CSIE30600/CSIEB0290 Database Systems

Shiow-yang Wu
Department of Computer Science and Information Engineering
National Dong Hwa University

Happy Lantern Festival (元宵節)

- Celebrating the **Lantern Festival** with sky lantern.

Course Objectives

- First course in database systems
- Cover the fundamental concepts
- Learn to use Database Management System (DBMS).
- Study the internals of DBMS (for a CS student)
- Learning state-of-the-art open source DBMS
- **Advanced topics 1**: Big Data, NoSQL, NewSQL, Distributed SQL, IoT Streaming Data, DB for AI
- **Advanced topics 2**: Cloud DB, mobile DB, XML DB, OODB, multimedia DB, parallel/distributed DB …

Course Information

- Course Title: Database Systems
- Course Number: CSIE30600/CSIEB0290
- Meeting Time: Thu 14:10~17:00
- Classroom: Science/Engineering Building II C303
- Instructor's Office: Sci/Eng Building II C308
- Office Hours: Thu 17:00 - 18:00
- Phone Number: (03) 8693020
- Email Address: **showyang@gms.ndhu.edu.tw**

Grading Policy

- Assignments 25%
- Midterm 25%
- Final Exam 25%
- Term project 25%

- (may change if necessary)

Web Pages

- Course web page: http://web.csie.ndhu.edu.tw/showyang/DB2024s/index.html
- **Not on "e學苑 - e-Learning@NDHU" !!**
- Instructor's homepage: http://web.csie.ndhu.edu.tw/showyang

Online Class (just in case)

- **Teams link**: https://teams.microsoft.com/l/meetup-join/19%3az6QLI1m3Hd8BHpPW0mDJpJjiS_-KLszoUlMPJZdrivc1%40thread.tacv2/1658531135758?context=%7b%22Tid%22%3a%22edba3211-8174-4411-b089-357c588fa127%22%2c%22Oid%22%3a%22e83708da-2e73-4b78-a037-e2bbca1f4d94%22%7d
- Join by ID:
  - Meeting ID: 440 956 272 439
  - Passcode: RkiWc9

Textbooks

- Avi Silberschatz, Henry F. Korth and S. Sudarshan. ***Database System Concepts, 7th Edition***. McGraw-Hill, 2019/2020. (https://www.db-book.com/)(DBSC7)
- Ramez Elmasri and Shamkant B. Navathe. ***Fundamentals of Database Systems, 7th Edition***. Pearson, 2016. (FDBS7) (https://www.pearson.com/us/higher-education/program/Elmasri-Fundamentals-of-Database-Systems-7th-Edition/PGM189052.html)
- Recommended but not required.

References

- C. J. Date. ***Database Design and Relational Theory: Normal Forms and All That Jazz, 2nd Edition***. Apress, 2019.
- Carlos Coronel and Steven Morris. ***Database Systems: Design, Implementation, & Management, 13th Edition***. Cengage Learning, 2018.
- Wilfried Lemahieu, Bart Baesens and Seppe vanden Broucke. ***Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data***. Cambridge University Press, 2018.
- Thomas Connolly and Carolyn Begg. ***Database Systems- A Practical Approach to Design, Implementation, and Management, 6th Edition***. Pearson, 2015.
- Garcia-Molina, J. D. Ullman, and J. Widom. ***Database Systems: The Complete Book, 2nd Edition***, Prentice Hall, 2008. (http://infolab.stanford.edu/~ullman/dscb.html)
- Jeffrey D. Ullman and Jennifer Widom. ***A First Course in Database Systems, 3rd Edition***, Prentice Hall, 2007. (http://infolab.stanford.edu/~ullman/fcdb.html)

SQL References

- Alan Beaulieu. ***Learning SQL: Generate, Manipulate, and Retrieve Data, 3rd Edition***. O'Reilly Media, 2020.
- Anthony Molinaro and Robert de Graaf. ***SQL Cookbook: Query Solutions and Techniques for All SQL Users 2nd Edition***. O'Reilly Media, 2020.
- Upom Malik, Matt Goldwasser, Benjamin Johnston. ***SQL for Data Analytics: Harness the power of SQL to extract insights from data, 3rd Edition***. Packt Publishing, 2022.
- Bobby Iliev. ***Introduction to SQL***. MIT License, 2020.
- GoalKicker.com. ***SQL Notes for Professionals***. (good desktop ref) (https://books.goalkicker.com/SQLBook/)

SQL References (cont.)

- Mike McGrath. ***SQL in Easy Steps, 4th Edition***. In Easy Steps Ltd. 2020.
- Kevin Kline, Daniel Kline and Brand Hunt. ***SQL in a Nutshell: A Desktop Quick Reference, 4th Edition.*** O'Reilly Media, Inc. 2020.
- Ben Forta. ***SQL in 10 Minutes a Day, Sams Teach Yourself, 5th Edition***. Sams Publishing, 2019.
- James R. Groff, Paul N. Weinberg, Paul Weinberg, James Groff. ***SQL: The Complete Reference, 3rd Edition***. McGraw-Hill, 2009.
- Alex Kriegel and Boris M. Trukhnov. ***SQL Bible, 2nd Edition***. Wiley, 2008.

PHP and MySQL References

- Robin Nixon. ***Learning PHP, MySQL & JavaScript, 7th Edition***. O'Reilly Media, Inc., 2024.
- Jon Ducket. ***PHP & MySQL: Server-side Web Development***. Wiley, 2022.
- W. J. Gilmore. ***Beginning PHP and MySQL: From Novice to Professional, 5th Edition***, Apress, 2018.
- Andrew Comeau and Stephen Burge. ***MySQL Explained: Your Step By Step Guide, 2nd Edition***, CreateSpace Independent Publishing Platform, 2017.
- Luke Welling and Laura Thomson. ***PHP and MySQL Web Development, 5th Edition***, Addison-Wesley Professional, 2016.
- Paul DuBois. ***MySQL, 5th Edition (Developer's Library).*** Addison-Wesley Professional, 2013.

Python Programming Books

- Eric Matthes. *Python Crash Course: A Hands-On, Project-Based Introduction to Programming, 3rd Edition*. No Starch Press, 2023.
- Steve Holden, Anna Ravenscroft and Alex Martelli. *Python in a Nutshell: A Desktop Quick Reference, 4th Edition*. O'Reilly Media, 2023.
- Johannes Ernesti and Peter Kaiser. *Python 3: The Comprehensive Guide to Hands-On Python Programming*. Rheinwerk Computing, 2022.
- Brett Slatkin. *Effective Python: 135 Specific Ways to Write Better Python, 3rd Edition*. Addison-Wesley Professional, 2024.
- Luciano Ramalho. *Fluent Python: Clear, Concise, and Effective Programming, 2nd Edition*. O'Reilly Media, 2022.
- Wes McKinney. *Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd edition*. O'Reilly Media, 2022.
- A Byte of Python(free online book)(https://python.swaroopch.com/)

On-line References

- Wikibooks, **Structured Query Language**. (https://en.wikibooks.org/wiki/Structured_Query_Language) (SQL:2011)
- Wikibooks, **MySQL**. (https://en.wikibooks.org/wiki/MySQL)
- Wikibooks, **PostgreSQL** (https://en.wikibooks.org/wiki/PostgreSQL)

On-line SQL Resources

- **Online SQL interpreter (for the DB Concepts book)**

(https://www.db-book.com/university-lab-dir/sqljs.html)

- **SQLite online**

(https://sqliteonline.com/)

- **The Try-SQL Editor (W3Schools)**

(https://www.w3schools.com/sql/trysql.asp?filename=trysql_asc)

- **Online SQL Compiler (Tutorialspoint)**

(https://www.tutorialspoint.com/execute_sql_online.php)

Individual Term Project

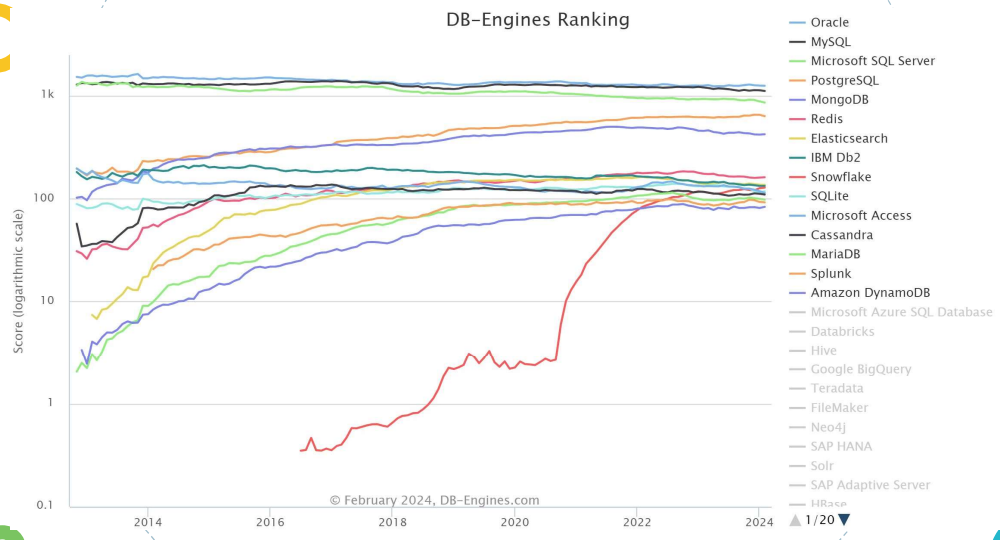
- An on-line database application (details will be announced in class)
- Use an open source DBMS as backend
- Use browser or smart phone as user interface
- Can use any technique to connect the database.
- We will discuss PHP+MySQL and/or Python+MySQL/PostgreSQL.
- Demonstration/report due date: June 20, 2024.

Why MySQL, PostgreSQL?

417 systems in ranking, February 2024

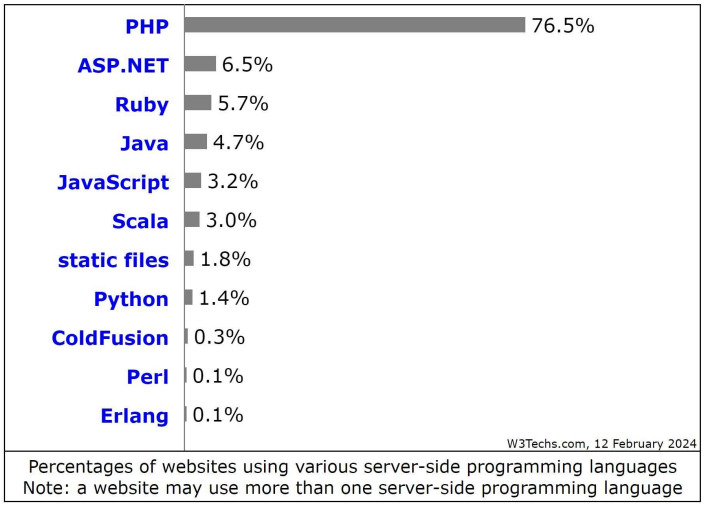
Rank Feb 2024	Rank Jan 2024	Rank Feb 2023	DBMS	Database Model	Score Feb 2024	Score Jan 2024	Score Feb 2023
1.	1.	1.	Oracle	Relational, Multi-model	1241.45	-6.05	-6.08
2.	2.	2.	MySQL	Relational, Multi-model	1106.67	-16.79	-88.78
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	853.57	-23.03	-75.52
4.	4.	4.	PostgreSQL	Relational, Multi-model	629.41	-19.55	+12.90
5.	5.	5.	MongoDB	Document, Multi-model	420.36	+2.88	-32.41
6.	6.	6.	Redis	Key-value, Multi-model	160.71	+1.33	-13.12
7.	7.	↑ 8.	Elasticsearch	Search engine, Multi-model	135.74	-0.33	-2.86
8.	8.	↓ 7.	IBM Db2	Relational, Multi-model	132.23	-0.18	-10.74
9.	9.	↑ 12.	Snowflake	Relational	127.45	+1.53	+11.80
10.	↑ 11.	↓ 9.	SQLite	Relational	117.28	+2.08	-15.38
11.	↓ 10.	↓ 10.	Microsoft Access	Relational	113.17	-4.50	-17.86
12.	12.	↓ 11.	Cassandra	Wide column, Multi-model	109.27	-1.77	-6.95
13.	13.	13.	MariaDB	Relational, Multi-model	97.23	-2.00	+0.42
14.	14.	14.	Splunk	Search engine	91.65	-1.07	+4.57
15.	↑ 16.	15.	Amazon DynamoDB	Multi-model	82.90	+1.96	+3.21
16.	↓ 15.	16.	Microsoft Azure SQL Database	Relational, Multi-model	79.56	-1.51	+0.81
17.	17.	↑ 19.	Databricks	Multi-model	76.91	-3.62	+16.58
18.	18.	↓ 17.	Hive	Relational	65.81	-1.15	-6.31
19.	19.	↑ 22.	Google BigQuery	Relational	63.63	+0.15	+11.17
20.	20.	↓ 18.	Teradata	Relational, Multi-model	51.24	-1.94	-11.79

Why MySQL, PostgreSQL?



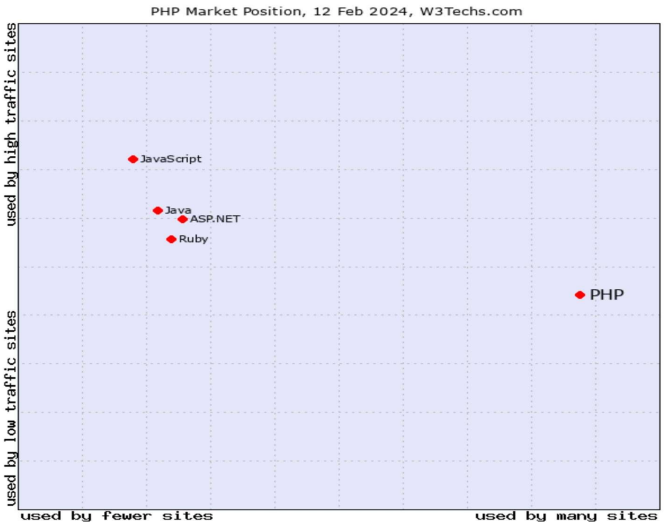
Why PHP ?

- Most popular **server-side** programming languages (https://w3techs.com/)

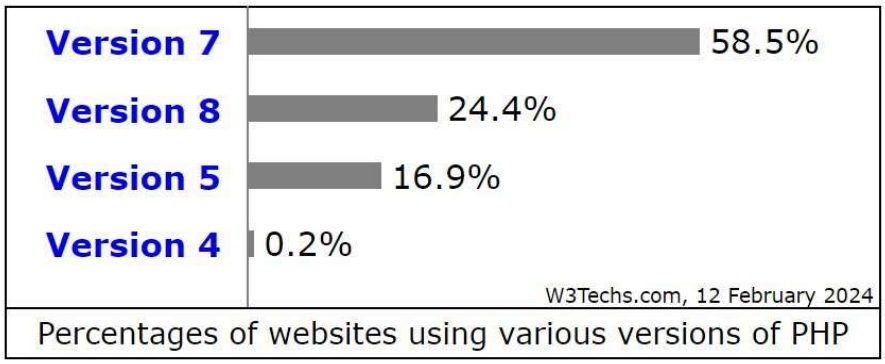


Percentages of websites using various server-side programming languages
Note: a website may use more than one server-side programming language

PHP Market Position



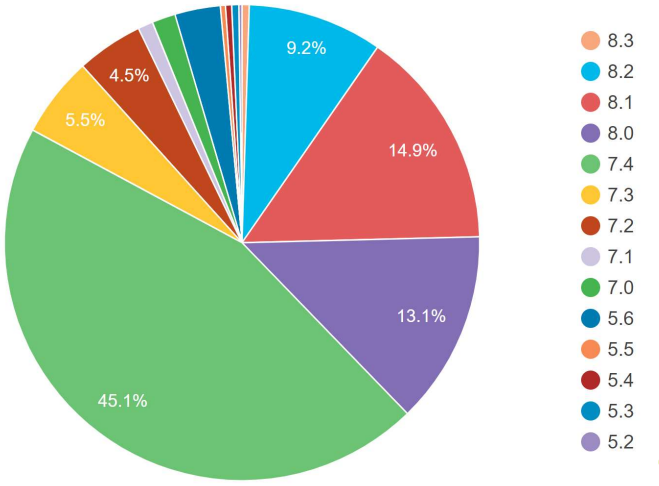
Which Version of PHP?



Percentages of websites using various versions of PHP

Versions of PHP ?

- Wordpress PHP version stats
- Should use 8.0 ↑
- At least 7.4



(https://wordpress.org/about/stats/)

Why Python ?

TIOBE Index

Feb 2024	Feb 2023	Change	Programming Language	Ratings	Change
1	1		Python	15.16%	-0.32%
2	2		C	10.97%	-4.41%
3	3		C++	10.53%	-3.40%
4	4		Java	8.88%	-4.33%
5	5		C#	7.53%	+1.15%
6	7	↑	JavaScript	3.17%	+0.64%
7	8	↑	SQL	1.82%	-0.30%
8	11	↑	Go	1.73%	+0.61%
9	6	↓	Visual Basic	1.52%	-2.62%
10	10		PHP	1.51%	+0.21%

PYPL Index

Worldwide, Feb 2024 :

Rank	Change	Language	Share	1-year trend
1		Python	28.11 %	+0.6 %
2		Java	15.52 %	-1.0 %
3		JavaScript	8.57 %	-1.0 %
4	↑	C/C++	6.92 %	+0.1 %
5	↓	C#	6.73 %	-0.1 %
6	↑	R	4.75 %	+0.7 %
7	↓	PHP	4.57 %	-0.6 %
8		TypeScript	2.78 %	-0.0 %
9		Swift	2.75 %	+0.5 %
10		Objective-C	2.37 %	+0.1 %

What about JavaScript ?

- Most popular client-side programming languages (https://w3techs.com/)

© W3Techs.com	usage	change since 1 January 2024
1. JavaScript	98.9%	+0.1%
2. Flash	1.2%	

percentages of sites

PHP vs JavaScript (1)

PHP	JavaScript
• Server-side scripting language	• Client-side scripting language
• Used for back-end development	• Mainly used for front-end development
• More secure (as is not visible in browser)	• Has tools for enhancing security but needs more effort to do so
• Helps to build high-level interactive web pages	• Helps to build user-friendly creative web pages
• Quite slow performance	• Fast performance
• More features available	• Less load on a server and less server traffic
• Combined with HTML	• Combined with HTML, XML, Ajax
• MariaDB, MySQL, and PostgreSQL;	• AngularJS and ReactJS:
• WordPress, Drupal, Joomla	• MongoDB, CouchDB, and NoSQL
• Best for e-commerce and other websites using CMS	• Best for dynamic SPAs

PHP vs JavaScript (2)

PHP VS JAVASCRIPT: KEY DIFFERENCES

Features	PHP	Javascript
Server-side language	Yes	Not without additional frameworks
Client-side language	No	Yes
OOP	Yes	Yes
Supports database	Yes	No
Open source	Yes	Yes
Performance	Comparatively slow	Fast
Works within browser	No	Yes
Garbage collection	Yes	Yes
Interchangeable objects and arrays	No	Yes
Accepts lower and uppercase variables	Yes	No

DB for JavaScript

- As a client-side language, built-in DB support is not a primary concern.
- Can still use embed DBs with small footprint.
- Oracle provides JavaScript support in the MySQL to write JavaScript stored programs. (2023/12/15)

JavaScript for Server-side?

Fastest growing server-side programming languages since 1 January 2024

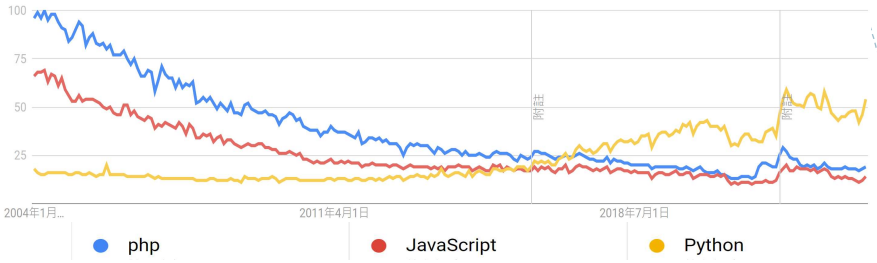
© W3Techs.com	sites
1. Ruby	31.8
2. JavaScript	13.6
3. Java	10.5

daily increase of number of sites per million

Need environment/frameworks such as Node.js, Express.js, … (out of the scope of this course)

Is PHP dying ?

- Google Trend



- Still used by > 76% of the websites (W3tech)
- Python and JavaScript/frameworks are rising.
- Will be invincible if you learn all THREE !!

Why Study Databases?

- Databases used to be *specialized applications*, now they are a *central component* in modern systems.
  - Knowledge of database concepts is essential for computer scientists
- Databases are everywhere, even when you don't see them explicitly
  - most activities involve data
    - Banking + credit cards: all transactions
    - Airlines: reservations, schedules
    - Universities: registration, grades
    - Telecommunications/networks

(more on next slide)

Why Study Databases?

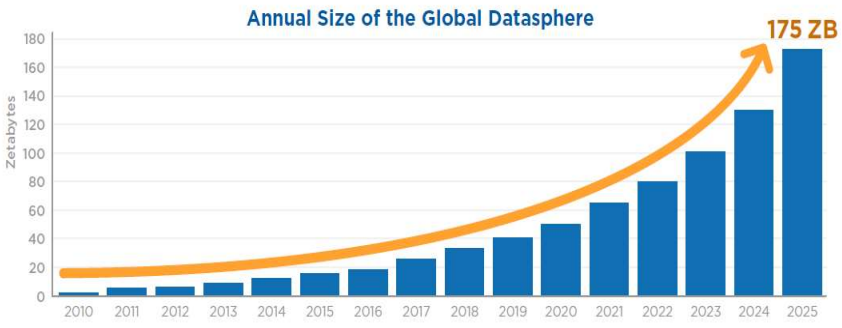
- Sales: customers, products, purchases
- Manufacturing: production, inventory, orders, supply chain
- Human resources: employee records, salaries, tax deductions
- Web sites: generated from databases; front-ends to databases
- Scientific research, e.g., studying the environment
- Your own data!

- Global data volume grows faster than ever! (next slide)
- Sky-high demand for Big data and NoSQL/NewSQL DB!
- Streaming data from IoT devices
- Data needs to be *managed*

Global DataSphere

- DataSphere: the sum of ALL data around the world!
- Global DataSphere will grow from 33ZB in 2018 to **175 ZB** by 2025 (IDC)

Figure 1 - Annual Size of the Global Datasphere



Why Study Databases?

- Because data is **valuable**:
  - E.g., bank account records, tax records, student records, personal information …
  - It must be protected - no matter what happens whether we have machine crashes, disk crashes, hurricanes/floods;
  - It also needs to be protected from people

Why Study Databases?

- Because data is often structured:
  - Bank account records all follow the same structure
  - We can exploit this regular structure
    - To retrieve data in effective ways (that is, we can use a query language)
    - To store data efficiently
- Dealing with unstructured data still needs database technologies.
- Big data needs database + new techniques

Why Study Databases?

- Because the database field has made significant contributions to basic computer science:
  - Concepts and techniques have been applied to different problems and different areas
- Because DBMS software is highly successful as a commercial technology (Oracle, SQL Server, …)
- Because DB research is highly active and VERY interesting!
  - Lots of opportunities to have practical impact

Syllabus

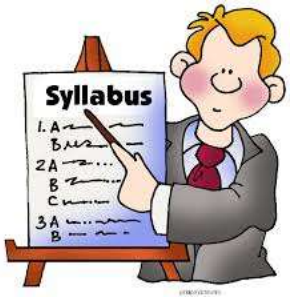
- Introduction
- Databases and database users
- Database system concepts and architecture
- Relational model and constraints
- Relational algebra (and calculus**)
- Basic & intermediate SQL (how to use a DB)
- Database design with ER/EER models
- ER/EER to relational mapping
- Open source RDBMS(MySQL, PostgreSQL, …)
- Web DB applications(with PHP, Python, …)

Syllabus (cont.)

- Relational database design I – Functional dependencies and normalization
- Relational database design II – Further normalization and design algorithms
- Big data processing**
  - Basic concepts, big data storage, MapReduce
  - NoSQL/NewSQL/Distributed SQL, graph databases
- Big data analytics**
  - Data warehousing
  - Online/realtime analytical processing
  - Data mining

Syllabus (cont.)

- Complex data types**
- Storage systems and structure**
- Indexing methods**
- Query processing & optimization
  - Query processing
  - Query optimization
- Transaction management
  - Transactions
  - Concurrency control
  - Recovery**



Syllabus (cont.)

- Object and object-relational databases**
- Semi-structured data and XML**
- Web databases

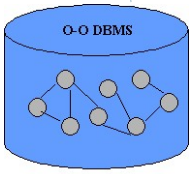


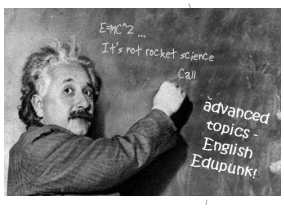
Figure 1: O-O Database Structure

Syllabus (cont.)**

- Parallel and distributed databases
- Cloud computing and data trends
- NoSQL and NewSQL databases
- Blockchain databases

Syllabus (Advanced Topics)**

- Database security
- Active databases
- Temporal and real-time databases
- Spatial databases
- Multimedia databases
- Deductive databases
- Information retrieval and Web search
- Mobile and pervasive data management
- Streaming data management/analytics



Accept the DB Challenges

- A very interesting and challenging class
- Be prepared for some theoretical discussion on the principles and algorithms.
- With homework and assignments.
- Don't be late for the class. You will NOT be able to keep up with the pace.
- Ask questions if you miss the point.
- Design your own examples.
- Join us on the DB challenges !!!