Papers for Presentations

Below are papers for your presentations/critiques. You should select one from Chapters 1-8 and one or more related papers from the remaining. Each would correspond to one presentation. While the first chapters refer to established work, the second part includes some of the latest research and brief position/papers and, therefore, it would need a more thorough coverage (perhaps additional papers). The first section of presentations will start Feb 15 and will cover 2 presentations a day. Make a decision of the papers you would like to present early. Selection is first come first served. For the second part, we will have 3-4 presentations per day. As selections are made, these papers will be posted for critiques also.

The papers listed in black are from the text. Those in green are not from the text but will be made available on the class web site.

Chapter 1: Data Models and DBMS Architecture
- Stonebraker and Hellerstein, "What Goes Around Comes Around."
- Hellerstein and Stonebraker, "Anatomy of a Database System."

Chapter 2: Query Processing

Chapter 3: Data Storage and Access Methods

Chapter 4: Transaction Management


Chapter 5: Extensibility


Chapter 6: Database Evolution


Chapter 7: Data Warehousing


Chapter 8: Data Mining


Chapter 9: Web Services and Data Bases


Chapter 10: Stream-Based Data Management

2. Donald Carney, U. ur et al. (Cetinsoyel, Mitch Cherniack, Christian Convey, Sangdon Lee, Greg Seidman, Michael Stonebraker, Nesime Tatbul, Stanley Zdonik). Monitoring Streams-A New Class of Data Management Applications. VLDB


Sensor Networks
5. Amol Deshpande† Carlos Guestrin‡ Samuel Madden§ † Joseph M. Hellerstein ‡ Wei Hong‡, Model-Driven Data Acquisition in Sensor Networks, VLDB 04

Overlay Networks / Privacy
7. Marcin Gomulicz, Marek Klonowski, and Miroslaw Kutyłowski. Provably Unlinkable Against Traffic Analysis After O(log(n)) Steps!?
9. Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, Yihong Xu, "Hippocratic Databases" VLDB 2002

Project possible themes

The following are some ideas. You can also propose a topic that is related to your research/interests as long as it is related to databases.

1. P2P Databases
   - single global schema
   - peers maintain their own subset of the database
• peers query other peers for more tuples
• insertions are done anywhere tagged by the tuple creator
• deletions are piggy-bagged to queries and tagged by the peer doing it (semantics of deletions always difficult)
• inserted tuples are hashed and their hashed value is stored in a DHT per relation (perhaps the peer creator is incorporated in the hashed value)
• queries are issued to other peers using random or probabilistic walks
• real challenge is where to direct the query to receive "new" tuples and "recent" deletions
• real challenge: avoid the return of duplicates i.e. tuples that have been seen by the query peer
• use of adm s+ - download and upload protocols? is there a primary site for a tuple?

2. Web-Protection using a Proxy Protector
• all web requests go thru a single web site which acts like a proxy-protector (PP)
• the protocol is like DCLK.com which redirects the url to the PP which retrieves the real url and checks for viruses, dangerous attachments, etc.
• the PP redirection can be optional i.e. it can be applied by SHIFT-CLICK
• the PP will keep a database of clean and dirty urls with metadata on their creation duration, and any other description that can be detected.
• since there may be multiple PP sites, they can be have different databases that can be merged, synced, and updated.
• are there any advantages in realizing PP in a p2p architecture?

3. Probabilistic Privacy Protection (PPP)
• Goal: data monitoring while preserving privacy
• TOR and other Onion Protocols
• Privacy Protocols (P3P, P4P, Hippocratic Databases)
• P2P overlays with probabilistic data propagation
• Rendezvous protocols

4. Friendly Reviewing (continuation from last year)
• DB publications, coauthor, and program committee data are all cleaned up in an Oracle database
• Looking for correlations in the database: PC committee and coauthors

5. Mobile DB on Cell phones and PDAs
• Import tinyDB on PDAs and Cell phones
• Similar to sensor's tinyDB project: but with a bit more functionality in the schema