Database Research Lab (DRL) at Polytechnic

Alex Delis

Department of CIS

Polytechnic University

e-mail: ad@naxos.poly.edu

CIS Research Overview Day

October 4th, 2002

Work Done with PhD Graduates

- Vinay Kanitkar, PhD'99 ,first job at Akamai Technologies
  - Real-Time Transaction Processing in a Network of Workstations.
  - Push and Two-stage Transaction Processing Protocols.
- George Kollios, PhD'00, faculty appointment at Boston University
  - Spatiotemporal Indexing Schemes for Animated Objects.
  - Indexing of Moving Objects.
- Je-Ho Park, PhD'00, first job at Voicemate.Com
  - Client-clustering in Networked Databases.
  - Dynamic Client-Clustering.
- Rick Regan, PhD'01, first job at IBM Research
  - Provision of a space efficient recovery mechanism based on the ephemeral logging protocol.

Database Services over Gigabit Networks

- Take advantage and use of
  - Available aggregate bandwidth
  - Quality of Service guarantees
  - Dynamic data replication/migration

  to create database systems that "never" fail.

Database Services over Gigabit Networks

- Issues to be examined:
  - Congestion control for database services (due to skewed access for data accesses, hot-spots, etc.)
  - Organization of the database catalog and coordination
  - Adaptive data replication & continuous data availability
  - Recovery from multiple-node failure
  - Database traffic characterization
  - Automatic configurability & self-tuning
  - Security

Data Management in IP-networks

- Database mirroring over very-wide area networks
  - * Commit protocols in the presence of high network latency (anything better than the usual 2-ϕ Commit?)
  - * Handling of massive and/or frequent updates
- Action-based transaction/request routing
  - * Efficiently route requests from users to servers that can satisfy them (assuming the data is partitioned in many servers)

Data Management in Ad-hoc communities

- Resource sharing in Internet user communities has yielded what has become known as Peer-to-Peer (P2P) computing.
- Key questions:
  1. how is (can be) information/data organized in such networks?
  2. naming of objects, indexing, updates?
  3. management of streaming data
  4. do databases have anything to gain from such "node alliances"?
  5. data consistency
  6. commercial applications?

Resource-Rich Disks for Data Management

CPU Processing + (not trivial) Memory Capacity in the disk drive (+ attachment to the Network):

- Earlier work has shown that "simple" tasks can be "pushed" from processing in main memory down to the disk (loading of application dependent code to the disk-unit CPU).
- What happens with more "complex" operations?
- Collaboration of algorithms running of resource-rich disk drives and traditional approaches running in the conventional memory hierarchy.
- Could farms of networked such disk-devices be used for the storage and (partial) processing of rapidly increasing data sets (such as satellite imagery, biological datasets, etc.)

"Value Added" Database Services

- Create collections of (multimedia) data objects that are not purely for entertainment purposes but carry some "value added" such as:
  - Libraries of Traffic Incidents & Public Emergency situations
  - Surveillance and news video-streams (spatio-temporal data)

"Value Added" Database Services

- What is needed:
  - New feature extraction and indexing methods as well as (possibly) new storage managers
  - Mechanisms that allow "navigation" through data at will (play forward, backward, zoom in/out, clip, etc.)
  - Answer whether specific types of objects appear in specific periods/instances efficiently (for example, "from the traffic incident library, find all vehicles that crashed between 10:00 pm and 12:00 noon")

Other Research Interests

- Parallelization of Data Warehouses.
- 3G Networks (UMTS) and Data Services.
- Organization of Scalable Publish/Subscribe services.
- Real-time Computing Systems.
- Evaluation of Computer Systems.
- Simulation Engines.

PhD Students at DRL

- Zhongqiang Chen
  - Manipulation of spatial GIS objects for the provision of accurate radio-predictions in urban environment while lowering computational overheads.
  - NOW-deployment and evaluation of such methods.
- Vassil Kriakov
  - LAN-based Storage System for Spatio-temporal data.
  - Management of continuous updates and on-the-fly workload balancing.
- Huseyin Akcan

Related WWW pages

naxos.poly.edu/∼ad

naxos.poly.edu/∼vinay

www.cs.bu.edu/∼gkollios

cis.poly.edu/∼jhpark