welcome to CS165!
DATA SYSTEMS
prof. Stratos Idreos
big data?
Every two days we create as much data as much we did from dawn of humanity to 2003 [Eric Schmidt, Google]
data systems are in the middle of all this
today

what is a db?

WANTED
Data Scientists

tomorrow
database system?
“relational databases are the foundation of western civilization”

Bruce Lindsay, IBM
ACM SIGMOD  Edgar F. Codd Inovations award 2012
dbs are everywhere...
how would 1 day without databases look like
declarative interface
ask “what” you want

db system

the system decides “how” to best store and access data

why is this good

5 decades of research
IBM, Microsoft, Oracle, Teradata, etc.
and a gazillion start-ups today
SQL queries

db system

>1 users concurrently

correct + complete answers

security/robustness
“Three things are important in the database world: performance, performance, and performance”

Bruce Lindsay, IBM
ACM SIGMOD  Edgar F. Codd Innovations award 2012
essential steps in using a database system

clean \[\rightarrow\] schema \[\rightarrow\] load \[\rightarrow\] tune

user/apps

query

experts/system admins
data systems architectures

data structures + algorithms

some problems:
how to store data
how to access data

how to best answer a complex query
(e.g., which data to access first and how)

how to answer millions of queries concurrently

how to guarantee correctness and availability
you will learn to design and implement db kernels!
you will learn to design and implement db kernels!

res = 0;
for (i = 0; i < 100; i++)
    if (d[i] > 10) res += d[i];

what this does is it “good”
data systems design (and research) is kind of an art
haven’t we be doing data analysis forever
so what is new
big data V’s
(it is not about size only)

volume  velocity  variety  veracity
actually all of CS is in the middle of the big data story!

major opportunity (and **responsibility**) for interdisciplinary research
"All models are wrong, but some are useful."

there are good chances we already have the data for the next big breakthroughs in say biology, medicine, etc. but we simply cannot extract the knowledge
tons of opportunities!

research - entrepreneurship - new fields
~1960s

late 1990s-early 2000: new designs start appearing

~2010-now: industry adoption and evolution

~2014

history/timeline

"dbs"
cs165 topics

modern systems
e.g., column-store and hybrid systems, shared nothing architectures, cache-conscious algorithms, hardware/software co-design, main memory systems, adaptive indexing, stream processing, scientific data management, and key value stores

past but still relevant topics
e.g., relational model, row-store database systems, optimization, indexing, concurrency control, recovery, SQL

how and why did we get here and where things might go
cs165 goals
understanding system design tradeoffs
be able to design and prototype a data system!

why
data system designer - researcher
any business - any science - any start-up

data scientists
logistics

Lectures twice a week + sections on demand

Quizzes + class participation 15%

Midterms (2) 20% (March 3 - April 14)

Running project 40% (design/build components of a data system)

Final project 25% (free topic, group project, research possible)

Extra points 10% on running project

Best projects 5% on running project (best 5)

Best projects 5% on final project (best 3)

10 late days for running project
more logistics

web site: http://stratos.seas.harvard.edu/classes/cs165-data-systems
(always check for updates+announcements)

notes: slides online + collaborative class notes

office hours: Stratos, Wed 2:30pm-4:00pm

TFs office hours: TBA

textbook: Database Management Systems,
by R. Ramakrishnan and J. Gehrke
+ several modern data systems surveys and research papers
prof. Stratos Idreos

other names: Efstratios Ydraios
Ευστράτιος Υδραίος, Στράτος Υδραίος

Grew up in Greece - fav non-cs hobby: windsurfing

Diploma and ME  Technical University of Crete, Greece
Ph.D.  University of Amsterdam, Netherlands
Research Intern: IBM Research California, Microsoft Research Redmond, EPFL Switzerland
Visiting Professor: National University of Singapore, EPFL Switzerland

Fav Awards:
ACM SIGMOD Jim Gray Dissertation Award
ERCIM Cor Baayen Award
**a db system**
allows you to answer queries fast

**a data exploration db system**
allows you to find fast which queries to ask
design db kernels for touch-based exploration

S. Idreos, E. Liarou.  
*dbTouch: Analytics at your fingertips.*  
Conference on Innovative Data Systems Research (CIDR), 2013
feedback

we want you to have fun!
data systems is an exciting field!

tell us how you are keeping up
tell us what you need to better follow the class
next class:
history - essential db properties
model - SQL - future

as of next week:
system architecture and design!
welcome to CS165!

DATA SYSTEMS

prof. Stratos Idreos