are not allowed in the SELECT clause in the CREATE VIEW statement.
* You may need to define intermediate views accordingly.

***Example:***

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

CREATE OR REPLACE VIEW school\_summary(
   schoolCode, schoolName, n\_departments) AS
SELECT s.schoolCode, s.schoolName,
COUNT(d.deptCode) AS n\_departments
FROM school AS s LEFT JOIN department AS d ON (s.schoolCode = d.schoolCode)
GROUP BY s.schoolCode, s.schoolName;

SHOW CREATE VIEW school\_summary;
-- Note something like "ALGORITHM=UNDEFINED DEFINER=`yue`@`localhost` SQL SECURITY DEFINER"
-- (default values) may be added.

DESC school\_summary;

SELECT \*
FROM school\_summary
WHERE n\_departments > 0;

DROP VIEW school\_summary;

Notes:

* "CREATE OR REPLACE" can replace existing view definition.
* There are no data types explicitly specified for the columns in the example.
* The column list is optional.
* MySQL supports specification of algorithm for processing views:
	+ MERGE: view definition merged into the containing query (inlining)
	+ TEMPTABLE: view results stored in temporary table to be used in the containing query.
* MySQL supports updatable views. Use them carefully. E.g., the view 'user' in 'mysql' is updatable.
	+ In general, "for a view to be updatable, there must be a one-to-one relationship between the rows in the view and the rows in the underlying table."

**3. Stored Subroutines**

* Subroutines stored by the DBMS can be called.
* Subroutines can be
	1. functions: return a value, or
	2. procedures:
		1. do not return a value, and
		2. work by side effects.
* Known as SQL/PSM (SQL/Persistent Stored Modules) in MySQL.
	1. Can execute SQL statements.
	2. Include general programming constructs.
* Some advantages of stored subroutines:
	1. Enforcing application constraints and requirements.
	2. Providing consistency and security control.
	3. Possible performance optimization by both the developers and the DBMS.
	4. Sharing among DB applications, especially when they use different languages.
* Some disadvantages of stored procedures:
	1. Use up DB server's resources.
	2. Potentially inefficient because of limitations in language constructs.
	3. Potentially harder to develop because of relative lack of libraries and difficulty in debugging.
	4. Can be vendor specific.
* In general, stored subroutines should be used more often.
* Some general points:
	1. A stored subroutine is associated with a particular database.
	2. Stored functions cannot be recursive.

**3.1 Stored Procedures**

***Example:***

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

-- A very simple stored procedure.
-- Redefine the delimiter to end the procedure.
DELIMITER //

CREATE OR REPLACE PROCEDURE deptInfo(IN dCode VARCHAR(4), OUT numFaculty INT)
BEGIN
   --  Display some information.
   SELECT d.deptName, d.SchoolCode, t1.n\_majors, t2.n\_minors
   FROM department AS d INNER JOIN
      (SELECT COUNT(stuId) AS n\_majors
      FROM student
      WHERE major = dCode) AS t1 INNER JOIN
      (SELECT COUNT(stuId) AS n\_minors
      FROM student
      WHERE minor = dCode) AS t2
   WHERE d.deptCode = dCode;

   -- MySQL does not direct sending output to console.
   -- It is necessary to use a SQL statement.
   SELECT 'Debuggin comment can be put here.';
   SELECT CONCAT('Faculty in the department: ', dCode) AS faculty;

   SELECT \*
   FROM faculty AS f
   WHERE f.deptCode = dCode;

   SELECT COUNT(f.facId) INTO numFaculty
   FROM faculty AS f
   WHERE f.deptCode = dCode;
END //

DELIMITER ;

SHOW CREATE PROCEDURE deptInfo;

SET @numFaculty = 0;
SET @dCode = 'CSCI';
CALL deptInfo(@dCode, @numFaculty);
SELECT @dCode, @numFaculty;

SET @dCode = 'ITEC';
CALL deptInfo(@dCode, @numFaculty);
SELECT @dCode, @numFaculty;

DROP PROCEDURE deptInfo;

Note:

1. Use of the DELIMITER command to redefine '//' as the delimiter indicating the end of the stored procedure. Otherwise, the default ';' is the delimiter indicating the end of the stored procedure.
2. Two parameters for this procedure: one using IN and one using OUT as the parameter passing mechanism.
3. A procedure does not return any value and accomplish its goal through *side effects*.
4. Side effects include:
	1. Return SELECT results in the procedure body.
	2. Copy numFaculty out upon completion.

***Example: using toyu***

DELIMITER //
CREATE OR REPLACE PROCEDURE AddNewCourse(
    IN course\_id INT,
    IN rubric CHAR(4),
    IN course\_number CHAR(4),
    IN course\_title VARCHAR(80),
    IN credits TINYINT
)
BEGIN
    INSERT INTO Course (courseId, rubric, number, title, credits)
    VALUES (course\_id, rubric, course\_number, course\_title, credits);
END //
DELIMITER ;

CALL AddNewCourse(3009, 'CSCI', '4436', 'Systems Administration', 3);

**3.2 Stored Functions**

***Example:***

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

-- A simple function
DELIMITER //

CREATE OR REPLACE FUNCTION n\_major(dCode varchar(4)) RETURNS INT
READS SQL DATA
BEGIN
   DECLARE count INT DEFAULT 0;

   SELECT COUNT(\*) INTO count
   FROM student
   WHERE major = dCode;

   RETURN count;
END //

DELIMITER ;

SHOW CREATE FUNCTION n\_major;

SELECT n\_major('CSCI');
SELECT n\_major('ITEC');

Note:

* A function returns a value and does not accomplish its goal through side effects.
* Note the argument passed into the function.
* Formal arguments of a function use the IN parameter passing mode.
* Note the variable declaration and assignment.
* The scope of the variable is within the function.
* 'READS SQL DATA' describes the characteristic of the subroutine: it reads but not writes SQL data.

***Example:***

The following example functions are included in the script of creating toyu. Function calls are added.

-- get the full name of a student.
DELIMITER //
CREATE OR REPLACE FUNCTION GetStudentFullName(
    student\_id INT
)
RETURNS VARCHAR(61)
DETERMINISTIC
BEGIN
    DECLARE full\_name VARCHAR(61);
    SELECT CONCAT(fname, ' ', lname) INTO full\_name
    FROM Student
    WHERE stuId = student\_id;
    RETURN full\_name;
END //
DELIMITER ;

SELECT GetStudentFullName(100000);
SELECT GetStudentFullName(100001);

-- get the full name of a department code
DELIMITER //
CREATE OR REPLACE FUNCTION GetDepartmentName(
    dept\_code CHAR(4)
)
RETURNS VARCHAR(30)
DETERMINISTIC
BEGIN
    DECLARE dept\_name VARCHAR(30);
    SELECT d.deptName INTO dept\_name
    FROM Department d
    WHERE d.deptCode = dept\_code;
    RETURN dept\_name;
END //
DELIMITER ;

SELECT GetDepartmentName('CSCI');
SELECT GetDepartmentName('ITEC');

-- Compute and return the GPA of a student
DELIMITER //
CREATE OR REPLACE FUNCTION GetStudentGPA(
    student\_id INT
)
RETURNS DECIMAL(3,2)
DETERMINISTIC
BEGIN
    DECLARE gpa DECIMAL(3,2);
    SELECT ROUND(SUM(g.gradePoint \* co.credits) / SUM(co.credits), 2) INTO gpa
    FROM Enroll e
    JOIN Class c ON e.classId = c.classId
    JOIN Course co ON c.courseId = co.courseId
    JOIN Grade g ON e.grade = g.grade
    WHERE e.stuId = student\_id;
    RETURN gpa;
END //
DELIMITER ;

SELECT GetStudentGPA(100000);
SELECT GetStudentGPA(100001);
SELECT GetStudentGPA(100002);

-- Return the top n students in a major department with GPA.
DELIMITER //
CREATE OR REPLACE FUNCTION GetTopStudentsInDepartment(
    dept\_code CHAR(4),
    top\_count INT
)
RETURNS VARCHAR(1000)
DETERMINISTIC
BEGIN
    DECLARE student\_list VARCHAR(1000);
    SET student\_list = '';

    WITH temp AS(
        SELECT s.stuId, CONCAT(s.fname, ' ', s.lname) AS student,
            ROUND(SUM(g.gradePoint \* co.credits) / SUM(co.credits), 2) AS gpa
        FROM Student s
        JOIN Department d ON s.major = d.deptCode
        JOIN Enroll e ON s.stuId = e.stuId
        JOIN Class c ON e.classId = c.classId
        JOIN Course co ON c.courseId = co.courseId
        JOIN Grade g ON e.grade = g.grade
        WHERE d.deptCode = dept\_code
        GROUP BY s.stuId, student
        ORDER BY gpa DESC
        LIMIT top\_count)
    SELECT GROUP\_CONCAT(CONCAT(temp.student, '(', temp.stuId, '):', temp.gpa) SEPARATOR ', ')
        INTO student\_list
    FROM temp
    LIMIT 1;

    RETURN student\_list;
END //
DELIMITER ;

SELECT GetTopStudentsInDepartment('CSCI', 2);
SELECT GetTopStudentsInDepartment('CSCI', 3);

**3.3 Cursors**

* In stored subroutines in MySQL, various high level language constructs are supported, such as variable declarations, conditional statements, control statements, etc.
* In particular, cursors are supported for allowing iteration through the result sets.

***Example:***

Execute the following code and ensure that you understand the result. The example is artifically constructed as there are better ways.

-- using cursor.
DELIMITER //

CREATE FUNCTION major\_students(dept VARCHAR(4))
RETURNS VARCHAR(1000)
READS SQL DATA
BEGIN
DECLARE result VARCHAR(1000) DEFAULT '';
DECLARE name VARCHAR(41) DEFAULT '';
DECLARE done INT DEFAULT FALSE;

DECLARE cursor\_1 CURSOR FOR
SELECT DISTINCT CONCAT(fName, ' ', lName) AS name
FROM student
WHERE major = dept;

DECLARE continue handler FOR NOT FOUND SET done = TRUE;

OPEN cursor\_1;
SET result = '';

compute\_loop: LOOP
-- This is needed as there are other SQL statements that may set done to true.
SET done = false;
FETCH cursor\_1 INTO name;
IF done THEN
LEAVE compute\_loop;
END IF;

IF (result <> '') THEN
SET result = CONCAT(result, ', ');
END IF;

SET result = CONCAT(result, name);
END LOOP;

CLOSE cursor\_1;

RETURN result;
END //

DELIMITER ;

SELECT major\_students('CSCI');
SELECT major\_students('CINF');

DROP FUNCTION major\_students;

* Note that the example is only used to demonstrate cursors in a stored function. A much better solution is to simply use GROUP\_CONCAT, such as:

SELECT GROUP\_CONCAT(CONCAT(fName, ' ', lName) SEPARATOR ', ') AS majors
FROM student
WHERE major = 'CSCI';

**4. Triggers**

* Triggers allow *event-driven* programming.
* A trigger is activated when certain events occur. Unlike stored routines, triggers are not explicitly called.
* Four questions to ask for an event model:
	1. What are the events?
	2. What information can be obtained from the events?
	3. What actions can be performed to handle events?
	4. How do the events propagate?

**4.1 The event model for triggers**

* Events:
	1. Inserting a row
	2. Updating a row
	3. Deleting a row
* Actions can be executed:
	1. Before the event
	2. After the event
* Information can be obtained through two keywords:
	1. old:
		+ the old value of the row before the event.
		+ for update and delete events
	2. new:
		+ the new value of the row after the event.
		+ for update and insert events
* Triggers do not propagate.

Advantages of triggers

1. The event driven model suits certain kinds of tasks better.
2. Can ease the implementation of consistency check.
3. Can ease the implementation of business logic and integrity check.

Disadvantages:

1. Relatively invisible, and possibly overlooked by developers.
2. Relatively difficult to debug.
3. Potential performance issues.
4. Potential complicated interactions.

***Example:***

An example of using trigger is for auditing, e.g. <https://vladmihalcea.com/mysql-audit-logging-triggers/>

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

-- trigger

CREATE TABLE courseUpdate(
   cuId        INT NOT NULL AUTO\_INCREMENT,
   `type`      CHAR(1),
    courseId    INT NOT NULL,
    rubric      CHAR(4) NOT NULL,
    number      CHAR(4) NOT NULL,
    title       VARCHAR(80) NOT NULL,
    credits     TINYINT NULL,
    ts          TIMESTAMP DEFAULT CURRENT\_TIMESTAMP ON UPDATE CURRENT\_TIMESTAMP,
    CONSTRAINT CourseUpdate\_courseId\_pk PRIMARY KEY (cuId),
    CONSTRAINT CourseUpdate\_deptCode\_fk FOREIGN KEY (rubric)
        REFERENCES Department(deptCode));

-- trigger example.

DELIMITER $$

CREATE TRIGGER update\_Course AFTER UPDATE ON Course FOR EACH ROW
BEGIN
   -- code should be more sophisticated.
   INSERT INTO courseUpdate(`type`, courseId, rubric, number, title, credits)
   VALUES('U', new.courseId, new.rubric, new.number, new.title, new.credits);
END $$

CREATE TRIGGER insert\_Course AFTER INSERT ON Course FOR EACH ROW
BEGIN
   -- code should be more sophisticated.
   INSERT INTO courseUpdate(`type`, courseId, rubric, number, title, credits)
   VALUES('I', new.courseId, new.rubric, new.number, new.title, new.credits);
END $$

DELIMITER ;

SELECT \* FROM Course;

UPDATE Course
SET number = 2315
WHERE courseId = 2000;

INSERT INTO Course(courseId, rubric, number, title, credits) VALUES
    (4000,'CSCI',3532,'Advanced Data Structures',3);

UPDATE Course
SET number = 3341
WHERE courseId = 4000;

SELECT \* FROM course;
SELECT \* FROM courseUpdate;

-- Clean up.
DROP TRIGGER update\_Course;
DROP TRIGGER insert\_Course;

DROP TABLE courseUpdate;

DELETE FROM Course
WHERE CourseId = 4000;

UPDATE Course
SET number = 3333
WHERE courseId = 2000;

SELECT \* FROM Course;

**5. System Catalog**

* Named collections of *meta-data* in SQL DBMS.
* Important for DB administrators and developers to know.
* Usually stored in relations.
* Common terms with similar meaning: data dictionary, catalog, etc.
* Three major databases that are included with MySQL: mysql, information\_schema and performance\_schema.

***Example:***

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information\_schema |
| mysql              |
| performance\_schema |
| swim              |
| toytu              |
| world              |
| yue\_exp            |
+--------------------+
7 rows in set (0.00 sec)
>

**5.1 mysql database**

* MySQL database stores information about the system. It is the system schema.
* See: <https://dev.mysql.com/doc/refman/8.1/en/system-schema.html>.
* However, many of these system tables cannot be accessed directly. They should be access through other means (such as information\_schema, see below)

***Example:***

SELECT DISTINCT u.user, u.password
FROM mysql.user AS u
WHERE u.host = 'localhost';

**5.2 information\_schema database**

* Store information about all databases in MySQL.
* Contents are system*views* using the MEMORY storage engine and they cannot be updated.
* Triggers cannot be defined on tables in information\_schema.

***Example:***

mysql> SELECT table\_name, table\_type, row\_format, table\_rows, avg\_row\_length
    -> FROM information\_schema.tables
    -> WHERE table\_schema = 'information\_schema'
    -> ORDER BY table\_name DESC;
+---------------------------------------+-------------+------------+------------+----------------+
| table\_name                            | table\_type  | row\_format | table\_rows | avg\_row\_length |
+---------------------------------------+-------------+------------+------------+----------------+
| VIEWS                                 | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| USER\_PRIVILEGES                       | SYSTEM VIEW | Fixed      |       NULL |           1986 |
| TRIGGERS                              | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| TABLE\_PRIVILEGES                      | SYSTEM VIEW | Fixed      |       NULL |           2372 |
| TABLE\_CONSTRAINTS                     | SYSTEM VIEW | Fixed      |       NULL |           2504 |
| TABLESPACES                           | SYSTEM VIEW | Fixed      |       NULL |           6951 |
| TABLES                                | SYSTEM VIEW | Fixed      |       NULL |           9450 |
| STATISTICS                            | SYSTEM VIEW | Fixed      |       NULL |           5753 |
| SESSION\_VARIABLES                     | SYSTEM VIEW | Fixed      |       NULL |           3268 |
| SESSION\_STATUS                        | SYSTEM VIEW | Fixed      |       NULL |           3268 |
| SCHEMA\_PRIVILEGES                     | SYSTEM VIEW | Fixed      |       NULL |           2179 |
| SCHEMATA                              | SYSTEM VIEW | Fixed      |       NULL |           3464 |
| ROUTINES                              | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| REFERENTIAL\_CONSTRAINTS               | SYSTEM VIEW | Fixed      |       NULL |           4814 |
| PROFILING                             | SYSTEM VIEW | Fixed      |       NULL |            308 |
| PROCESSLIST                           | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| PLUGINS                               | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| PARTITIONS                            | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| PARAMETERS                            | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| KEY\_COLUMN\_USAGE                      | SYSTEM VIEW | Fixed      |       NULL |           4637 |
| INNODB\_TRX                            | SYSTEM VIEW | Fixed      |       NULL |           4534 |
| INNODB\_LOCK\_WAITS                     | SYSTEM VIEW | Fixed      |       NULL |            599 |
| INNODB\_LOCKS                          | SYSTEM VIEW | Fixed      |       NULL |          31244 |
| INNODB\_CMP\_RESET                      | SYSTEM VIEW | Fixed      |       NULL |             25 |
| INNODB\_CMPMEM\_RESET                   | SYSTEM VIEW | Fixed      |       NULL |             29 |
| INNODB\_CMPMEM                         | SYSTEM VIEW | Fixed      |       NULL |             29 |
| INNODB\_CMP                            | SYSTEM VIEW | Fixed      |       NULL |             25 |
| GLOBAL\_VARIABLES                      | SYSTEM VIEW | Fixed      |       NULL |           3268 |
| GLOBAL\_STATUS                         | SYSTEM VIEW | Fixed      |       NULL |           3268 |
| FILES                                 | SYSTEM VIEW | Fixed      |       NULL |           2677 |
| EVENTS                                | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| ENGINES                               | SYSTEM VIEW | Fixed      |       NULL |            490 |
| COLUMN\_PRIVILEGES                     | SYSTEM VIEW | Fixed      |       NULL |           2565 |
| COLUMNS                               | SYSTEM VIEW | Dynamic    |       NULL |              0 |
| COLLATION\_CHARACTER\_SET\_APPLICABILITY | SYSTEM VIEW | Fixed      |       NULL |            195 |
| COLLATIONS                            | SYSTEM VIEW | Fixed      |       NULL |            231 |
| CHARACTER\_SETS                        | SYSTEM VIEW | Fixed      |       NULL |            384 |
+---------------------------------------+-------------+------------+------------+----------------+
37 rows in set (0.08 sec)

* It is more flexible than the SHOW command.

***Example:***

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

-- System Catalog
-- Getting selected columns from information\_schema
SELECT table\_name, table\_type, row\_format, table\_rows, avg\_row\_length
FROM information\_schema.tables
WHERE table\_schema = 'information\_schema'
ORDER BY table\_name DESC;

-- databases and tables
SELECT t.TABLE\_SCHEMA AS `schema`, COUNT(t.TABLE\_NAME) AS num\_tables
FROM information\_schema.tables AS t
GROUP BY `schema`
ORDER BY num\_tables DESC;

SELECT t.TABLE\_SCHEMA AS `schema`, t.ENGINE, COUNT(t.TABLE\_NAME) AS num\_tables
FROM information\_schema.tables t
GROUP BY `schema`, t.ENGINE
ORDER BY `schema`, num\_tables DESC;

***Example:***

DROP SCHEMA IF EXISTS dbtool;
CREATE SCHEMA dbtool;
USE dbtool;

-- Return the number of columns of a table in a schema (database) in the output parameter column\_count
DELIMITER //
CREATE OR REPLACE PROCEDURE count\_columns(
    IN schema\_name VARCHAR(64),
    IN table\_name VARCHAR(64),
    OUT column\_count INT
)
BEGIN
    SET @\_\_cc\_query = CONCAT('SELECT COUNT(\*) INTO @\_\_cc\_column\_count FROM INFORMATION\_SCHEMA.COLUMNS WHERE TABLE\_SCHEMA = ? AND TABLE\_NAME = ?');
    PREPARE stmt FROM @\_\_cc\_query;
    SET @\_\_cc\_schema\_name = schema\_name;
    SET @\_\_cc\_table\_name = table\_name;

    EXECUTE stmt USING @\_\_cc\_schema\_name, @\_\_cc\_table\_name;
    SET column\_count = @\_\_cc\_column\_count;
    DEALLOCATE PREPARE stmt;
END //
DELIMITER ;

CALL count\_columns('toyu', 'student', @column\_count);
SELECT @column\_count;
CALL count\_columns('swim', 'swimmer', @column\_count);
SELECT @column\_count;

**5.3 performance\_schema database**

* contain performance related server data.

**An Introduction to Python**

by K. Yue

**1. Resources**

* Python's installation: Please install Python *v3.10* or below (Good for CGI) for development in this course. Do not install the later version.
* 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 the successful installation screen.
* 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 many 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 reasonably 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. (as opposed to static typing)
	4. Strongly typed: less implicit type conversion.
	5. Native high level data structures: list, tuple, set, dictionary, etc.
	6. Interpreted
	7. Automatic garbage collection
	8. Open source
	9. Use good software engineering principles.

Python 3.10.8 (tags/v3.10.8:aaaf517, Oct 11 2022, 16:50:30) [MSC v.1933 64 bit (AMD64)] on win32

Type "help", "copyright", "credits" or "license()" for more information.

""" Multiline strings can be written

 using three "s, and are often used

 as documentation.

"""

' Multiline strings can be written\n using three "s, and are often used\n as documentation.\n'

p = """ Multiline strings can be written

 using three "s, and are often used

 as documentation.

"""

p

' Multiline strings can be written\n using three "s, and are often used\n as documentation.\n'

print('hello, world')

hello, world

x = 'Bun Yue likes ice cream'

x

'Bun Yue likes ice cream'

x.upper()

'BUN YUE LIKES ICE CREAM'

y =673864871

y

673864871

id(x)

2503942659712

y =784975982478957234895723948579847594379

id(y)

2503942470224

y = 'a summer night'

* 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 meanings. 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 Concepts to learn in 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
6. 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/Spring2025/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 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/Spring2025/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/Spring2025/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/Spring2025/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/Spring2025/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/Spring2025/notes/emSQL/CodeSumNB.zip). Unzipp to get CodeSumNB.ipynb. Put it in your notebook working directory and open it with juypter-lab.

**MySQL Programming in Python**

by K. Yue

**1. MySQL Drivers for Python**

* There are many MySQL drivers for Python. Some examples:
	+ MySQLDB Python: <http://sourceforge.net/projects/mysql-python/>
	+ MySQL Connector Python: <https://dev.mysql.com/downloads/connector/python/>: written with pure Python.
	+ MySQL Connector Python's Developer Guides: <http://dev.mysql.com/doc/connector-python/en/index.html>



* Selecting the right driver is important and not straightforward. Examples:
	+ An example comparing driver performance: <https://gist.github.com/methane/90ec97dda7fa9c7c4ef1>
	+ PyMySQL evaluation: <https://wiki.openstack.org/wiki/PyMySQL_evaluation>
* Because of some problems of MySQL Connector in Python 3, we use*PyMySQL.*

**2. PyMySQL:**

* Installation:
	+ Install pip (Package Installer for Python), if necessary: <https://pip.pypa.io/en/stable/installing/>.
	+ Command: "pip install PyMySQL"
* Documentation: <http://pymysql.readthedocs.io/en/latest/>
* Compliant with Python Database API Specification v2.0: <https://www.python.org/dev/peps/pep-0249/#module-interface>.

|  |
| --- |
| **Additional Materials: MySQL Connector Python (*not* used in this course, but it is a good resource)*** API reference: <http://dev.mysql.com/doc/connector-python/en/connector-python-reference.html>
* Read MySQL Connector Python's Guideline for developer: [http://dev.mysql.com/doc/connector-python/en/connector-python-coding.htm](http://dev.mysql.com/doc/connector-python/en/connector-python-coding.html)
* Examples of good tips and insight:
	1. "Use config.py module to store database connection information."
	2. "Any application that accepts input must expect to handle bad data."
	3. "Data that you choose to store in MySQL instead is likely to have special characteristics."
	4. "you can use Python's triple-quoting mechanism to enclose the entire statement."
	5. "Oracle recommends the ENGINE=INNODB clause for most tables, and makes InnoDB the default storage engine in MySQL 5.5 and up."
 |

**3. Python Programming with MySQL**

* Embedded SQL: embedding SQL in a host language.
* SQL is a DB language, not a general purpose language.
* Host languages:
	1. general purpose languages
	2. usually have vast libraries and tools.
	3. usually have a good community in a specific application area.
* Examples of popular host languages: Java, JavaScript, ts, GOLANG, C#, C, C++, PHP, Perl, Python, Ruby, etc.
* Preparation: create a MySQL account in your computer for development if you have not already done so. (You may use PHPMyAdmin or MySQL Workbench)

**3.1 Basic steps**

* Basic steps of embedded SQL programming.
	1. Create connections to the database.
	2. Prepare and execute SQL statements in the connections by using a cursor.
	3. Use the results of SQL statements.
	4. Perform house cleaning and close connections.
* Step (2) => learn your SQL well.
* Some issues in step (3):
	1. Data type mismatch (impedance mismatch): relations/tables are usually not supported natively in the host languages.
	2. Performance consideration: DB operations are usually much more expensive (take longer time) than CPU operations. Examples:
		1. Minimize DB connections, especially active ones.
		2. Optimize the numbers of SQL statements and their efficiencies.
	3. Security consideration: DB breaches are serious.

**3.2 Connection and Cursor Objects**

* To solve the data type mismatch problem, the concept of *cursor* is usually used.
* A cursor allows the programmer to *iterate* through the result set, one row at a time.
* Cursor reference of PyMySQL: <http://pymysql.readthedocs.io/en/latest/modules/cursors.html>
* There may be different cursor types for performance, security, and access method considerations.
* PyMySQL's cursor types:
	1. SSCursor: unbuffered cursor
	2. DictCursor: return a dictionary
	3. SDictCursor: an unbuffered cursor returning a dictionary

***Example:***

[student1.py](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2025/notes/emSQL/student1.py.txt): (All .py and .ini files have an extra extension, .txt, which should be removed when downloading).

import pymysql

# [1] Making connection to the MySQL server
cnx = pymysql.connect(user='demo', password='abcdef',
    host='localhost',
    database='toyu')

# Create a cursor using the connection.
cursor = cnx.cursor()

# [2] Prepare a SQL query for the problem
query = '''
SELECT CONCAT (s.fname, ' ', s.lname) AS student,
    d.deptName,
    CONCAT(f.fname, ' ', f.lname) as advisor
FROM student AS s LEFT JOIN department AS d
        ON (s.major = d.deptCode)
     LEFT JOIN faculty AS f
        ON (s.advisor = f.facId);
'''

# Execute the query
cursor.execute(query)

# [3] Use the result in the query
for (student, major, advisor) in cursor:
    print("{}: major={}; advisor={}".format(student, major, advisor))

# [4] Housekeeping
cursor.close()
cnx.close()

The output of running this program:

C:\...>python student1.py
Tony Hawk: major=Computer Science; advisor=Paul Smith
Mary Hawk: major=Computer Science; advisor=Paul Smith
David Hawk: major=Computer Science; advisor=Mary Tran
Catherine Lim: major=Information Technology; advisor=None
Larry Johnson: major=Information Technology; advisor=Deborah Gump
Linda Johnson: major=Computer Information Systems; advisor=Daniel Kim
Lillian Johnson: major=Computer Information Systems; advisor=Andrew Byre
Ben Zico: major=None; advisor=None
Bill Ching: major=Arts; advisor=None
Linda King: major=Arts; advisor=Art Allister
Cathy Johanson: major=None; advisor=Art Allister

* Some tips:
	1. Test your SQL statements thoroughly first.
	2. Use typical input parameters for testing.
	3. Be mindful of:
		1. security, especially SQL injection
		2. special characters of the languages involved
		3. performance

**3.3 Configuration files**

* It is better to use configuration module and data ini file (e.g. dbconfig.py and dbconfig.ini). Why?
	1. Separation of data and program
	2. Reuse and maintenance
	3. Security
* In general, using a configuration data file is a good programming practice.

[dbconfig.py](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2025/notes/emSQL/dbconfig.py.txt):

import configparser

#  simplistic and no error handling.
def get\_mysql\_param(filename='dbconfig.ini', section='mysql'):

    config = configparser.ConfigParser()
    config.read(filename)

    return config[section]

If running in Windows IIS (such as in the dcm server):

import configparser
from pathlib import Path

#  simplistic and no error handling.
def get\_mysql\_param(filename='dbconfig.ini', section='mysql'):
    config = configparser.ConfigParser()
    file\_path = (Path(\_\_file\_\_).parent / filename).resolve()
    config.read(file\_path)

    return config[section]

[dbconfig.ini](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2025/notes/emSQL/dbconfig.ini.txt):

[mysql]
host = localhost
port = 3306
database = toyu
user = your\_account
password = your\_password

In your Python program:

from dbconfig import \*
import pymysql

db = get\_mysql\_param()
cnx = pymysql.connect(user=db['user'],
      password=db['password'],
      host=db['host'],
      database=db['database'],
      port=int(db['port']))
cursor = cnx.cursor()

If you use the default port (i.e., 3306), you may adjust by

1. remove the line "port=3306" in dbconfig.ini, and
2. remove the argument "port" in the call of pymysql.connect() in the main program.

In your Python program,

[student2.py](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2025/notes/emSQL/student2.py.txt):

from dbconfig import \*
import pymysql

db = get\_mysql\_param()

print (db)
cnx = pymysql.connect(user=db['user'], password=db['password'],
      host=db['host'],
      database=db['database'],
      port=int(db['port']))

cursor = cnx.cursor()

# [2] Prepare a SQL query for the problem
query = '''
SELECT CONCAT (s.fname, ' ', s.lname) AS student,
    d.deptName,
    CONCAT(f.fname, ' ', f.lname) as advisor
FROM student AS s LEFT JOIN department AS d
        ON (s.major = d.deptCode)
     LEFT JOIN faculty AS f
        ON (s.advisor = f.facId);
'''

# Execute the query
cursor.execute(query)

# [3] Use the result in the query
for (student, major, advisor) in cursor:
    print(f"{student}: major={major}; advisor={advisor}")

# [4] Housekeeping
cursor.close()
cnx.close()

* See Format Specification Mini-Language: <https://docs.python.org/3.11/library/string.html#formatspec>
* Your SQL statement may use placeholder parameters: %s.
* Parametrized statements are preferred.

Example from <http://dev.mysql.com/doc/connector-python/en/connector-python-api-mysqlcursor-execute.html>:

insert\_stmt = (
  "INSERT INTO employees (emp\_no, first\_name, last\_name, hire\_date) "
  "VALUES (%s, %s, %s, %s)"
)
data = (2, 'Jane', 'Doe', datetime.date(2012, 3, 23))
cursor.execute(insert\_stmt, data)

select\_stmt = "SELECT \* FROM employees WHERE emp\_no = %(emp\_no)s"
cursor.execute(select\_stmt, {'emp\_no': 2})

***Example:***

Get the school code as the command line argument.

[student3.py](https://dcm.uhcl.edu/yue/courses/joinDB/Spring2025/notes/emSQL/student3.py.txt)

from dbconfig import \*
import pymysql
import sys

# Use the school code in command line arguments
# to list all students majoring in a department
# in the school.
# [1] Making connection to the MySQL server

db = get\_mysql\_param()
cnx = pymysql.connect(user=db['user'], password=db['password'],
      host=db['host'],
      database=db['database'],
      port=int(db['port']))
cursor = cnx.cursor()

# [2] Prepare a SQL query for the problem
query = '''
SELECT CONCAT (s.fname, ' ', s.lname) AS student,
    d.deptName,
    CONCAT(f.fname, ' ', f.lname) as advisor
FROM student AS s LEFT JOIN department AS d
        ON (s.major = d.deptCode)
     LEFT JOIN faculty AS f
        ON (s.advisor = f.facId)
WHERE d.schoolCode = %s;
'''

# [2b] Get input values
school\_code = 'CSE'
if len(sys.argv) > 1:
    school\_code = sys.argv[1]

# Execute the query
cursor.execute(query, school\_code)

# [3] Use the result in the query
for (student, major, advisor) in cursor:
    print(f"{student}: major={major}; advisor={advisor}")

# [4] Housekeeping
cursor.close()
cnx.close()

Result:

C:\...>python student3.py CSE
Linda Johnson: major=Computer Information Systems; advisor=Daniel Kim
Lillian Johnson: major=Computer Information Systems; advisor=Andrew Byre
Tony Hawk: major=Computer Science; advisor=Paul Smith
Mary Hawk: major=Computer Science; advisor=Paul Smith
David Hawk: major=Computer Science; advisor=Mary Tran
Catherine Lim: major=Information Technology; advisor=None
Larry Johnson: major=Information Technology; advisor=Deborah Gump

Notes:

* %s is used as a placeholder.