

Spring 2017

ICS421 Database Systems

Dr. Lipyeow Lim

Information & Computer Sciences

University of Hawai`i at Mānoa

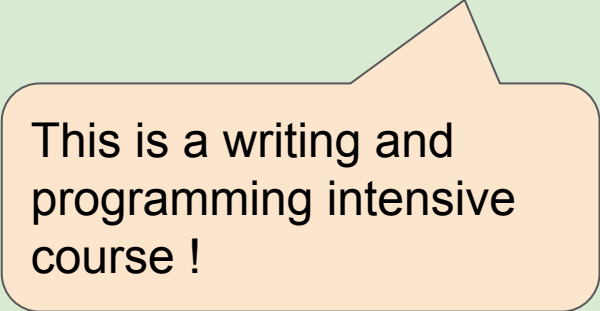
What you will learn in this course

How to manage very large data

- Parallel & distributed Database Technology
- Key concepts in NoSQL technology
- Architectural patterns for building big data systems

Pre-requisites

- Basic DBMS concepts from ICS321
- Software engineering from ICS314
- Multi-threaded programming



This is a writing and programming intensive course !

Rhythm of the Course

- Read the assigned readings BEFORE class
- Take the quiz BEFORE class
- Attend and participate DURING class
 - Summary lecture
 - In-class exercises
- Start early and work on the homework assignments
 - Writing component
 - Programming component
- Start early and work on the project

Homework Assignments

There will be about 3 homework assignments that are designed to be a single system building exercise. You will be building a *Parallel SQL Processing System* that will

- Create tables stored on a cluster of machines
- Load data into those tables
- Run SQL queries on those tables
- Support parallel join algorithms on those tables

You will be writing extensive design and user documentation to fulfil the writing intensive requirement

Computing & Software Environment

- Google credits will be provided for testing of programming assignments & project
 - Cluster of VMs
- Your choice of programming language, DBMS & API
 - Java+JDBC recommended, Python is ok.
- ANTLR will be used to generate the parser for SQL

Development should be done on your laptop

You are expected to learn these platforms, software, APIs, tools, etc. on your own!

Course Project

There will be some choice, but must involve

- A large real data set
- Non-trivial processing (SQL or other analytics) on that data set
- A comparison between:
 - The Parallel SQL Processing System you have built
 - At least one other open source or commercial big data technology



EVERY DAY WE CREATE

2,500,000,
000,000,
000,000

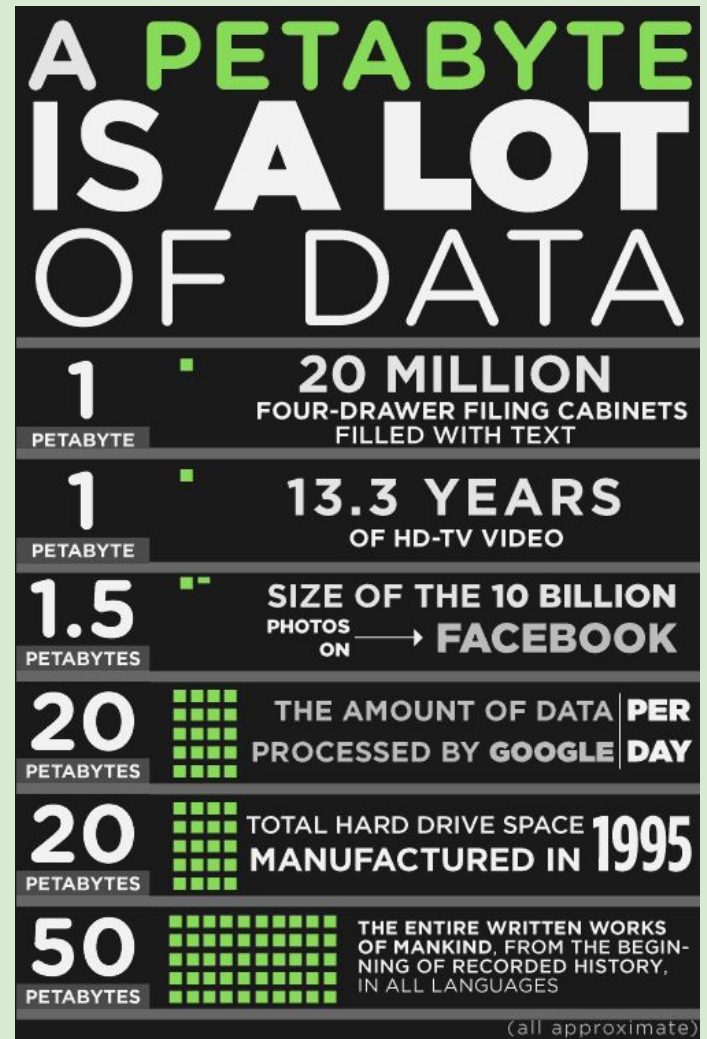
(2.5 QUINTILLION) BYTES OF DATA

*This would fill 10 million blu-ray discs,
the height of which stacked, would measure
the height of 4 Eiffel Towers on top of one another.*



Multiples of Bytes

Multiples of bytes						V·T·E
Decimal			Binary			
Value	Metric		Value	IEC	JEDEC	
1000	kB	kilobyte	1024	KiB	kibibyte	KB kilobyte
1000 ²	MB	megabyte	1024 ²	MiB	mebibyte	MB megabyte
1000 ³	GB	gigabyte	1024 ³	GiB	gibibyte	GB gigabyte
1000 ⁴	TB	terabyte	1024 ⁴	TiB	tebibyte	—
1000 ⁵	PB	petabyte	1024 ⁵	PiB	pebibyte	—
1000 ⁶	EB	exabyte	1024 ⁶	EiB	exbibyte	—
1000 ⁷	ZB	zettabyte	1024 ⁷	ZiB	zebibyte	—
1000 ⁸	YB	yottabyte	1024 ⁸	YiB	yobibyte	—



WHERE IS DATA COMING FROM?

