Next Generation Information Systems

Avi Silberschatz

Department of Computer Science

Yale University

URL: www.cs.yale.edu/~avi

The Digital Age

- Digital information forms the glue for blending the fields of computing, communication and entertainment.
- At the center of this revolution is data that is stored, accessed and delivered in digital format. Some of the major issues surrounding this type of data are:
  - Data is to be available to the users anytime and anywhere and with the desired QoS.
  - Data access must adhere to privacy and security policies.
  - Data Interoperability.
  - Fast access to data, which implies support for queries with approximate answers.
  - Data analysis and mining capabilities over very large datasets.
- Many of the advances in information systems are due to development of new technologies. These advances, in turn, are pushing the developments of even newer technologies.

Research Challenges

- Storage retrieval and delivery of multimedia data
  - Storage System Issues
  - QoS issues of continuous media data (e.g., video and audio)
- Approximate answers
  - useful for very large data sets
  - useful for Web searching
- Data mining
  - Discovering “interesting” patterns in very large data sets
  - Discovering “interesting” patterns from incomplete information
- Data Interoperability
- Privacy and security
- Next generation Networks
  - Converged networks
  - Network Management

Multimedia Data

- Regular Data
  - text, binary, image
- Database Data
  - tuples, objects
- Continuous Media Data
  - Video Data
    - The display (playback) of the data must be continuous with a fixed rate, which is typically 30 frames/second.
    - A viewer may wish to control the way the data is to be displayed by applying various VCR-type operations to the video data.
  - Audio Data
    - The playback must be continuous with a fixed rate, which is dependent on the sample rate.
    - A listener may wish to control the way the data is played back.

Storage System Issues

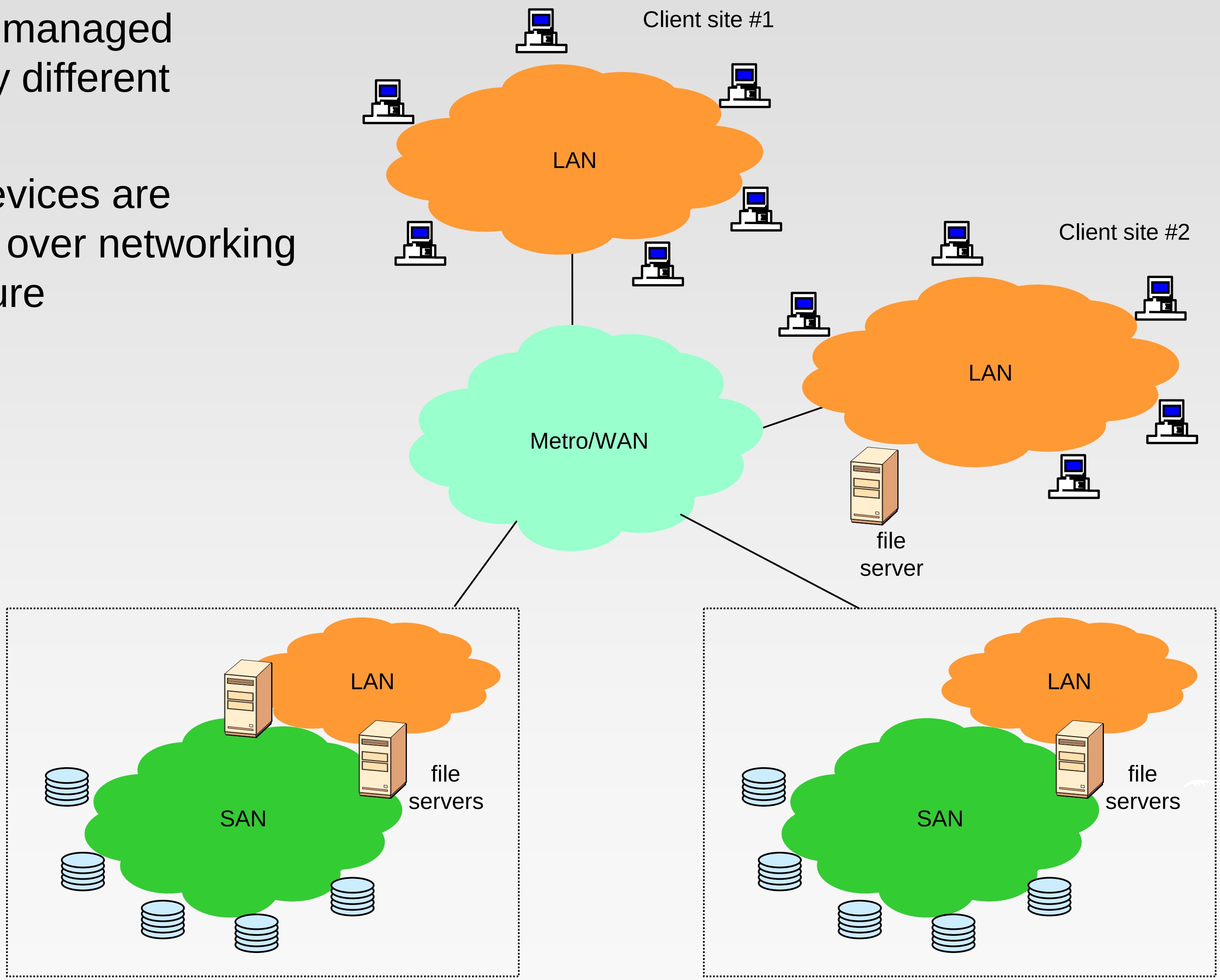
- Rapid growth in storage capacity demand
  - world-wide installed storage:
    - 738 PetaByte in 2000
    - over 75% per year storage capacity increase over the next 5 years
    - reaches ZettaByte in 2009
  - data stored at Global 2500 companies double every 18 months
  - data stored at e-commerce companies grow at 400% a year
- Management
  - 40-50% of company IT budget is spent on storage
  - fraction of IT budget spent on storage is expected to grow
  - cost for storage management exceeds cost of storage equipment
    - management: $300 per GB per year
    - low-end storage: $14 - $50 per GB (packaged, powered, networked)
  - management cost is expected to grow
- Storage Requirement
  - 24 x 7
  - Disaster recover

Storage is Moving Into the Network

- Motivation
  - Use commodity IP based networks
  - IT staff know-how
  - Distance and universal access
- Applications
  - Disaster recovery
  - Archiving
  - Backups
  - Content Distribution
  - Managed storage
  - Value added storage services
  - Consolidation of storage

IP-Based Network Storage

- Storage is managed possible by different domains
- Storage devices are connected over networking infrastructure



IP-based Network Storage (Cont.)

- IETF standards are being drafted
  - Most popular: iSCSI and FCIP
  - Almost all networking and storage companies are participating in these standards
- Issues
  - Performance
  - Reliability
- Future
  - end-to-end iSCSI;
    - end-to-end IP storage networking?
    - demise of FC?
  - Hybrid?
    - FC (InfiniBand) SAN islands connected over IP networks
    - FC SANs in data centers accessed by IP networks

Network Storage Security

- Customers may not trust the storage service provider (SSP)
- Storage consolidation over different customers is essential to make storage outsourcing viable. However, customers may not trust each other
- Threat model
  - Disclosure of data to an eavesdropper intercepting communication
  - Disclosure of data to storage service provider (SSP) and to other customers of the SSP
  - Manipulation of communication by an attacker
  - Manipulation of data by the SSP or other customers of the SSP
- Challenges
  - high throughput encryption (e.g., 1Gbps, 10 Gbps)
  - security without hindering performance

Multimedia Storage and Delivery Issues

- The size of some databases is enormous, especially those that are used for data mining (e.g., cash register transactions).

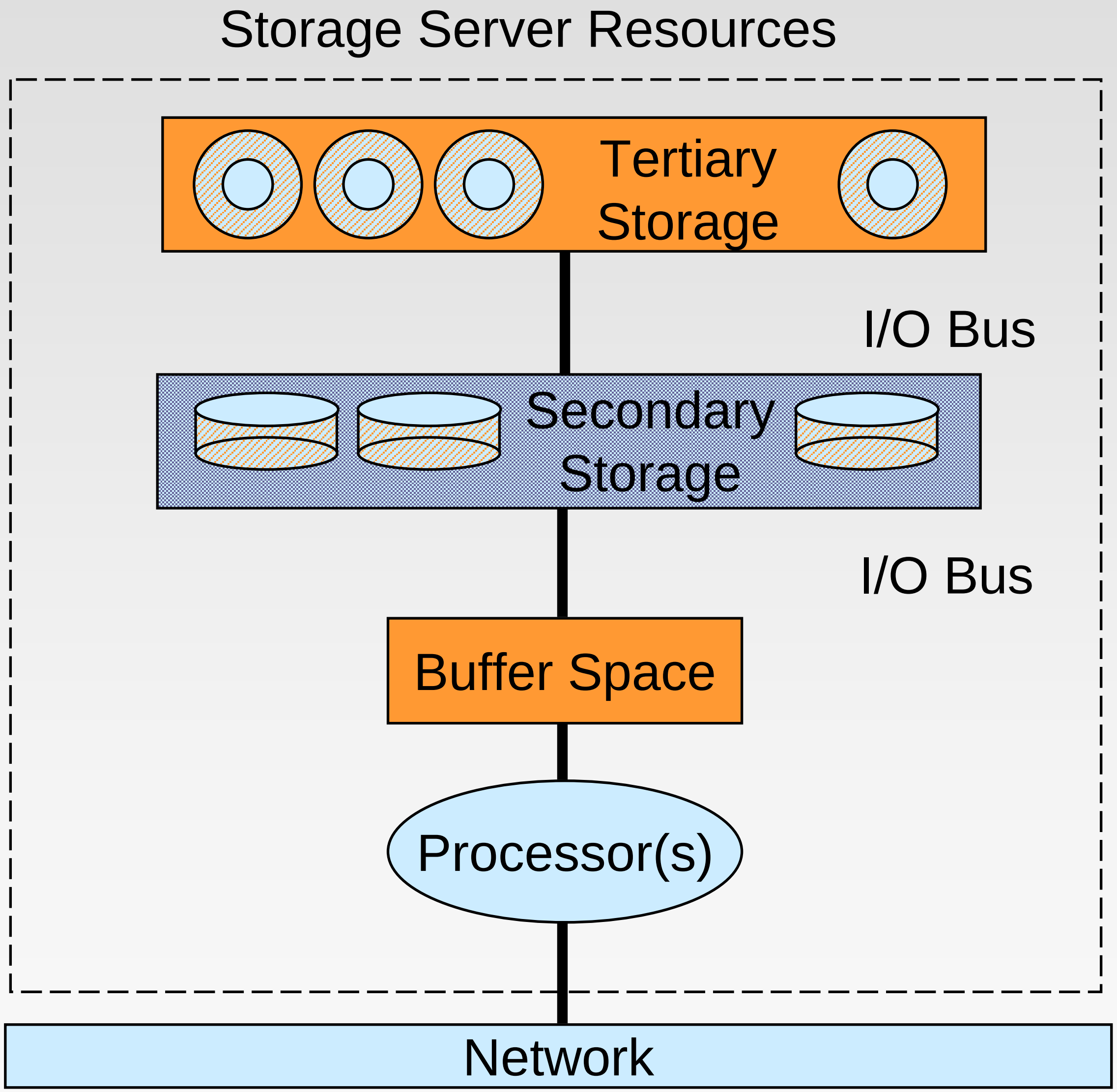
30 terabytes largest commercial database

- Some information sources generate data at an astonishing rate (e.g., satellite images).

EOS – 1-2 terabytes per day

- The BBC is planning to digitize the last 50 years of programming.
- Continuous media data is voluminous:
  - 100 minute MPEG-1 video requires 1.125GB
  - 100 minute HDTV video requires 15GB
- Continuous media data require support for QoS.

System Resources to be Managed for QoS



Research Issues

- Admission control
- Disk Scheduling
- Buffer Management
- Storage Management
  - data layout
  - varying disk transfer rates
  - disk striping
  - meta data
  - fault-tolerance
- Tertiary storage

Cycle-based Scheduling

- Let T be the length of a service cycle
- Maintain a queue of requests $R_1, R_2 \ . \ . \ . \ R_n$. Each R_i corresponding to a request to view a CM clip. Each request has an associated rate r_i.
- For each request, a buffer is allocated of size $2 * T * r_i$.
- Requests in the queue are served in a cyclic order using double buffering. In each cycle I:
  - get data from disk to buffer (I mod 2)
  - transfer data from the (I + 1 mod 2) buffer to the client

Disk Scheduling

- Request are serviced in service cycles (rounds).
- In the beginning of a service cycle requests are ordered in C-SCAN order.
- In the beginning of every service cycle, it is ensured that

  - $\sum 2 \cdot T \cdot r_i \leq B$
  - $\sum \left(\frac{T \cdot r_i}{r_{disk}} + t_{rot} + t_{settle} \right) + 2 \cdot t_{seek} \leq T$

  hold. (where t_{rot}, t_{settle}, t_{seek} are the rotational delay, settle time, and seek time, respectively, and B is the buffer pool size).
- The value of T is adjusted depending on the workload.
- In every service cycle,

  $$\min \left\{ T \cdot r_i,\ 2 \cdot T \cdot r_i - (\text{offset of last retrieved - offset of last consumed}) \right\}$$

  bits of data retrieved for each request.

Admissions Control

Queue is bounded by an admission control scheme

- For each request, the service time for a request is estimated.
- A request is admitted only if the sum of the estimated service times for all admitted requests does not exceed the duration of service cycle T.

Admission Control (cont.)

- Reserve a fraction of service cycle T, say $\rho \cdot T \ (0 < \rho \leq 1)$ for continuous media requests.
- A request (real-time, non-real-time), is admitted if
$$\sum\left(\frac{T \cdot r_i}{r_{disk}} + t_{rot} + t_{settle}\right) + \sum\left(\frac{n_i}{r_{disk}} + t_{rot} + t_{settle}\right) + 2 \cdot t_{seek} \leq T$$
- A real-time request is admitted if
$$\sum\left(\frac{T \cdot r_i}{r_{disk}} + t_{rot} + t_{settle}\right) + 2 \cdot t_{seek} \leq \rho \cdot T$$
- Above scheme ensures
  - both continuous and non-continuous media requests are allocated time during a service cycle.
  - any time during a service cycle unused by continuous media requests is allocated to non-continuous media requests.

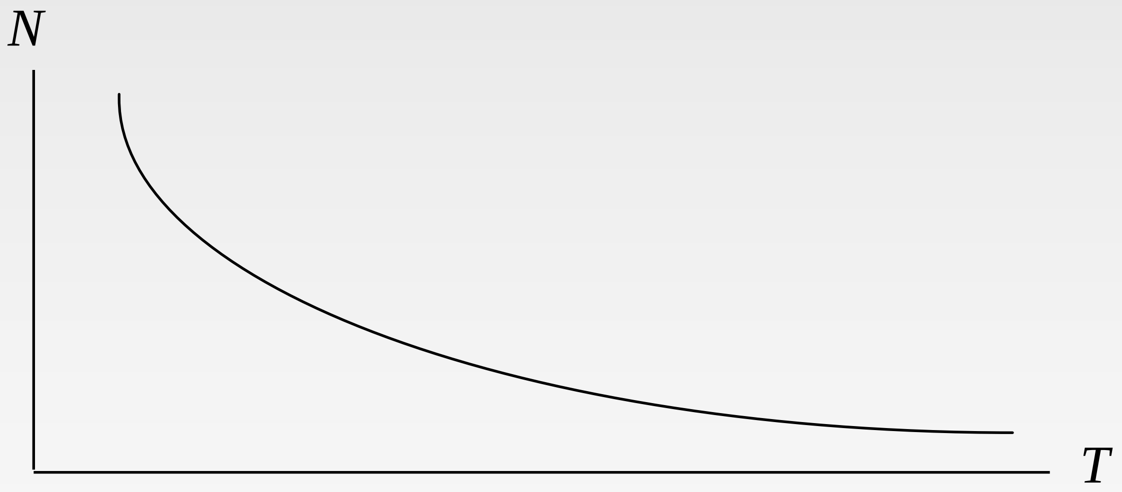
Length of T

What about the length of T?

Buffer Space Constraints

- Let B be the available buffer size
- Let N be the number of admitted clients
- Assume infinite disk bandwidth
- Requirements:

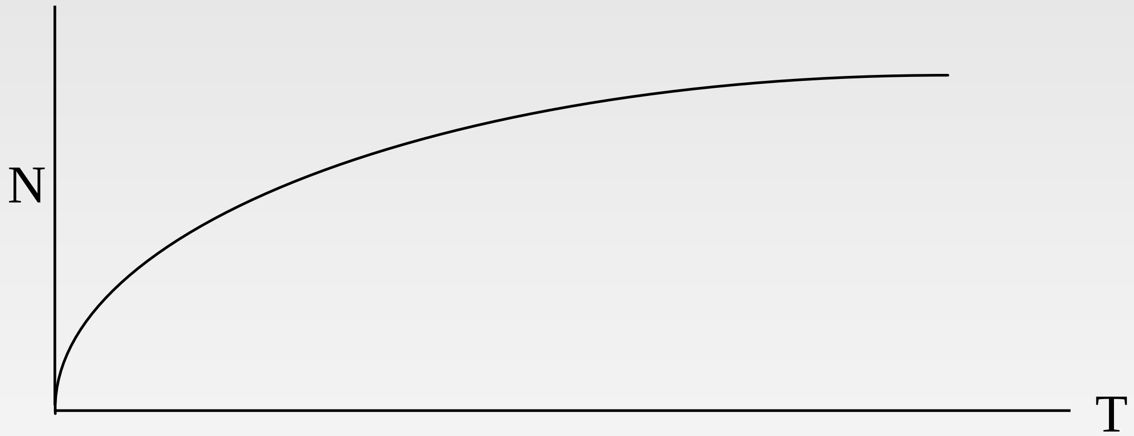
$$\sum_{i=1}^{N} 2 * T * r_i \leq B$$



For a given buffer size B, the larger T, the fewer clients can be admitted.

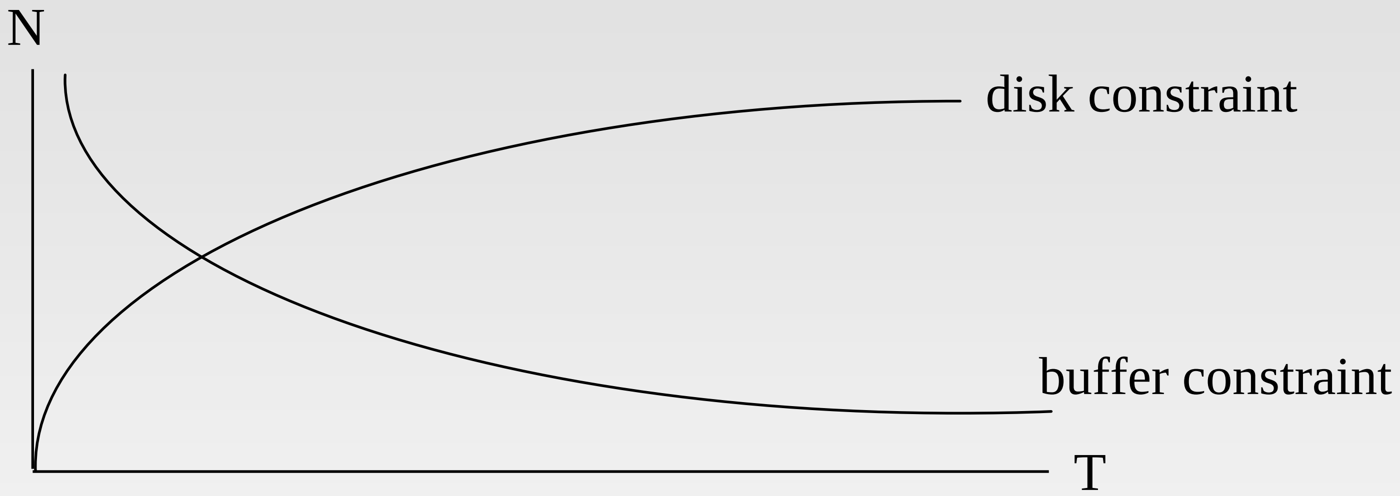
Disk Bandwidth Constraints

- Assume infinite buffer space
- Use C-SCAN disk scheduling
- Requirements: $2 * t_{settle} + N * (t_{rot} + t_{settle}) + \sum_{i=1}^{N} \frac{T * r_i}{r_{disk}} \le T$



The larger T the larger N is

Combining Disk & Buffer Constraints



The optimal T is obtained by solving a quadratic equation of the disk and buffer space constraints.

Minimizing Response Time

- Under some workloads (e.g., request with small r_i's such as 64 Kbps), the value of T that maximizes throughput can be high (e.g., 20 secs.).
- This might yield high response times.
- Solution:
  - maintain small T values
  - in order not to degrade throughput, for each request R_i data is prefetched from disk in every k_i service cycles (instead of in every service cycle)
  - The maximum amount of data prefetched is $k_i \cdot T \cdot r_i$
  - buffer space allocated to R_i is $(k_i + 1) \cdot T \cdot r_i$

Minimizing Response Time (contd.)

- Issues:
  - Calculation of k_i's
  - Admission control:
    - $lcm(k_1, k_2, \ldots, k_n)$ service cycles to manage
    - For a request R_i, finding the least loaded service cycles

$$u_i + k_i \cdot l,\ 0 \le l \le \frac{lcm(k_1, k_2, \ldots k_n)}{k_i} - 1$$

    - In order to reduce response time, start a new request R_i in the first possible service cycle and then move it incrementally to the selected least loaded service cycle.
- This solution also provides higher throughput for workloads with small r_i's

Querying Huge Data Sets

- Give me all objects (e.g., images) that look like this.
- If we are dealing with PetaBytes of data, this may take days or weeks.
- One solution is to capture “meta data” information about the stored objects as the objects are stored in the database.
  - Querying is done against the “meta data”.
  - Major issue – nature of the meta data.
- Another solution is to provide support for “approximate answers”.

Providing Approximate Answers

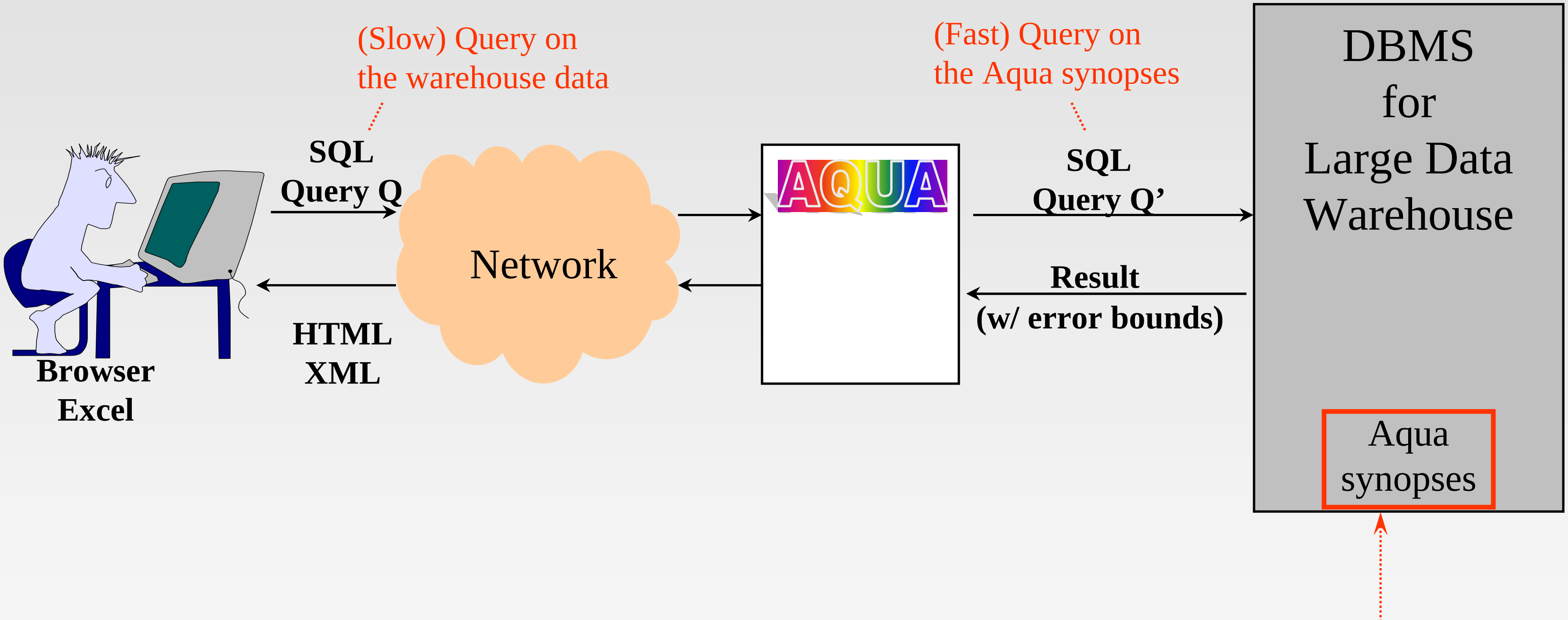
- Traditional databases provide exact answers to queries, but...
- In massive data environments, can take minutes to hours due to disk I/Os
- In distributed environments, data may be remote or currently unavailable
- In real-time environments, even single I/O may be too slow

Providing Approximate Answers (Cont.)

- Trade-off accuracy for performance: e.g., 30 minutes for exact answer vs. 3 seconds for an approximate answer with 5% error
- Examples where fast approximate answers are preferred:
  - drill-down query sequence in data mining: searching for the “interesting” queries
  - tentative answer when base data unavailable
  - leading digits suffice (e.g., 3.5 million vs. 3.512 million)
- Can proceed to the exact answer, if desired

The AQUA System

Approximate Query Engine for data warehousing



- Aqua precomputes and maintains small synopses of the data
- Aqua provides approximate answers with accuracy guarantees, by rewriting user queries as depicted above

Aqua Synopses: The Key Ingredient

- (Small) Surrogate for the actual data.
- Must accurately estimate the exact answers from the synopses.
- As data is updated, must keep synopses up-to-date.

We developed new techniques for summarizing data, and for adapting these summaries to changes in both the data and the query mix.

First system to provide fast, highly-accurate approximate answers for a broad class of queries arising in data warehousing scenarios

Private, Public, and Sensitive Information in a Wired World

- Private information
  - Only the data subject has a right to it.
- Public information
  - Everyone has a right to it.
- Sensitive information
  - “Legitimate users” have a right to it.
  - It can harm data subjects, data owners, or data users if it is misused.

Erosion of Privacy

- "You have zero privacy. Get over it." – Scott McNealy, 1999
- Changes in technology are making privacy harder.
  - increased use of computers and networks
  - reduced cost for data storage
  - increased ability to process large amounts of data
- Becoming more critical as public awareness, potential misuse, and conflicting goals increase.

"Public Records" in the Internet Age

- Depending on State and Federal law, "public records" can include:
  - Birth, death, marriage, and divorce records
  - Court documents and arrest warrants (including those who were acquitted)
  - Property ownership and tax-compliance records
  - Driver's license information
  - Occupational certification

  They are, by definition, "open to inspection by any person."
- Traditionally: Many public records were "practically obscure."
  - Stored at the local level on hard-to-search media, e.g., paper, microfiche, or offline computer disks.
  - Not often accurately and usefully indexed.

  Now: More and more public records, especially Federal records, are being put on public web pages in standard, searchable formats.
- Issues
  - Should some Internet-accessible public records be only conditionally accessible?
  - Should data subjects have more control?
  - Should data collectors be legally obligated to correct mistakes?

Examples of Sensitive Information

- Copyright works
- Certain financial information
- Health Information
- Question: Should some information now in “public records” be reclassified as “sensitive”?

State of Technology

- We have the ability (if not always the will) to prevent improper access to private information. Encryption is very helpful here.
- We have little or no ability to prevent improper use of sensitive information. Encryption is less helpful here.

The PORTIA Project

- PORTIA: Privacy, Obligations, and Rights in Technology of Information Assessment
- Large ITR grant from NSF. It is five-year multi-institutional, multi-disciplinary, multi-modal research project on end-to-end handling of sensitive information in a wired world
- Researchers from:
  - Stanford: Dan Boneh, Hector Garcia-Molina, John Mitchell, Rajeev Motwani
  - Yale: Joan Feigenbaum, Ravi Kennan, Avi Silberschatz
  - University of NM: Stephanie Forrest
  - Stevens Institute: Rebecca Wright
  - NYU: Helen Nissenbaum
  - Plus participation by software industry, key user communities, advocacy organizations, and non-CS academics.
- http://crypto.stanford.edu/portia

PORTIA Goals

- Produce a next generation of technology for handling sensitive information that is qualitatively better than the current generation's.
- Enable end-to-end handling of sensitive information over the course of its lifetime.
- Formulate an effective conceptual framework for policy making and philosophical inquiry into the rights and responsibilities of data subjects, data owners, and data users.

Five Major Research Themes

- Privacy-preserving data mining and privacy-preserving surveillance
- Database policy enforcement tools
- Sensitive data in P2P systems
- Policy-enforcement tools for database systems
- Identity theft and identity privacy

Network System Challenges

- Next-generation network -- will be simpler, lower cost, and will provide customized services for consumers and businesses
- Converged networks -- will incorporate the best features of today's voice and data networks
- Network management – automate many of the functions that are currently done by people.

Network Management Challenges

- Managing today's networks is extremely challenging due to their increased complexity
  - Networks contain hundreds of network elements and thousands of physical links
  - Network elements follow a multitude of protocols (e.g., BGP, OSPF, ISIS, RIP)
  - Networks are heterogeneous and contain equipment from multiple different vendors
- Manually managing networks
  - is tedious, labor-intensive, time-consuming and error-prone
  - is not cost-effective due to severe shortages of and high costs of skilled labor
- Critical need for software tools that automate network management tasks

Next-Generation Network Management

- Next-Generation network management software functionality includes
  - Keeping track of network inventory and topology
  - Monitoring network link bandwidth and latency
  - Storing, analyzing and reporting network performance data
  - Load balancing by appropriately configuring network parameters
  - Automating and simplifying network configuration tasks (e.g., VPNs)
- Value Proposition:
  - Ease management and configuration of ISP networks
  - Optimize utilization of network resources
- Goal: Make networks self-administering and self-tuning