


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)**.
- Study the **internals** of DBMS
- Learning state-of-the-art **open source DBMS**
- **Advanced topics 1:** Cloud DBs, Big Data, NoSQL, NewSQL, Distributed SQL, Streaming Data*
- **Advanced topics 2:** XML DB, OODB, mobile DB, multimedia DB, parallel/distributed DB ...

CSIE30600/CSIEB0290 Database Systems Course Information 2

Course Information



- Course Title: Database Systems
- Course Number: CSIE30600/CSIEB0290
- Meeting Time: Thu 14:10~17:00
- Classroom: Science/Engineering Building II A331
- 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/DB2022f/index.html>
- Not on “e學苑 - e-Learning@NDHU” !!
- Instructor's homepage:
<http://web.csie.ndhu.edu.tw/showyang>

Online Class (if necessary)



- Teams link:
https://teams.microsoft.com/l/meetup-join/19%3az6QLI1m3Hd8BHpPW0mDJpJjS_-KLszoUIMPJZdrivc1%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



- 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: Perform fast and efficient data analysis with the power of SQL**. Packt Publishing, 2019.
- 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.

SQL References (cont.)



- Alan Beaulieu. **Learning SQL: Generate, Manipulate, and Retrieve Data, 3rd Edition**. O'Reilly Media, 2020.
- 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



- Jon Ducket. *PHP & MySQL: Server-side Web Development*. Wiley, 2022.
- Robin Nixon. *Learning PHP, MySQL & JavaScript, 6th Edition*. O'Reilly Media, Inc., 2021.
- 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



- Luciano Ramalho. *Fluent Python: Clear, Concise, and Effective Programming, 2nd Edition*. O'Reilly Media, 2022.
- Eric Matthes. *Python Crash Course, 2nd Edition: A Hands-On, Project-Based Introduction to Programming*, No Starch Press, 2019.
- Steve Holden, Anna Ravenscroft and Alex Martelli. *Python in a Nutshell, 3rd Edition*. O'Reilly Media, Inc. 2017.
- Wes McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython, 2nd Edition*. O'Reilly Media, 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.

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



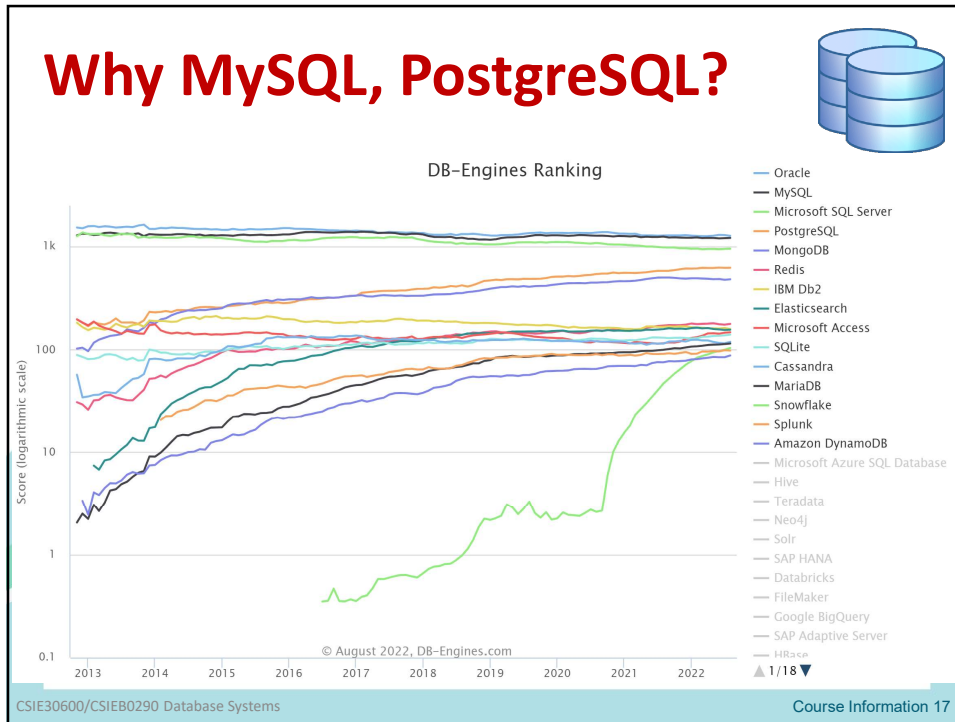
- An on-line database application (details will be announced in class)
- Use an open source DB as backend database
- Use browser or smart phone as user interface
- Can use any technique to connect the database.
- We will discuss PHP+MySQL and/or Python+MySQL/PostgreSQL.
- Demonstration and report due date: Dec 29, 2022.

Why MySQL, PostgreSQL?



395 systems in ranking, August 2022

Rank			DBMS	Database Model	Score		
Aug 2022	Jul 2022	Aug 2021			Aug 2022	Jul 2022	Aug 2021
1.	1.	1.	Oracle +	Relational, Multi-model i	1260.80	-19.50	-8.46
2.	2.	2.	MySQL +	Relational, Multi-model i	1202.85	+7.98	-35.37
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model i	944.96	+2.83	-28.39
4.	4.	4.	PostgreSQL +	Relational, Multi-model i	618.00	+2.13	+40.95
5.	5.	5.	MongoDB +	Document, Multi-model i	477.66	+4.68	-18.88
6.	6.	6.	Redis +	Key-value, Multi-model i	176.39	+2.77	+6.51
7.	7.	7.	IBM Db2	Relational, Multi-model i	157.23	-3.99	-8.24
8.	8.	8.	Elasticsearch	Search engine, Multi-model i	155.08	+0.75	-2.01
9.	9.	↑10.	Microsoft Access	Relational	146.50	+1.41	+31.66
10.	10.	↓9.	SQLite +	Relational	138.87	+2.20	+9.06
11.	11.	11.	Cassandra +	Wide column	118.15	+3.74	+4.49
12.	12.	12.	MariaDB +	Relational, Multi-model i	113.89	+1.37	+14.92
13.	13.	↑23.	Snowflake +	Relational	103.12	+3.97	+56.58
14.	14.	↓13.	Splunk	Search engine	97.44	-0.76	+6.84
15.	↑16.	↑16.	Amazon DynamoDB +	Multi-model i	87.26	+3.32	+12.36
16.	↓15.	↓15.	Microsoft Azure SQL Database	Relational, Multi-model i	86.18	+1.28	+11.02
17.	17.	↓14.	Hive +	Relational	78.66	-0.82	-5.27
18.	18.	↓17.	Teradata +	Relational, Multi-model i	69.07	-1.85	+0.25
19.	19.	↓18.	Neo4j +	Graph	59.35	+0.94	+2.40
20.	20.	20.	Solr	Search engine, Multi-model i	55.78	+0.08	+4.71

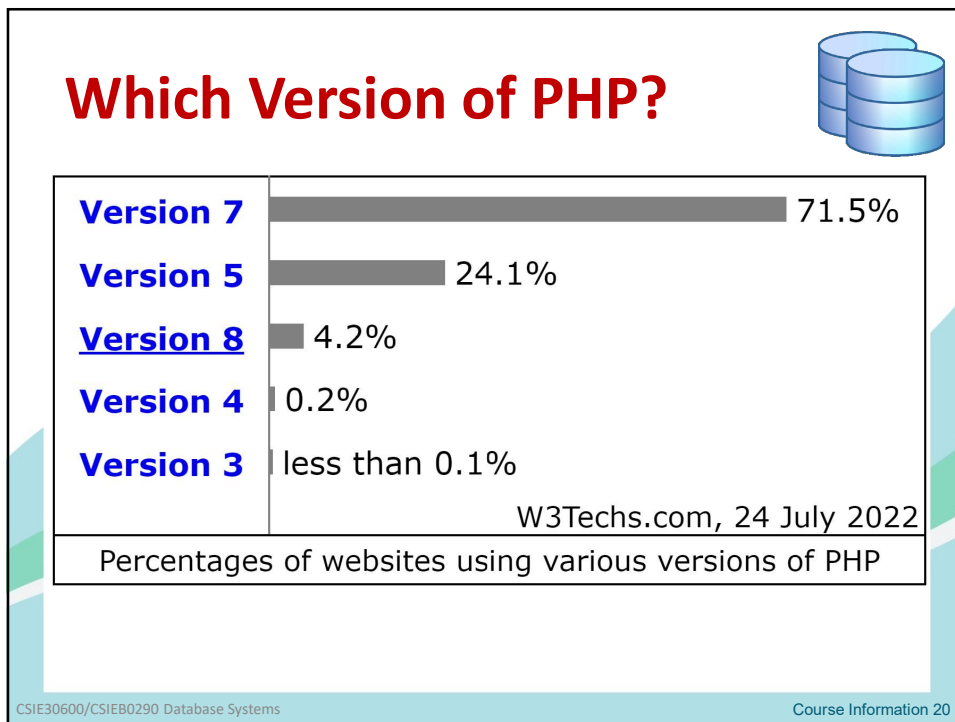
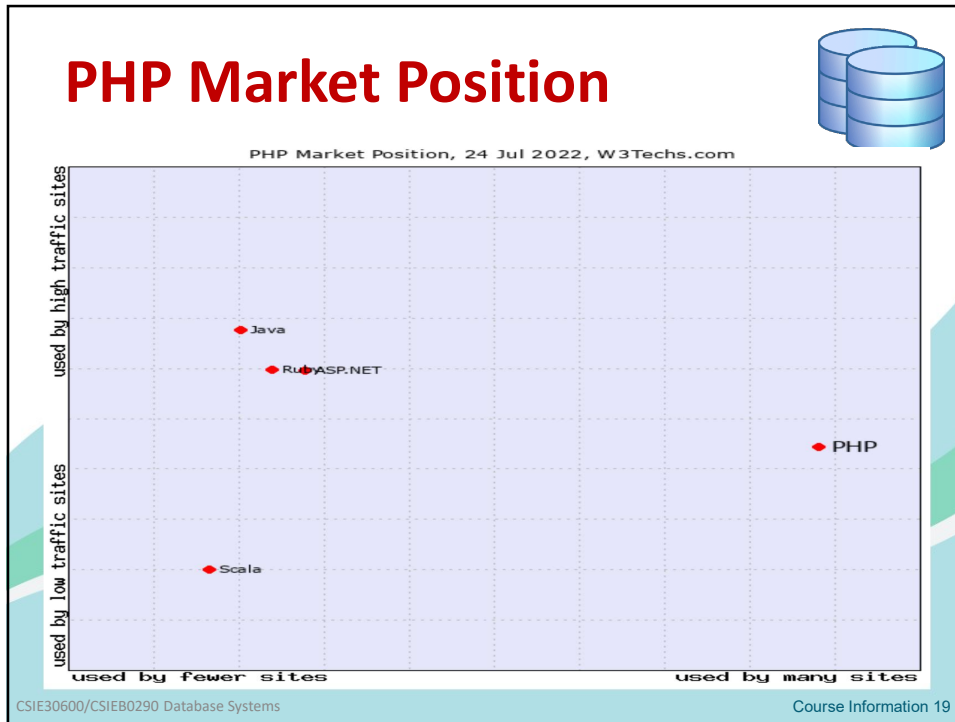


Why PHP ?

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

© W3Techs.com	usage	change since 1 July 2022
1. PHP	77.4%	-0.2%
2. ASP.NET	7.6%	-0.2%
3. Ruby	5.7%	-0.1%
4. Java	4.4%	+0.3%
5. Scala	2.7%	

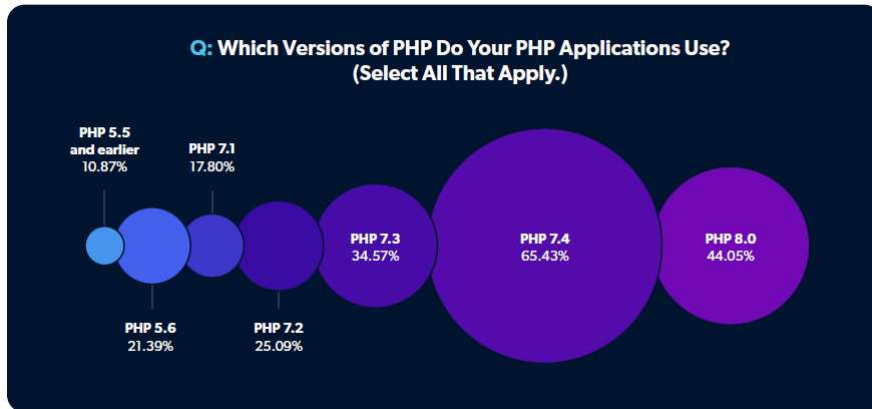
percentages of sites



Versions of PHP ?



- Another survey



(<https://www.zend.com/blog/state-php-2022>)

Why Python ? (TIOBE Index)



Jul 2022	Jul 2021	Change	Programming Language	Ratings	Change
1	3	↑	Python	13.44%	+2.48%
2	1	↓	C	13.13%	+1.50%
3	2	↓	Java	11.59%	+0.40%
4	4		C++	10.00%	+1.98%
5	5		C#	5.65%	+0.92%
6	6		Visual Basic	4.97%	+0.47%
7	7		JavaScript	1.78%	-0.93%
8	9	↑	Assembly language	1.65%	-0.76%
9	10	↑	SQL	1.64%	+0.31%
10	16	↑	Swift	1.27%	+0.20%
11	8	↓	PHP	1.20%	-1.38%
12	13	↑	Go	1.14%	-0.03%
13	11	↓	Classic Visual Basic	1.07%	-0.32%
14	20	↑	Delphi/Object Pascal	1.06%	+0.21%
15	17	↑	Ruby	0.99%	+0.04%
16	21	↑	Objective-C	0.94%	+0.17%
17	18	↑	Perl	0.78%	-0.32%
18	14	↓	Fortran	0.76%	-0.36%
19	12	↓	R	0.76%	-0.57%
20	19	↓	MATLAB	0.73%	-0.15%

Why Python ? (PYPL Index)



Worldwide, Jul 2022 compared to a year ago:

Rank	Change	Language	Share	Trend
1		Python	28.38 %	-2.3 %
2		Java	17.5 %	-0.7 %
3		JavaScript	9.29 %	+0.1 %
4		C#	7.63 %	+0.5 %
5		C/C++	6.48 %	-0.1 %
6		PHP	5.32 %	-1.0 %
7		R	4.13 %	+0.4 %
8	↑↑	TypeScript	2.55 %	+0.8 %
9	↓	Objective-C	2.13 %	+0.3 %
10	↑	Swift	1.95 %	+0.3 %

What about JavaScript ?




Most popular client-side programming languages

© W3Techs.com	usage	change since 1 July 2022
1. JavaScript	98.0%	
2. Flash	1.4%	


percentages of sites

PHP vs JavaScript (1)



PHP	JavaScript
<ul style="list-style-type: none"> • Server-side scripting language • Used for back-end development 	<ul style="list-style-type: none"> • Client-side scripting language • Mainly used for front-end development
<ul style="list-style-type: none"> • More secure (as is not visible in browser) • Helps to build high-level interactive web pages • Quite slow performance • More features available • Combined with HTML • MariaDB, MySQL, and PostgreSQL; • WordPress, Drupal, Joomla • Best for e-commerce and other websites using CMS 	<ul style="list-style-type: none"> • Has tools for enhancing security but needs more effort to do so • Helps to build user-friendly creative web pages • Fast performance • Less load on a server and less server traffic • Combined with HTML, XML, Ajax • AngularJS and ReactJS: • MongoDB, CouchDB, and NoSQL • Best for dynamic SPAs

PHP vs JavaScript (2)




PHP VS JAVASCRIPT: KEY DIFFERENCES

Features	PHP	PHP	Javascript	JS
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	


>hackr.io

PHP vs JavaScript (3)



**PHP VS JAVASCRIPT
EMPLOYMENT OPPORTUNITIES**


PHP	JS
PHP	JavaScript:
Annual Salary: \$91,215	Annual Salary: \$96,533
# of Open Jobs on Indeed: 104,000+	# of Open Jobs on Indeed: 12,000+



>hackr.io

<https://hackr.io/blog/php-vs-javascript>

JavaScript for Server-side?



Fastest growing server-side programming languages since 1 July 2022

Rank	Language	sites
1.	Java	602
2.	Python	159
3.	JavaScript	105

daily number of additional sites
in the top 10 million

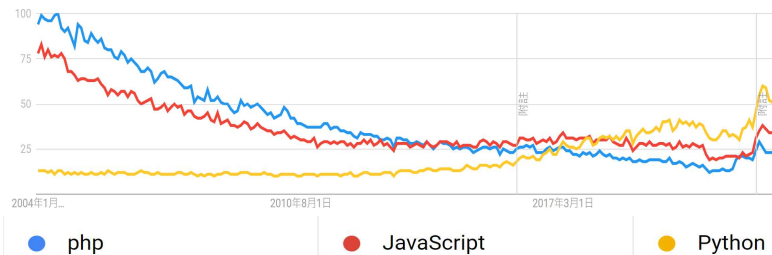
Need environment/frameworks such as Node.js, Express.js, ... (out of the scope of this course)

CSIE30600/CSIEB0290 Database Systems Course Information 28

Is PHP dying ?



- Google Trend



- Still used by > 77% 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 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?



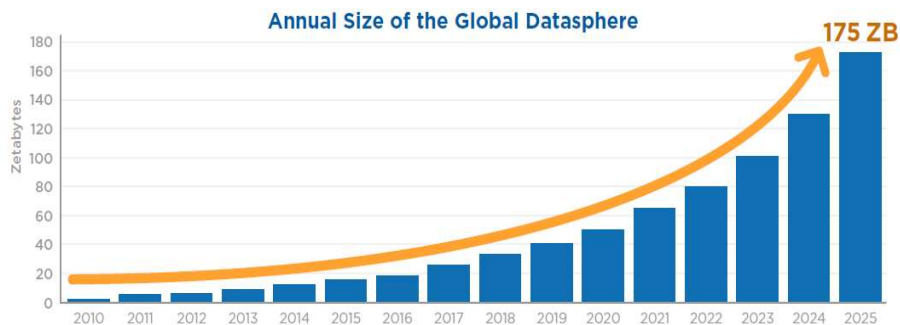
- Sales: customers, products, purchases
- Manufacturing: production, inventory, orders, supply chain
- Human resources: employee records, salaries, tax deductions
- Web sites: generated from databases; front-ends to databases
- Scientific research, e.g., studying the environment
- Your own data!
- Global data volume grows faster than ever! (next slide)
- Sky-high demand for Big data and NoSQL/NewSQL DB!
- Streaming data from IoT devices
- **Data needs to be *managed***

Global DataSphere



- **DataSphere**: the sum of ALL data around the world!

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 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

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, SQL Server, DB2, ...)
- 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



© Can Stock Photo

CSIE30600/CSIEB0290 Database Systems

Course Information 37

Syllabus (cont.)



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



CSIE30600/CSIEB0290 Database Systems

Course Information 38

Syllabus (cont.)

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


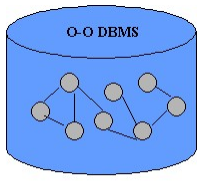







Figure 1: O-O Database Structure



CSIE30600/CSIEB0290 Database Systems Course Information 39

Syllabus (cont.)**

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

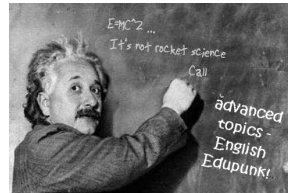




CSIE30600/CSIEB0290 Database Systems Course Information 40

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**.
- You must **keep up** with the pace.
- **Ask questions** if you miss the point.
- Design your own **examples**.
- Join us on the **DB challenges** !!!

