CS 6320 Intro

Immanuel Trummer
itrummer@cornell.edu

Course Organization

- **Lecture Times**

  - Tuesdays & Thursdays

  - 1:25 PM to 2:40 PM

  - Upson Hall 216

- **Office Hours**

  - Wednesday 3 PM to 4 PM

  - 411b Gates Hall

- **Web site** (online this evening):

  - http://www.cs.cornell.edu/courses/cs6320/2019sp/

Course Components

- Paper presentations & discussion (50% of Grade)

- Course project (50% of Grade)

Presentation

- Two (to three) papers on related topics

- Often mixing seminal with recent papers

- Duration: 1h15 for presentation **& discussion**

- Two students per presentation

- Need to send in slides at least **one day** in advance!

Presentation Hints

- One common story, not two separate papers

- Presentation should encourage discussions

  - Don't hesitate to throw questions at the audience!

- Make sure to leave enough time for discussions

- Time should be approximately split between papers

- Ideally: presentation teams of senior/junior students

Participation Hints

- Read the papers in advance!

- Don't hesitate to ask questions during the presentation

- Will **check attendance** starting from next week!

Course Project

- **<u>Up to three students</u>** can work on the same project

- Topic must relate to the **<u>broad database area</u>**

  - Can be a topic you're working on anyway

- Some high-level topic **<u>ideas</u>**

  - Deterministic approximation

  - Reinforcement learning for query optimization

  - Voice query interfaces

  - …

Project Timeline

- First **two weeks**: select a topic, write one page summary

- Until **March 15**: progress report (2 pages)

- Until **May 7**: final report (6 pages)

- Ideally your report turns into a research paper …

- Send all reports to itrummer@cornell.edu

Section 1: Foundations

- Indexing

- Join algorithms

- Query optimization

- Concurrency control

- Logging & recovery

- Buffer management

Section 2: Efficient Query Processing

- Main memory databases

- Query compilation

- Approximate processing

- Processing on novel hardware

- Massively parallel processing

Section 3: Transaction Processing

- CAP Theorem and NoSQL

- NewSQL systems

- Deterministic DBMS

- Coordination avoidance

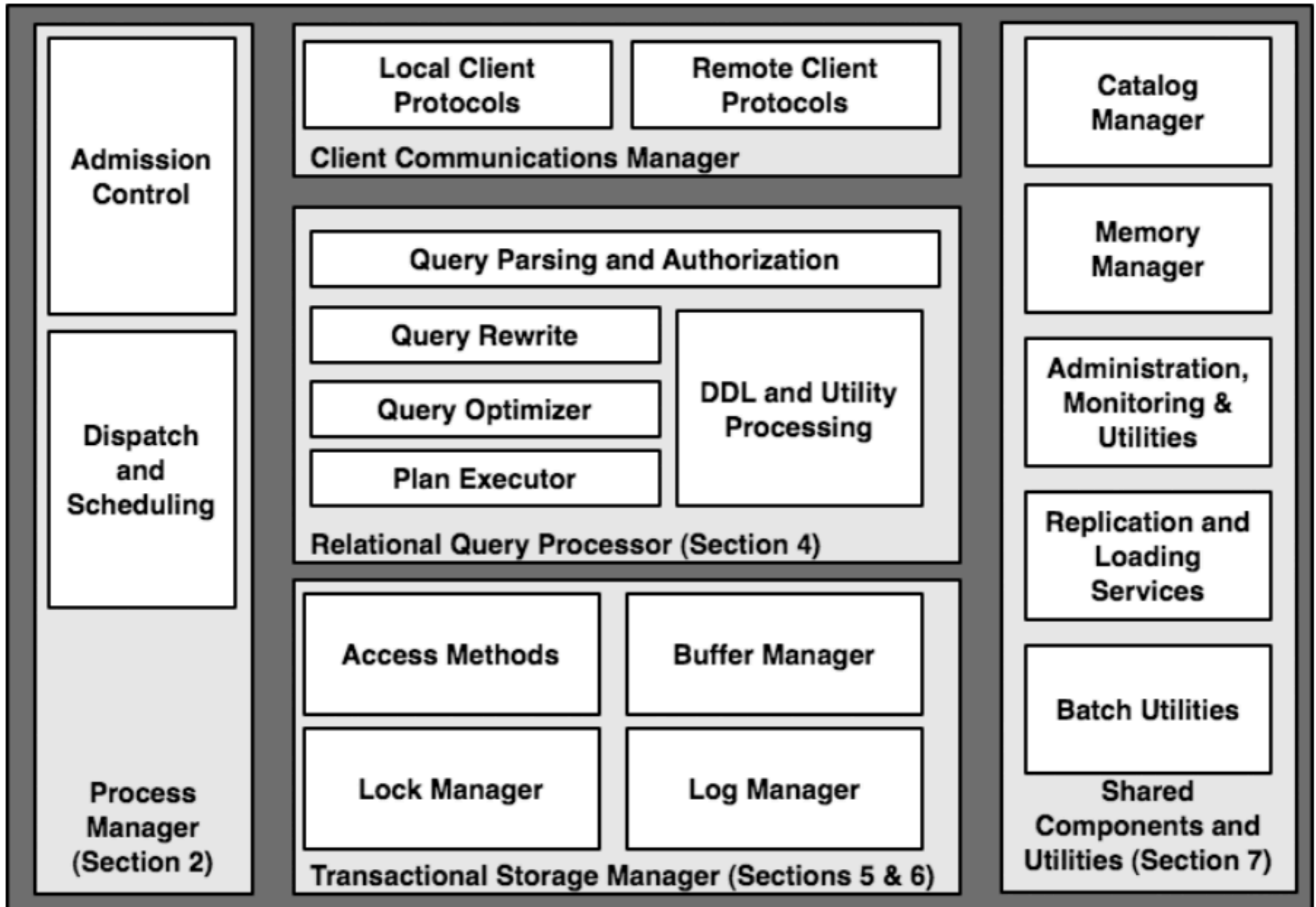- Concurrency control on multi-cores

Section 4: Beyond Relational Data Processing

- Graph databases

- Databases for time series

- Stream processing

- Spatial databases

- Systems for declarative ML

Section 5: Interfaces

- Data visualization

- Voice-based interfaces

- Query by example

- Gestural query interfaces and Augmented Reality

Next Lecture

Process Manager (Section 2)
- Admission Control
- Dispatch and Scheduling

Client Communications Manager
- Local Client Protocols
- Remote Client Protocols

Relational Query Processor (Section 4)
- Query Parsing and Authorization
- Query Rewrite
- Query Optimizer
- Plan Executor
- DDL and Utility Processing

Transactional Storage Manager (Sections 5 & 6)
- Access Methods
- Buffer Manager
- Lock Manager
- Log Manager

Shared Components and Utilities (Section 7)
- Catalog Manager
- Memory Manager
- Administration, Monitoring & Utilities
- Replication and Loading Services
- Batch Utilities

http://db.cs.berkeley.edu/papers/fntdb07-architecture.pdf

Questions?