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"*.
- Nick Roussopoulos, Chung-Min Chen, Stephen Kelley, Alex Delis, Yannis Papakonstantinou: The ADMS Project: View R Us. IEEE Data Eng. Bull. 18(2): 19-28 (1995)

Chapter 2: Query Processing

1. P.G. Selinger, M.M. Astrahan, D.D. Chamberlin, R.A. Lorie, and T.G. Price. Access path selection in a relational database management system. In SIGMOD, 1979.
2. Leonard Shapiro. "Join Processing in Database Systems with Large Main Memories" ACM TODS Sept. 1986.
3. David J. DeWitt, Jim Gray. Parallel Database Systems: The Future of High Performance Database Processing Communications of the ACM.
4. Goetz Graefe. Encapsulation of Parallelism in the Volcano Query Processing System. SIGMOD Conference 1990: 102-111.
5. Lothar F. Mackert, Guy M. Lohman. R* Optimizer Validation and Performance Evaluation for Distributed Queries. VLDB 1986
6. Michael Stonebraker, Paul M. Aoki, Witold Litwin, Avi Pfeffer, Adam Sah, Jeff Sidell, Carl Staelin, Andrew Yu. Mariposa: A Wide-Area Distributed Database System. VLDB J. 5(1): 48-63 (1996).

Chapter 3: Data Storage and Access Methods

1. Norbert Beckmann, Hans-Peter Kriegel, Ralf Schneider, and Bernhard Seeger. "The R*-tree: An efficient and robust access method for points and rectangles," Proceedings of ACM SIGMOD Int'l. Conf. on Management of Data, pp. 322-331, 1990.
2. T. Sellis, N. Roussopoulos, and C. Faloutsos, The R+-Tree: A Dynamic Index For Multi-Dimensional Objects, VLDB 1987.
   http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/vldb87_R+Trees.pdf
3. N. Roussopoulos and D. Leifker. Direct Spatial Search on Pictorial Databases Using Packed R-Trees. In Proc. ACM SIGMOD, pp. 17-31, 1985.
   http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/Packed_R-Trees_Sigmod85.pdf
4. Jim Gray, Goetz Graefe. "The Five-Minute Rule Ten Years Later, and Other Computer Storage Rules of Thumb." ACM SIGMOD Record 26(4): 63-68 (1997).
5. David A. Patterson, Garth A. Gibson, Randy H. Katz: A Case for Redundant Arrays of Inexpensive Disks (RAID). SIGMOD 1988: 109-116.

Chapter 4: Transaction Management

1. Jim Gray, Raymond A. I_orie, (iianfrancu R. Iutzolu, Irving 1.. Traiger. Granularity of Lucks and Degrees of Consistency in it Shared Data Base. IFIP Working Conference oil Modelling in Data Base Management Systems 1976: 365–3c)4.
2. Il.T. Kung in (] John ']'. Ruhinum. "An Optimistic Methods for C:ortcurrency Control," TODS 6/81 1,. 213–226.
3. C. Mohan, Don Fladerle, Bruce Lindsay, Hamid Pirahesh, and Peter Schwartz. "ARIES: A Transaction Recovery Method Supporting Fine-Granularity Locking and Partial Rollbacks Using Write-ahead Logging," ACM TODS 17(1):94–162, 1993.
4. C. Mohan, Bruce Lindsay, and R. Obermiarck. "Transaction Management in the R* Distributed Database Management Systems." ACM TODS 11(4):378–396, 1986.
5. Jim Gray, Pat Helland, Patrick O Neil, and Dennis Slursha. The Dangers of Replication and a Solution. SIGMOD 96.

Chapter 5: Extensibility

1. Michael Stonebraker. Inclusion Of New Types In Relational Data Base Systems. Proceedings of the International Conference on Data Engineering, 1986.
2. Joseph M. Hellerstein, Jeffrey F. Naughton, Avi Pfeffer. Generalized Search Trees for Database Systems. Proc. 21st hit. Conf. Very Large Data Bases, VLDB 1995.
3. John Grant, Witold Litwin, Nick Roussopoulos, Timos K. Sellis: Query Languages for Relational Multidatabases VLDB J. 2(2): 153–171 (1993)
4. Manuel Rodriguez-Martinez, Nick Roussopoulos: MOCHA: A Self-Extensible Database Middleware System for Distributed Data Sources. SIGMOD Conference 2000: 213–224

Chapter 6: Database Evolution

1. Chung-Min Chen, Nick Roussopoulos: Adaptive Database Buffer Allocation Using Query Feedback. VLDB 1993: 342–353
http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/AdaptiveBufferMNGT_Query_Feedback_vldb93.pdf
2. Chung-Min Chen, Nick Roussopoulos: Adaptive Selectivity Estimation Using Query Feedback. SIGMOD Conference 1994: 161–172
http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/AdaptiveBufferMNGT_Query_Feedback_vldb93.pdf
3. Chung-Min Chen, Nick Roussopoulos: The Implementation and Performance Evaluation of the ADMS Query Optimizer: Integrating Query Result Caching and Matching. EDBT 1994: 323–336
http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/ADMS_Query_Optimizer_edbt94.pdf
4. Surajit Chaudhuri and Vivek Narasayya. "AutoAdmin 'What-if' Index Analysis Utility." Proc 1998 SIGMOD.
5. Philip A. Bernstein, "Applying Model Management to Classical Meta Data Problems," Proc. CIDR 2003 pp. 2()9–220.
6. C. Mohan and Inderpal Narang. "Algorithms for Creating Indexes for Very Large Tables Without Quiescing Updates," Proc 1992 SIGMOD.

Chapter 7: Data Warehousing

1. Surajit Chaudhuri and Cirneshwar Dayal. An overview of data warehousing and OLAP technology. SIGMOD Record, 26(1):65–74, 1997.
2. Patrick O Neil and Dallan Quass. Improved Query Performance with Variant Indexes. Proc. of ACM SIGMOD Conf., Tucson, Arizona, 1997.
3. Jim Gray, Surajit Chaudhuri, Adam Bosworth, Andrew Layman, Don Reichart, Murali Venkatrao. DataCube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. Data Mining and Knowledge Discovery 1, 29–53, 1997.

4. Yannis Kotidis and Nick Roussopoulos. DynaMat: A Dynamic View Management System for Data Warehouses. Proceedings ACM SIGMOD International Conference on Management of Data, June 1-3, 1999.
5. Yannis Sismanis, Antonios Deligiannakis, Nick Roussopoulos, Y. Kotidis: "Dwarf: shrinking the PetaCube." SIGMOD Conference 2002: 464-475 http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/SIGMOD2002_DWARF-shrinking-the-petacube.pdf
6. Yannis Sismanis, Nick Roussopoulos: The Complexity of Fully Materialized Coalesced Cubes. VLDB 2004: 540-551 http://www.cs.umd.edu/class/spring2006/cmsc724/SECURE/PAPER_LIST/VLDB_2004_DWARF_complexity.pdf
7. Nick Roussopoulos: Materialized Views and Data Warehouses. KRDB 1997: 12.1-12.6
8. Peter J. Haas, Joseph M. Hellerstein: Ripple Joins for Online Aggregation. SIGMOD Conference 1999: 287-298
9. Joseph M. Ftellerstein, Ron Avnur, Vij.ryshankar Raman. Informix under CONTROL: Online Query Processing. Data A4inin,y and Knuw Icd,yc !)i,coowrv 2000.

Chapter 8: Data Mining

1. Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. SIGMOD Conference 1996: 103-114.
2. John Shafer, Rakesh Agrawal, and Manish Mehta. SPRINT: A Scalable Parallel Classifier for Data Mining. Proc. 22nd hit. Conf. Very Large Databases.
3. Rakesh Agrawal and Ramakrishnan Srikant. Fast Algorithms for Mining Association Rules. Proc. 20th Int. Conf. Very Large Data Bases, VLDB 1994.
4. Surajit Chaudhuri, Vivek Narasayya, and Sunita Sarawagi. Efficient Evaluation of Queries with Mining Predicates. Proc. of the 18th Int'l Conference on Data Engineering (ICDE), 2002.

Chapter 9: Web Services and Data Bases

1. Eric A. Brewer. Combining Systems and Databases: A Search Engine Retrospective.
2. Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale I Iypertextual Web Search Engine. Proc. 7th International World Wide Web Conference (WWW 7) Computer Networks, 30(1-7):107-117, 1998.
3. Sergej Sizov, Michael Biwer, Jens Graupmann, Stefan Siersdorfer, Martin Theobalcl, Gerhurd Wcikum, and Patrick 7,inuoer. "The BINGO! System for Informiation Portal Generation and Expert Web Search," CIDR 2003.
4. Serge Abiteboul. Querying Semi-Structured Data. In Proc. of the International Conference on Database Theory ICDT'97, pp. 1-17, 1997.
5. Roy Goldman and Jennifer Widom. DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases. Proceedings of 23rd International Conference on Very Large Data Bases (VLDB 97).
6. Jianjun Chen, David DeWitt, Fend Tian, and Yuan Wang. NiagaraCQ: A Scalable Continuous Query System for the Internet Databases. ACM SIGMOD, page 379, 2000.
7. Alexandros Labrinidis, Nick Roussopoulos: Adaptive WebView Materialization. WebDB 2001: 85-90
8. Alexandros Labrinidis, Nick Roussopoulos: Balancing Performance and Data Freshness in Web Database Servers. VLDB 2003: 393-404

Chapter 10: Stream-Based Data Management

1. Ron Avnur and Joseph M. Flellerstein. Eddies: Continuously Adaptive Query Processing. SIGMOD Conference 2000: 261-272.

2. Donald Carney, U_ur Cetintemel, Mitch Cherniack, Christian Convey, Sangdon Lee, Greg Seidman, Michael Stonebraker, Nesime Tatbul, Stanley Ldonik. Monitoring Slreams-A New Class of Data Management Applications. VLDB
3. Swarup Acharya, Rafael Alonso, Michael J. Franklin, Stanley B. Zdonik: Broadcast Disks: Data Management for Asymmetric Communications Environments. SIGMOD Conference 1995: 199-210
4. Konstantinos Stathatos, Nick Roussopoulos, John S. Baras: Adaptive Data Broadcast in Hybrid Networks. VLDB 1997: 326-335

Sensor Networks

1. G. Manku and R. Motwani. Approximate frequency counts, over data streams. In VLDB, 2002.
2. Antonios Deligiannakis, Yannis Kotidis, Nick Roussopoulos: Compressing Historical Information in Sensor Networks. SIGMOD Conference 2004: 527-538
3. Antonios Deligiannakis, Minos N. Garofalakis, Nick Roussopoulos: A Fast Approximation Scheme for Probabilistic Wavelet Synopses. SSDBM 2005: 243-252
4. Amit Manjhi, Suman Nath, Phillip B. Gibbons, "Tributaries and Deltas: Efficient and Robust Aggregation in Sensor Network Streams", Sigmod 85
5. Amol Deshpande† Carlos Guestrin‡ Samuel R. Madden§ ‡ Joseph M. Hellerstein† ‡ Wei Hong‡, Model-Driven Data Acquisition in Sensor Networks, VLDB 04

Overlay Networks / Privacy

1. Dimitris Tsoumakos, Nick Roussopoulos "Adaptive Probabilistic Search in P2P" Peer-to-Peer Computing 2003: 102-109
2. Ryan Huebsch, Brent Chun, Joseph M. Hellerstein, Boon Thau Loo, Petros Maniatis, Timothy Roscoe, Scott Shenker, Ion Stoica and Aydan R. Yumerefendi. The Architecture of PIER: an Internet-Scale Query Processor. CIDR 2005.
3. Ryan Huebsch, Joseph M. Hellerstein, Nick Lanham, Boon Thau Loo, Scott Shenker, Ion Stoica. Querying the Internet with PIER. VLDB 2003.
4. Boon Thau Loo, Joseph M. Hellerstein, Ryan Huebsch, Scott Shenker, Ion Stoica. Enhancing P2P File-Sharing with an Internet-Scale Query Processor. VLDB 2004
5. Boon Thau Loo, Ryan Huebsch, Ion Stoica and Joseph M. Hellerstein. The Case for a Hybrid P2P Search Infrastructure. 3rd International Workshop on Peer-to-Peer Systems (IPTPS 04), San Diego, CA, Feb 2004.
6. Jinyang Li, Boon Thau Loo, Joseph Hellerstein, Frans Kaashoek, David Karger, Robert Morris. On the Feasibility of Peer-to-Peer Web Indexing and Search. 2nd International Workshop on Peer-to-Peer Systems (IPTPS 03), Berkeley, CA, Feb 2003
7. Marcin Gomułkiewicz, Marek Klonowski, and Mirosław Kutyłowski, Provable Unlinkability Against Traffic Analysis Already After O (log(n)) Steps!?
8. Roger Dingledine Nick Mathewson Paul Syverson "Tor: The Second-Generation Onion Router" 13th USENIX Security Symposium, San Diego, CA, USA, August 9-13, 2004.
9. Rakesh Agrawal Jerry Kiernan Ramakrishnan Srikant Yirong 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 adms+- 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 attachements, 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 meta data 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
- Randezvous 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 Cellphones and PDAs
- Import tinyDB on PDAs and Cellphones
- Similar to sensor's tinyDB project but with a bit more functionality in the schema