CSIE30600/CSIEB0290 Database Systems

Shiow-yang Wu

Department of Computer Science and
Information Engineering
National Dong Hwa University

Course Objectives

- First course in database systems
- Cover the fundamental concepts
- Using Database Management System (DBMS).
- Learning state-of-the-art open source DBMS
- Study the underline implementation of DBMS
- **Advanced topics**: Cloud DBs and NoSQL for Big Data (topics for Big Data Systems course)*
- **Advanced topics**: XML DB, OODB, mobile DB, multimedia DB, parallel/distributed DB ...

Course Information

- Instructor's Office: Eng Building C308
- Office Hours: Tue 17:00 - 18:00
- Phone Number: (03) 8693020
- Email Address: **showyang@gms.ndhu.edu.tw**
- Grading Policy: (may change if necessary)
  - Assignments   25%
  - Midterm   25%
  - Final Exam   25%
  - Term project   25%

CSIE30600/CSIEB0290 Database Systems   Course Information 3

Web Pages

- Course webpage:
  http://web.csie.ndhu.edu.tw/showyang/DB2019s/index.html
- Not on "東華e學苑" !!
- Instructor's homepage:
  http://web.csie.ndhu.edu.tw/showyang

CSIE30600/CSIEB0290 Database Systems   Course Information 4

Textbooks

- Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems, 7th Edition*. Pearson, 2016. (https://www.pearson.com/us/higher-education/program/Elmasri-Fundamentals-of-Database-Systems-7th-Edition/PGM189052.html)

- Avi Silberschatz, Henry F. Korth and S. Sudarshan. *Database System Concepts, 7th Edition*. McGraw-Hill, 2019. (https://www.db-book.com/db7/)

- Recommended but not required.

References

- Thomas Connolly and Carolyn Begg. *Database Systems- A Practical Approach to Design, Implementation, and Management, 6th Edition*. Pearson, 2015.

- Abraham Silberschatz, Henry F. Korth, and S. Sudarshan. *Database System Concepts, 6th Edition*. McGraw-Hill, 2011.

- Garcia-Molina, J. D. Ullman, and J. Widom. *Database Systems: The Complete Book, 2nd Edition*, Prentice Hall, 2008.

- Jeffrey D. Ullman and Jennifer Widom. *A First Course in Database Systems, 3rd Edition*, Prentice Hall, 2007.

SQL References

- Andrew Johansen. *SQL: The Ultimate Beginner's Guide!* CreateSpace Independent Publishing Platform, Nov 2015.
- John Viescas and Michael J. Hernandez. *SQL Queries for Mere Mortals: A Hands-On Guide to Data Manipulation in SQL (3rd Edition).* Addison-Wesley Professional, Jun 2014.
- Joe Celko. *Joe Celko's SQL for Smarties, 5th Edition: Advanced SQL Programming*. Morgan Kaufmann. Dec 2014.
- Stephane Faroult. *SQL Success – Database Programming Proficiency*. RoughSea Ltd, 2013.
- Ben Forta. *SQL in 10 Minutes, Sams Teach Yourself, 4th Edition*. Sam Publishing, Nov 2012.
- Mike McGrath. *SQL in Easy Steps, 3rd Edition*. In Easy Steps Ltd. 2012.

SQL References (cont.)

- Alan Beaulieu. *Learning SQL, 2nd Edition.* O'Reilly Media, Inc. 2009.
- James R. Groff, Paul N. Weinberg, Paul Weinberg, James Groff. *SQL: The Complete Reference, 3rd Edition*. McGraw-Hill, 2009.
- Kevin Kline, Daniel Kline and Brand Hunt. *SQL In A Nutshell, 3rd Edition.* O'Reilly Media, Inc. 2008.
- Alex Kriegel and Boris M. Trukhnov. *SQL Bible, 2nd Edition.* Wiley, 2008.

PHP and MySQL References

- Luke Welling and Laura Thomson. ***PHP and MySQL Web Development, 5th Edition***, Addison-Wesley Professional, 2016.
- W. J. Gilmore. ***Beginning PHP and MySQL: From Novice to Professional, 5th Edition***, Apress, 2016.
- Mr Andrew Comeau and Stephen Burge. ***MySQL Explained: Your Step By Step Guide***, CreateSpace Independent Publishing Platform, Nov 2015.
- Paul DuBois. ***MySQL, 5th Edition (Developer's Library).*** Addison-Wesley Professional, 2013.
- Alan Forbes. ***The Joy of PHP: A Beginner's Guide to Programming Interactive Web Applications with PHP and mySQL, 2nd Edition***, Plum Island Publishing LLC, 2013.
- Larry Ullman. ***PHP and MySQL for Dynamic Web Sites: Visual QuickPro Guide, 4th Edition***, Peachpit Press, 2011.

Python Programming Books

- Steve Holden, Anna Ravenscroft and Alex Martelli. *Python in a Nutshell, 3rd Edition*. O'Reilly Media, Inc. 2017.
- Mark Lutz. *Learning Python, 5th Edition*. O'Reilly Media, 2013.
- David Beazley and Brian K. Jones. *Python Cookbook, 3rd edition*. O'Reilly Media, 2013.
- Luciano Ramalho. *Fluent Python*. O'Reilly Media, 2015.

On-line Resources

- 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)
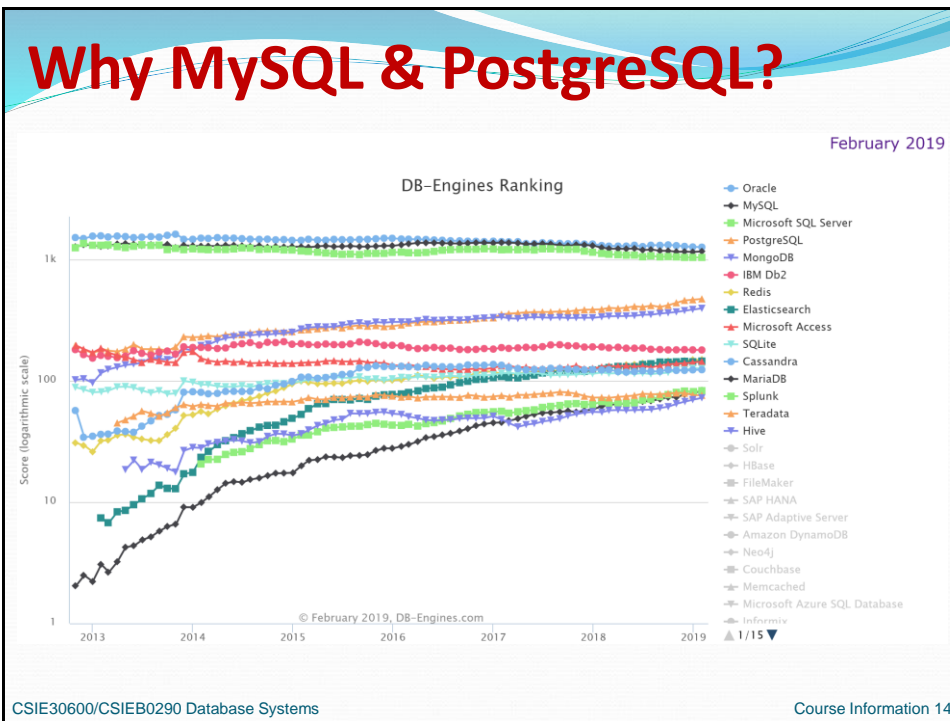
Individual Term Project

- An on-line database application (details will be announced in class)
- Use an open source DB as backend database
- Use browser or smart phone as user interface
- Can use any technique to connect the database.
- We will discuss PHP+MySQL and/or Python+PostgreSQL.
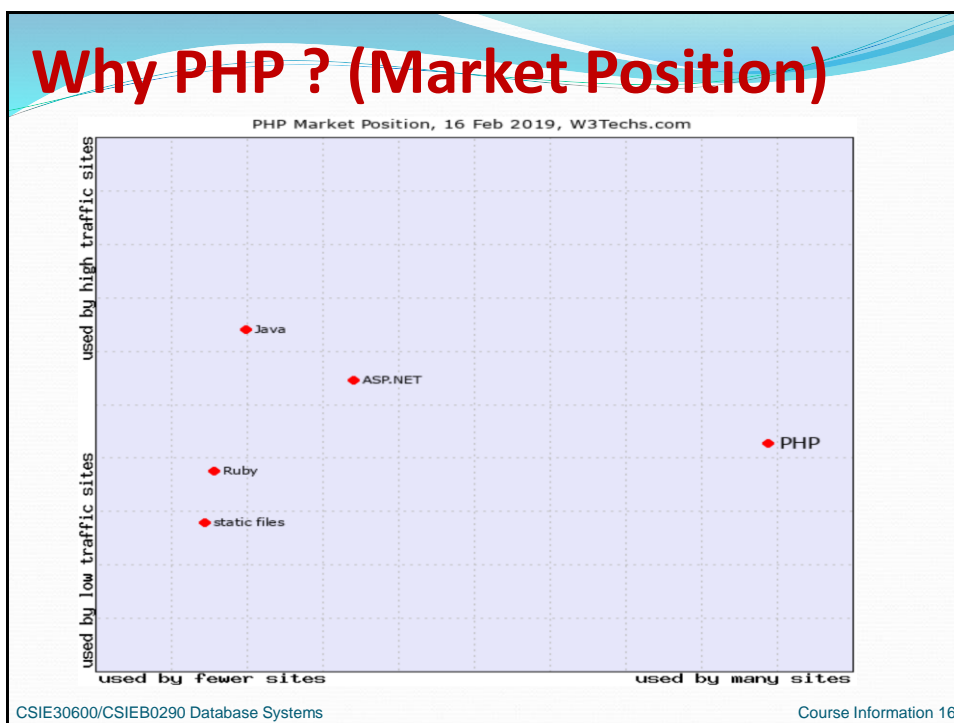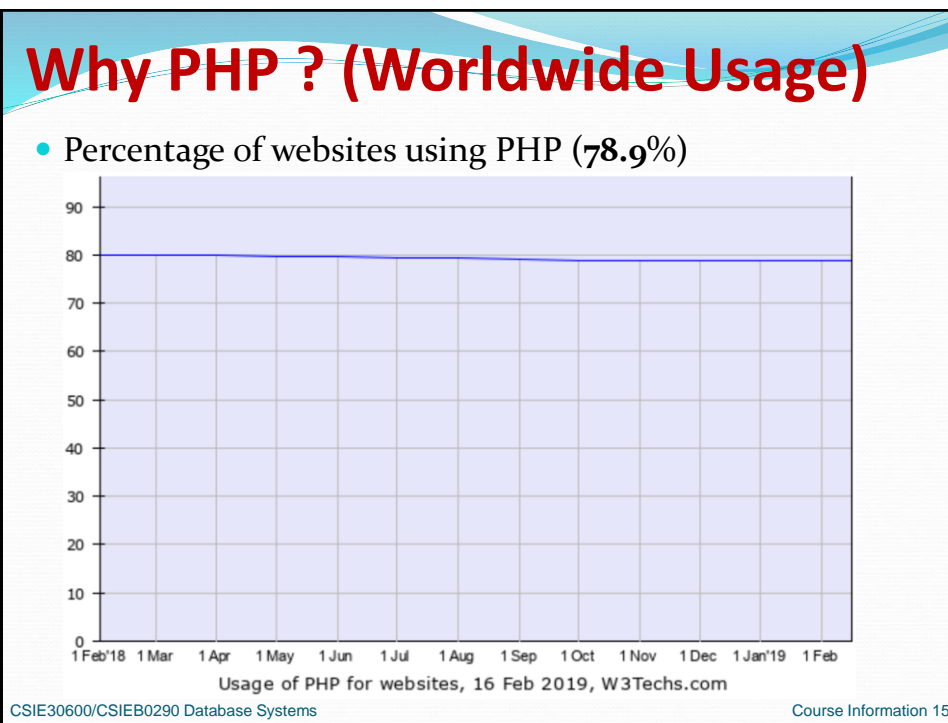- Demonstration and report due date: June 28, 2019.

Why MySQL & PostgreSQL?
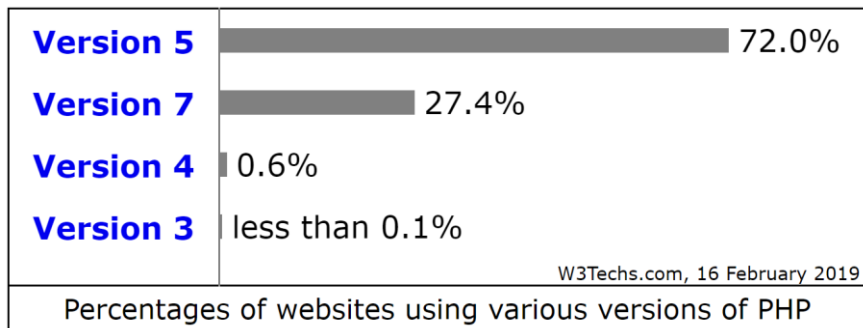
343 systems in ranking, February 2019

Rank			DBMS	Database Model	Score		
Feb 2019	Jan 2019	Feb 2018			Feb 2019	Jan 2019	Feb 2018
1.	1.	1.	Oracle	Relational, Multi-model	1264.02	-4.82	-39.26
2.	2.	2.	MySQL	Relational, Multi-model	1167.29	+13.02	-85.18
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1040.05	-0.21	-81.98
4.	4.	4.	PostgreSQL	Relational, Multi-model	473.56	+7.45	+85.18
5.	5.	5.	MongoDB	Document	395.09	+7.91	+58.67
6.	6.	6.	IBM Db2	Relational, Multi-model	179.42	-0.43	-10.55
7.	7.	↑8.	Redis	Key-value, Multi-model	149.45	+0.43	+22.43
8.	8.	↑9.	Elasticsearch	Search engine, Multi-model	145.25	+1.81	+19.93
9.	9.	↓7.	Microsoft Access	Relational	144.02	+2.41	+13.95
10.	10.	↑11.	SQLite	Relational	126.17	-0.63	+8.89
11.	11.	↓10.	Cassandra	Wide column	123.37	+0.39	+0.59
12.	↑13.	↑17.	MariaDB	Relational, Multi-model	83.42	+4.60	+21.77
13.	↓12.	13.	Splunk	Search engine	82.81	+1.39	+15.55
14.	14.	↓12.	Teradata	Relational	75.97	-0.22	+2.98
15.	15.	↑18.	Hive	Relational	72.29	+2.38	+17.23
16.	16.	↓14.	Solr	Search engine	60.96	-0.52	-2.91
17.	17.	↓16.	HBase	Wide column	60.28	-0.12	-1.43
18.	18.	↑19.	FileMaker	Relational	57.79	+0.64	+3.43
19.	19.	↑20.	SAP HANA	Relational, Multi-model	56.55	-0.09	+9.19
20.	↑21.	↓15.	SAP Adaptive Server	Relational	55.75	+0.71	-7.74

Why MySQL & PostgreSQL?

Why PHP ? (Worldwide Usage)

- Percentage of websites using PHP (**78.9**%)

Usage of PHP for websites, 16 Feb 2019, W3Techs.com

Why PHP ? (Market Position)

PHP Market Position, 16 Feb 2019, W3Techs.com

Note 8

Which Version of PHP?

Version 5	████████████████ 72.0%
Version 7	████ 27.4%
Version 4	▌0.6%
Version 3	less than 0.1%

W3Techs.com, 16 February 2019

Percentages of websites using various versions of PHP

Why Python ?

Feb 2019	Feb 2018	Change	Programming Language	Ratings	Change
1	1		Java	15.876%	+0.89%
2	2		C	12.424%	+0.57%
3	4	^	Python	7.574%	+2.41%
4	3	v	C++	7.444%	+1.72%
5	6	^	Visual Basic .NET	7.095%	+3.02%
6	8	^	JavaScript	2.848%	-0.32%
7	5	v	C#	2.846%	-1.61%
8	7	v	PHP	2.271%	-1.15%
9	11	^	SQL	1.900%	-0.46%
10	20	^^	Objective-C	1.447%	+0.32%
11	15	^^	Assembly language	1.377%	-0.46%
12	19	^^	MATLAB	1.196%	-0.03%
13	17	^^	Perl	1.102%	-0.66%
14	9	vv	Delphi/Object Pascal	1.066%	-1.52%
15	13	v	R	1.043%	-1.04%
16	10	vv	Ruby	1.037%	-1.50%
17	12	vv	Visual Basic	0.991%	-1.19%
18	18		Go	0.960%	-0.46%
19	49	^^	Groovy	0.936%	+0.75%
20	16	vv	Swift	0.918%	-0.88%

Note 9

Why Study Databases?

- Databases used to be *specialized applications,* now they are a *central component* in computing environments
  - Knowledge of database concepts is essential for computer scientists
- Databases are everywhere, even when you don't see them
  - 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 DB!
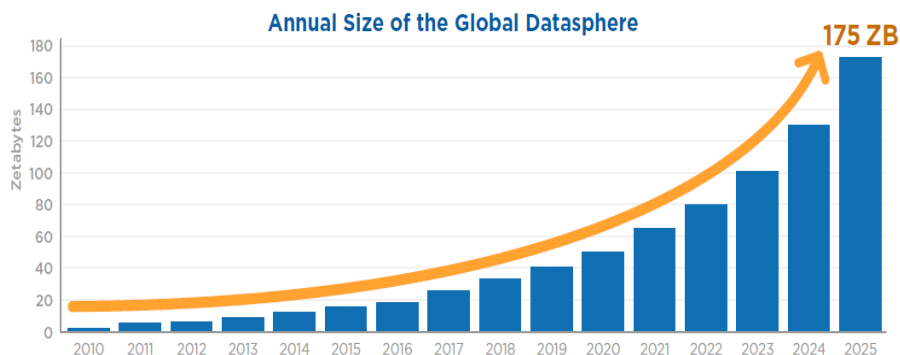
- Data needs to be *managed*

Note 10

Global DataSphere

- DataSphere: the sum of ALL data around the world!
- Global DataSphere will growh to **175 ZB** by 2025, according to IDC.

Figure 1 – Annual Size of the Global Datasphere

Annual Size of the Global Datasphere

175 ZB

CSIE30600/CSIEB0290 Database Systems　　　　　　　　Course Information 21

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**

CSIE30600/CSIEB0290 Database Systems　　　　　　　　Course Information 22

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 useful 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

CSIE30600/CSIEB0290 Database Systems   Course Information 23

Why Study Databases?

- Because the database field has made significant contributions to basic computer science:
  - *Understand concepts and apply to different problems and different areas*
- Because DBMS software is highly successful as a commercial technology (Oracle, DB2, SQL Server...)
- Because DB research is highly active and VERY interesting!
  - Lots of opportunities to have practical impact

CSIE30600/CSIEB0290 Database Systems   Course Information 24

Syllabus

- Introduction
- Databases and database users
- Database system concepts and architecture
- Basic SQL (how to use a DB)
- ER/EER models and conceptual design
- Relational model and constraints
- Relational algebra and calculus**
- 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
- Storage structure and indexing methods**
- Query processing and optimization
- Physical design and tuning**
- Transaction processing
- Concurrency control
- Recovery**

Syllabus (cont.)

- Object and object-relational databases**
- Semi-structured data and XML**
- Web database
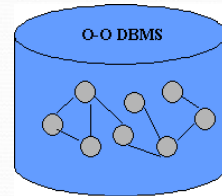
Figure 1: O-O Database Structure

Syllabus (cont.)*

- Distributed and parallel databases
- Cloud computing and big data trends
- Big data systems and NoSQL databases
- Big data processing and analytics

© Can Stock Photo

Syllabus (Advanced Topics)*

- Database security
- Active databases
- Temporal and real-time databases
- Spatial databases
- Multimedia databases
- Deductive databases
- Information retrieval and Web search
- Data mining and data warehousing
- Mobile and pervasive data management
- Streaming data management/analytics
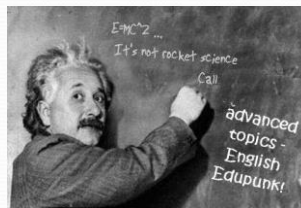
Accept the DB Challenges

- A very interesting and challenging class
- Be prepared for some theoretical discussion on the principles and algorithms.
- You must keep up with the pace.

- Ask questions if you miss the point.
- Design your own examples.

- Join us with the DB challenges !!!

Note 15