Final Notes

Alvin Cheung
Aditya Parameswaran

The Class Mission (Day 1 Slide 2)

This class ~~will cover~~ *covered* how to develop systems to *efficiently manage, maintain, process, query, transact with, and make sense of data*

You ~~will learn~~ *have learned*...

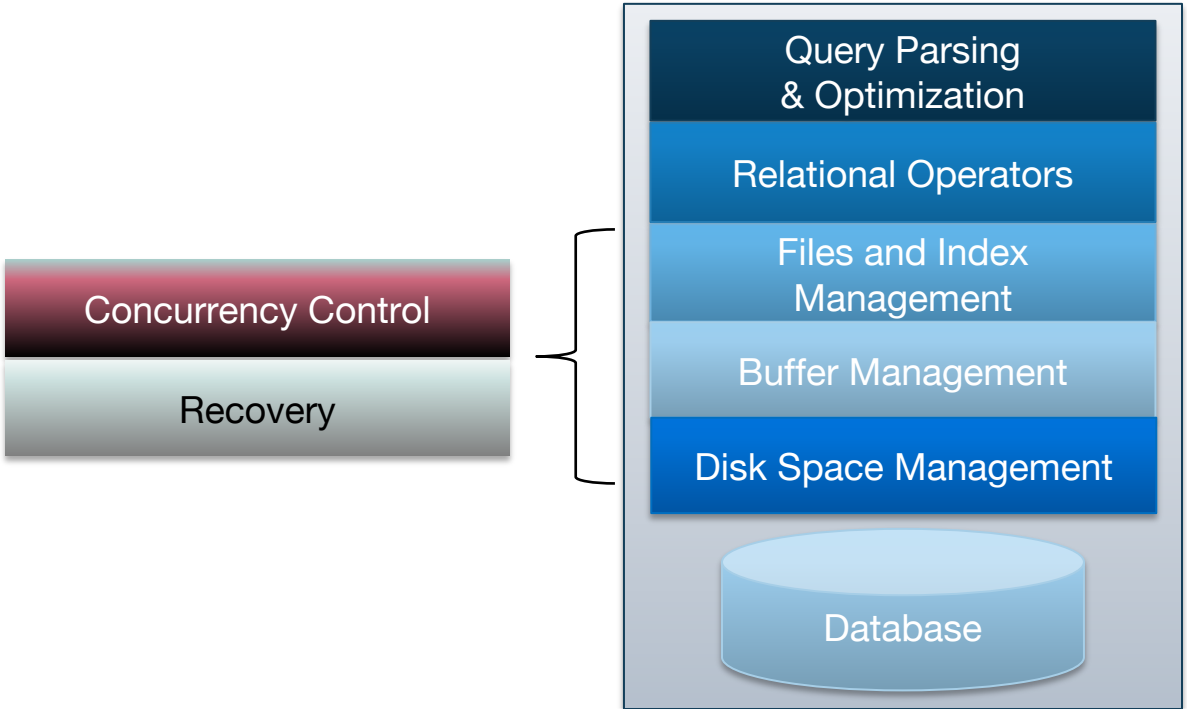
- Data Oriented Programming with SQL
- Foundations of Data System Design
  - Storage, indexing
  - Query processing and optimization
- Transactions
  - Concurrency, Consistency, Recovery
- Data Modeling
  - Application-level representations of data

Principles

- Data Independence
- Declarative Programming
- Rendezvous in Time and Space
- Isolation and consistency
- Data representations

Systems

We ~~will examine~~ *examined* various levels of a DBMS



And we extended the stack!

- DB Design
- Parallel query processing
- Distributed transactions
- NoSQL & MongoDB
- MapReduce & spark
- OLAP and column stores

You DID It! Well, almost …

- Mastered SQL & Worked with Sqlite
- Implemented B+Trees
- Implemented Join Algorithms and Selinger Qopt
- Implemented (Multi-granul.) Locking
- Implemented ARIES
- Mastered MQL & Worked with MongoDB

If you're curious to learn more….

- Classes, grad
  - CS286B: Graduate database systems
  - CS198-12: Data systems and foundations seminar
  - CS262: Advanced topics in computer systems
  - CS294-170: Programming the cloud
  - CS294-162: ML systems

- Classes, undergrad
  - Data100: Data science
  - CS162: OS/System programming
  - CS164: PL and compilers
  - CS188/189: AI/ML
  - CS161: Security

If you're curious to learn more….

- Consider doing research!
- The field of data management has continued to evolve since the 70s…
  - One of the most exciting areas of research today!
  - Strong overlaps with Systems, Arch, HCI, PL, ML, Security, Theory, ….
  - Strong overlaps with industry & open-source community
  - Vibrant conferences: VLDB & SIGMOD
  - Thriving research community

If you're curious to learn more….

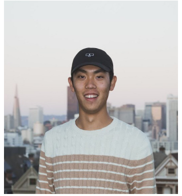
- Consider doing research *with us*!
- We're part of the "Data Systems & Foundations Group"
  - w/ Natacha Crooks, Joey Gonzalez & Joe Hellerstein
  - And many affiliates
  - https://dsf.berkeley.edu/
  - Reach out to us if you are interested in research projects!

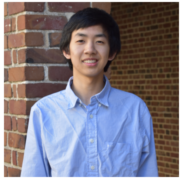
A Smattering of Research Projects

- Alvin C.
  - LightDB: A DBMS for video and visual data!
  - Chestnut: autogen. in-memory data layouts & query plans for speeding up applications
- Natacha C.
  - Obladi: txns in the cloud while hiding access patterns
  - Enforcing consistency guarantees on streaming data
- Joe H.
  - Flor: materialization of program state (views!) to speed up debugging of ML pipelines
  - Hydro: data programming for the serverless cloud
- Joey G.
  - Clipper: efficient serving of ML models at scale
  - RayTune: multi-query optimization and resource sharing for ML model search
- Aditya P.
  - DataSpread: spreadsheets-meets-DBMSs
  - Orpheus: a database that supports versioning

Thank you, and we hope you had fun!

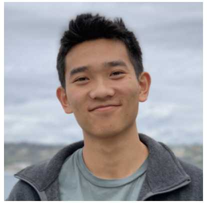




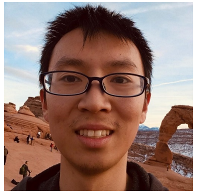













Berkeley
cs186