Lecture #01: Course Introduction and History of DBMSs

15-721 Advanced Database Systems (Spring 2018)
http://15721.courses.cs.cmu.edu/spring2018/
Carnegie Mellon University
Prof. Andy Pavlo

1 Database History Observations

- Stonebraker argues that a lot of the issues relevant a long time ago are still relevant today [11].
- The "SQL vs. NoSQL" debate of the 2010s is reminiscent of the "Relational vs. CODASYL" debate from the 1970s [8].

2 The 1960s

- **First DBMS - Integrated Data System (IDS)** [1, 6]
  - Developed internally at GE in the early 1960s by Charles Bachman.
  - GE sold their computing division to Honeywell in 1969.
  - Based on the *Network Data Model* that supported *Tuple-at-a-time* query execution [3].
- **CODASYL** [12]
  - Proponents of COBOL and the network data model people got together and proposed a standard for how programs will access a database.
  - The *Network Data Model* made it difficult to write complex queries due to the sets that maintained relationships. It also made these early systems more susceptible to corruption [11].
- **IBM Information Management System (IMS)** [7]
  - Early database system developed to keep track of purchase orders for Apollo moon mission.
  - Based on the *Hierarchical Data Model* that organized collections of data with parent/child relationships.
  - Programmer-defined physical storage format (e.g., hash table vs. tree) and tuple-at-a-time query execution.

3 The 1970s

- **Relational Model** [5]
  - Ted Codd was a mathematician working at IBM Research who saw developers constantly changing their codebase whenever the database's schema changed.
  - Codd created the relational model abstraction to avoid this maintenance based on three key ideas:
    1. Store database in simple data structures
    2. Access data through high-level language
    3. Physical storage left up to implementation
- Early implementations of relational DBMSs:

- IBM Research: **System R**
- U.C. Berkeley (Mike Stonebraker): **INGRES**
- Relational Software, Inc. (Larry Ellison): **Oracle**

4   The 1980s

- The relational model wins the database marketplace over CODASYL.
  - IBM released their first relational DBMS (**DB2**) in 1983.
  - System R's "SEQUEL" declarative query language becomes the standard (later renamed to "SQL").
  - Many new enterprise DBMSs are invented (**Informix**, **Sybase**, **TeraData**) but Oracle wins marketplace.
  - Stonebraker leaves INGRES, returns to Berkeley and starts the Postgres project.
- **Object-Oriented Databases** [13]
  - Argued that how people wrote code and how data is stored in a database is different
  - Avoid "relational object impedance mismatch" by tightly coupling objects and database.
  - Few of these original DBMSs from the 1980s still exist today. These systems performed poorly when executing complex queries. There was also no standard API or programming language.

5   The 1990s

- No major advancements in database systems or application workloads.
- Microsoft forks **Sybase** and creates **SQL Server** for Windows NT.
- **MySQL** is written as a replacement for **mSQL**.
- Illustra (the commercial version of **Postgres**) gets bought by **Informix**. Graduate students at Berkeley take the original academic **Postgres** code and adds support for SQL.
- **SQLite** started in early 2000.

6   The 2000s

- **Data Warehouses**
  - Distributed / Shared-Nothing
  - Relational / SQL
  - Usually closed-source
  - Significant performance benefits from using *Decomposition Storage Model* (i.e., columnar storage).
- **NoSQL Systems** [4]
  - Focus on high-availability and high-scalability
  - Schema-less
  - Non-relational data models
  - No ACID transactions

- Custom APIs instead of SQL
- Usually open-source

7   The 2010s

- **NewSQL Systems** [2, 9]
    - Strive to provide same scalability and performance for OLTP workloads as NoSQL DBMSs without giving up ACID.
    - Relational / SQL
    - Distributed
    - Usually closed-source
- **Hybrid Transactional-Analytical Processing (HTAP)** [10]
    - Execute fast OLTP like a NewSQL system while also executing complex OLAP queries like a data warehouse system.
    - Distributed / Shared-Nothing
    - Relational/SQL
    - Mixed open/closed-source
- **Cloud-based Database Systems**
    - First database-as-a-service (DBaaS) offerings were "containerized" versions of existing DBMSs (e.g., Amazon RDS).
    - There are new DBMSs that are designed from scratch explicitly for running in a cloud environment.
- **Specialized Database Systems**
    - Shared Disk (HDFS, EBS)
    - Embedded
    - Time-Series
    - Multi model
    - Blockchain

References

[1] Oral-History:Charles Bachman. `http://ethw.org/Oral-History:Charles_Bachman`, April 2011.

[2] M. Aslett. How will the database incumbents respond to NoSQL and NewSQL? The 451 Group, April 2011. URL `http://www.the451group.com/report_view/report_view.php?entity_id=66963`.

[3] C. W. Bachman. On a generalized language for file organization and manipulation. *Commun. ACM*, 9 (3):225–226, 1966. URL `http://doi.acm.org/10.1145/365230.365272`.

[4] R. Cattell. Scalable sql and nosql data stores. *SIGMOD Rec.*, 39:12–27, 2011. URL `http://doi.acm.org/10.1145/1978915.1978919`.

[5] E. F. Codd. A relational model of data for large shared data banks. *Commun. ACM*, 13(6):377–387, June 1970. URL `http://doi.acm.org/10.1145/362384.362685`.

[6] T. Haigh. How charles bachman invented the dbms, a foundation of our digital world. *Commun. ACM*, 59(7):25–30, June 2016. URL `http://doi.acm.org/10.1145/2935880`.

[7] B. Klein, R. A. Long, K. R. Blackman, D. L. Goff, S. P. Nathan, M. M. Lanyi, M. M. Wilson, J. Butterweck, and S. L. Sherrill. *An Introduction to IMS: Your Complete Guide to IBM Information Management System*. IBM Press, 2nd edition, 2012.

[8] A. S. Michaels, B. Mittman, and C. R. Carlson. A comparison of the relational and codasyl approaches to data-base management. *ACM Comput. Surv.*, 8(1), Mar. 1976. URL `http://doi.acm.org/10.1145/356662.356668`.

[9] A. Pavlo and M. Aslett. What's really new with newsql? *SIGMOD Rec.*, 45(2):45–55, Sept. 2016. URL `http://doi.acm.org/10.1145/3003665.3003674`.

[10] M. Pezzini, D. Feinberg, N. Rayner, and R. Edjlali. Hybrid Transaction/Analytical Processing Will Foster Opportunities for Dramatic Business Innovation. `https://www.gartner.com/doc/2657815/`, 2014.

[11] M. Stonebraker and J. Hellerstein. What goes around comes around. *Readings in Database Systems*, 4, 2005.

[12] R. W. Taylor and R. L. Frank. Codasyl data-base management systems. *ACM Comput. Surv.*, 8(1): 67–103, Mar. 1976. URL `http://doi.acm.org/10.1145/356662.356666`.

[13] S. B. Zdonik and D. Maier, editors. *Readings in Object-oriented Database Systems*. Morgan Kaufmann Publishers Inc., 1990. ISBN 0-55860-000-0.