**CSCI 4333.2 Classroom Notes and Demonstrations**

8/20/2025

**Introduction to the course  
CSCI 4333**

by K. Yue

**1. Promotion**

* This course is (hopefully) one of the more useful CS/DS courses for students.
* World data is estimated to double every two years.

**2. How to be successful in the course**

General Course Suggestions:

1. Course expectation is demanding.
2. Please consider forming the habit of listening carefully and asking a lot of questions.

General Professionalism:

1. Attitude
2. Be considerate.
3. Be helpful and useful to others.
4. Be a good listener.
5. Be responsive.
6. Hardworking
7. Attention to details.
8. Focus: uni-tasking

Some general tips:

1. Engagement: Participate. Ask questions, a lot of them. Help others. Plan ahead.
2. Preparation: start as early as possible and do not fall behind.
3. Don’t copy and paste. Instead, copy, integrate, and apply.
4. [SEE-I](https://en.everybodywiki.com/SEE-I): State, Elaborate, Exemplify and illustrate.
5. Form good habits.

Some good traits of Computer and Data Scientists:

1. Habits of trying to make sense of stuff.
2. Intellectual curiosity.
3. Tinkering and experimentation.
4. Open-minded, not dogmatic.
5. A large tool set.

**3. Resources**

* Companion materials of our textbook: please consult the course page in Canvas for additional resources related to the textbook.
* Contents of the course will be based mostly on
  1. Lecture notes posted in the course website: [../../index.html](https://dcm.uhcl.edu/yue/courses/joinDB/Fall2025/index.html).
  2. Classroom demonstrations.
  3. Assignments.
* Please read the appropriate pages in the textbook and lecture notes in this site *before* coming to the class.
* Document your learning. Bring a notebook to the class. Otherwise, it may be a good idea to print out the notes and bring them to the class so you can make notes during the class.

**4. Introduction**

* *Persistent* data is the backbone of many applications.
  1. Data remains after software closes.
  2. Examples: profiles in social media, Instagram. In cloud, in files, in DBMS.
  3. Examples of non-persistent (transient) data: stored in RAM, registers, … variable values in Java prorams.
* Three main choices of storing persistent data:
  1. Files
  2. Databases: focus of this course.
  3. Cloud-based storage and database.
* Some advantages of DBMS (according to Ricardo, the optional textbook of this class):
  1. Sharing of data
  2. Control of redundancy
  3. Data consistency
  4. Improved data standards
  5. Better data security
  6. Improved data integrity
  7. Balance of conflicting requirements: DBMS allows the specification of problem requirements much better than file systems.

Requirements:

Bun Yue’s Bday is 6/14/1995

2/30/1988?

8/abc/hkd

1/4/1780

* 1. Faster development of new applications
  2. Better data accessibility
  3. Economy of scale: larger scales implies better efficiency.
     + USA better than smaller countries in certain things.
  4. Scaling well:
     + Size of problem get bigger, solution is still effective.
  5. More control over concurrency:
     + File systems allow very little concurrency to preserve data integrity.
     + DBMS allows more concurrency with nuanced access control.
  6. Better backup and recovery procedures
* How do we *make sense* of these 12 different advantages?
  1. Different textbooks may have different collections of the advantages of DBMS because of different classifications.
  2. No need to memorize them.
  3. Better to assimilate them and construct your own list.
  4. Make your own notes. Use [SEE-I](https://en.everybodywiki.com/SEE-I) (In your own words, state, elaborate, and exemplify with examples, and illustrate the concept.)
     + However, do not overuse metaphors.
  5. Learning through documentation, communications, and teaching.
* What are some disadvantages of DBMS?
  1. Complexity
  2. Cost
  3. Learning curve
  4. Possible single points of failure and bottleneck

1. **A Simple Introduction to the Relational Model**

**Installing XAMPP: for development.**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* Relational databases are the most popular databases: <https://db-engines.com/en/ranking>. It is based on the relational model.
* There are many other data models.
* In layman's term: A *table* (relation) is the basic unit of a relational database.
* A table is composed of many *rows* (tuples).
* Each row has many *column* (attribute) values.

A screenshot of a computer

AI-generated content may be incorrect.

* A primary key is roughly a *minimal* set of columns in a table that*uniquely identify* a row.
* Two tables can be related to each other by *foreign keys*. A foreign key is roughly a column in a table in which its value must be equal to the referenced value of the primary key in another table (called the paren or referenced table).
* Relational DBMS is the most popular DBMS. Examples:
  + DB-engine ranking: <https://db-engines.com/en/ranking>
  + Top 10 DBMS in Data Science: <https://towardsdatascience.com/top-10-databases-to-use-in-2021-d7e6a85402ba>
* SQL is the 'glue' in many DB systems.

***Classroom discussion***

Please ask questions about the toy University DB ([toyu](https://dcm.uhcl.edu/yue/courses/joinDB/Fall2025/notes/toyu/toyu.html))