History of the field of databases

- Late 60's: **network** (CODASYL) & **hierarchical** (IMS) DBMS.
  - Low-level “record-at-a-time” DML, i.e. physical data structures reflected in DML (no data independence)
- 1970: **Codd's paper -- the relational model**. The most influential paper in DB research.
  - Set-at-a-time DML. Data independence. Allows for schema and physical storage structures to change under the covers. Truly important theory, led to "paradigm shift" in thinking and in practice.
  - Papadimitriou: "as clear a paradigm shift as we can hope to find in computer science".
  - **Turing award**

History of the field of databases

- early-to-mid-70's
  - raging debate between the two camps.
  - "great debate" in 1975
- mid 70's: 2 full-function (sort of) prototypes
  - Ingres
  - System R
  - Ancestors of essentially all today's commercial systems

History of the field of databases

- Ingres: UCB 1974-77
  - a ``pickup team", including Stonebraker & Wong early and pioneering. Led to Ingres Corp (CA), Sybase, MS SQL Server, Britton-Lee, Wang's PACE.
- System R: IBM San Jose (now Almaden)
  - 15 PhDs. Led to IBM's SQL/DS & DB2, Oracle, HP's Allbase, Tandem's Non-Stop SQL. System R arguably got more stuff ``right"
- Both were viable starting points, proved practicality of relational approach. Beautiful example of theory -> practice!!

History of the field of databases

- early 80's
  - commercialization of relational systems
- mid 80's
  - SQL becomes “intergalactic standard”.
  - DB2 becomes IBM's flagship product.
  - IMS “sunseted”

History of the field of databases

- 90's: the age of maturity
  - network & hierarchical essentially dead (though commonly in use!)
  - relational becomes mainstream
  - improvements in terms of transactional facilities, performance and stability
  - Scale, scale, scale…

5

Scale, scale, scale…

- EOSDIS*: 1 Tb/day, keep it all for 15 years (they need tertiary storage for that)

  **NASA's Earth Observing System Data and Information System*
- WalMart: 365 node system, 6Tb online, 4billion row table, 200million updates daily, 4000 queries/day, 1500 users/week, 4 min DS response time w/ avg. 60000 rows

*Databases make the world go round, mainly due to their ability to handle **HUGE** amounts of data, **RELIABLY!!!***

***Large scale** is our business…*

6

History of the field of databases

- Late 90's: object relational & the web
  - SQL-1999 & early implementations
  - support for ADT's
  - RDBMS's as back-end for internet front-ends
  - Application Servers and middleware

7

TODAY - The Lowell report 2003

- Senior database researchers gather every few years to assess the state of database research and to recommend problems and problem areas that deserve additional focus.
- The previous meetings were held in Laguna Beach, Ca. in 1989, in Palo Alto, Ca. (Lagunitas) in 1990, in Palo Alto, Ca. (Lagunitas II) in 1995, and at Asilomar, Ca. in 1998.
- The sixth ad-hoc meeting was held May 4-6, 2003 in Lowell, Mass., USA.

http://research.microsoft.com/~Gray/Lowell/

8

Issues for future research

- (data)Bases for everything
- Information Fusion
- Multimedia Querying
- Uncertain data & Personalization
- Data Mining
- Privacy & Trustworthy Systems
- New User Interfaces
- 100 year storage

… no more data bases …

…, it is time to stop grafting new constructs onto the traditional architecture of the past. Instead, we should rethink basic DBMS architecture with an eye toward supporting:

- Structured data
- Text, space, time, image, and multimedia data
- Procedural data, that is data types and the methods that encapsulate them
- Triggers
- Data Streams and queues

as co-equal first class components within the DBMS architecture — both its interface and its implementation — rather than as afterthoughts grafted on a relational core.

The participants were adamant that one should start with a clean sheet of paper.

Issues for future research

- Information Fusion: Therefore, one must perform information integration on-the-fly over perhaps millions of information sources. … the thorny problem of semantic heterogeneity remains …
- Multimedia Querying: … to create easy ways to analyze, summarize, search, and view the “electronic shoebox” of a person’s multimedia information.
- Uncertain data: …query processing must move from a deterministic model, where there is an exact answer for every query, to a stochastic one, where the query processor performs evidence accumulation to get a better and better answer to a user query.

11

Issues for future research

- Data mining: users … wish for tools that generate some “pearls of wisdom”.
- A challenge for data mining research is to develop algorithms and structures for sifting through the databases looking for such pearls, while running in background and consuming excess system resources.
- Another important challenge is to integrate data mining with database querying, optimization, and other facilities such as triggers.

12

Issues for future research

- Privacy: our community can work on security systems that include a component dealing with the prospective use to which the data will be put. Access decisions should be based not only on who is requesting the data but also on what use it will be put to.
- New User Interfaces: There is a crying need for better ideas in this area.

13

Issues for future research

- 100 year storage: even archived information is disappearing, because it was captured on a medium that is deteriorating (e.g. photographic film or magnetic tape) or because it was captured on a medium that requires obsolete devices (e.g. special storage drives), or because the application that is needed to interpret the information no longer works (e.g. troff).
- [we need] mechanisms for migration, to copy information from deteriorating or obsolete media, and for emulation, to capture methods that can interpret information that is stored for long periods (e.g. troff renderer)

14